Title
Essays on Public Finance and Development

Author
Cohen, Isabelle M

Publication Date
2021

Peer reviewed|Thesis/dissertation

Essays on Public Finance and Development

by

Isabelle M Cohen

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Frederico Finan, Chair
Professor Edward Miguel
Professor Ernesto Dal Bó

Spring 2021

Essays on Public Finance and Development

Abstract

Essays on Public Finance and Development

by

Isabelle M Cohen

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Frederico Finan, Chair

This dissertation consists of three contributions to the literature at the intersection of development economics and public economics, set in the context of Uganda but with broader applications. In each of the three chapters, I use rigorous identification and a rich variety of administrative data to answer interesting and challenging questions about a variety of ways in which the public sector influences and is influenced by economic development.

The first chapter studies the subnational fiscal response to international aid, and examines how other donors and the Ugandan government respond to exogenous increases in the amount of World Bank funds allocated to local projects. While similar analysis has been done before at the country level, my analysis is the first that is focused on the subnational level. I find evidence of a "crowding in" of donors, whereby some other donors follow the lead of the World Bank in determining where to allocate funding across Uganda's smaller administrative units. The crowding in of aid, and the subsequent potential for overspending and diminishing marginal returns, may help explain why international aid sees large returns viewed locally, but smaller returns on national levels.

The second chapter focuses on a randomized control trial which I conducted with the Uganda Revenue Authority (URA), studying tax compliance in the context of low state capacity. Although the ability of the state to collect tax revenue is crucial for development, relatively little is known about what drives tax compliance and how it can be improved in poor, low-capacity countries. I show that low-cost, easily implementable tax compliance interventions can succeed, testing a highly effective tax encouragement scheme in conjunction with URA. I find a 6x rate of return to sending a simple text message to prospective taxpayers in the days before taxes are due, which increases to 13x when considering an enforcement-focused message; this intervention is estimated to have raised over $12,500 USD. To further understand the levers of tax compliance, I build an extensive, granular, nationwide dataset of public good provision, and utilize machine learning tools to estimate state capacity using this dataset; I then examine the heterogeneity of the treatment effect across different levels

of state capacity. I find that the treatment was most effective in areas where state capacity is low, and particularly in areas where state capacity was low and there have been recent investments in public goods. These results suggest that while deterrence-based methods may be most effective across the board, there is a crucial role for fiscal exchange in encouraging tax compliance in low-capacity environments.

The third chapter focuses on the effects of administrative unit proliferation, or the creation of new subnational administrative units, on the provision of governmental services and the processes of economic development. I utilize a variety of rich data sources on budgets, educational personnel and infrastructure, household well-being, economic activity, political attitudes and voting behavior in a differences-in-differences framework to examine the effects of the 2009-2010 wave of district creation in Uganda. I compare parent districts, which inherit the administrative capacity of their predecessors; newly formed districts, which must build administrative capacity from the ground up; and non-splitting districts, which are unchanged, while controlling for a variety of underlying static and temporally varying characteristics. I find that parent districts benefit from the split in terms of service provision and medium-run economic development, suggesting that smaller districts may have been better able to respond adequately to the needs of the population. However, new districts benefit from the split only in terms of service provision and not in terms of economic development, suggesting that a lack of adminstrative capacity may have hampered their internal growth. These findings suggest that even under favorable conditions for the creation of new districts, a lack of adequate staffing capability may hinder the benefits of decentralization.

Jointly, these three papers address three topics crucial to the fields of both development and public economics: fiscal response, taxation, and decisions on governmental size. In other words, this dissertation addresses how countries raise funding, both internally and externally, and where and how capably they spend it. Although I study these questions in the context of Uganda, all three papers have broader implications, whether on the dynamics of international aid, the importance of state capacity to taxation (and vice versa), or the dynamics of economic development under insufficient administrative capacity.

In memory of my grandmother, Marna Cohen

Contents

List of Figures

List of Tables

Acknowledgments

I am deeply and profoundly grateful to the many people who have helped me along the journey to completing this Ph.D.. I want to acknolwedge the tremendous support provided by my advisor, Fred Finan, along so many different dimensions, for whose encouragement I am thorougly grateful. I also want to deeply thank my committee members Edward Miguel and Ernesto Dal Bó, who have been endlessly generous with their time and support. I am also grateful to all those who have provided support and feedback on these and other projects that have gone in to my Ph.D., including Katerina Linos, Supreet Kaur, Danny Yagan, Emmanuel Saez, and many others in the Department of Economics and beyond. I also thank the staff of the Department of Economics, particularly Patrick Allen, who have provided a wonderful wealth of support.

In addition, I want to acknowledge the enormous role that many of my classmates have played in academic and non-academic support. Those to whom I'm grateful include Christina, Junyi, Eric, Sam, Max, and Preston, among many others, for invaluable advice and guidance, and many fond memories of my time in Berkeley. I am also grateful for the D-Lab for providing me with a vibrant and thoughtful community outside of my own department, and helping me broaden my scope.

There are many in Uganda who have worked to make this research possible. I am profoundly grateful to my colleague, mentor and friend Kizito Omala, and to all the support I have been so fortunate to receive from him. I also owe an enormous debt of gratitude to Nicholas Musoke at the Uganda Revenue Authority, whose investigative spirit and deep commitment to research I am thoroughly grateful for, as well as his colleagues Tina Kaidu Barugahara and Milly I. Nalukwago, and former colleague Dennis Godwin Tumusiime, all of whom have made possible our collaborative research efforts. I also extend a special thanks to the domestic tax department for their assistance. I am grateful as well to Innovations for Poverty Action (IPA) Uganda, including in particular Carin Mirowitz, Vianney Mbonigaba, and Elly Atuhumuza; I also thank each of the twenty seven surveyors who spent weeks calling thousands of Ugandan businessmen in the summer of 2019.

I also want to thank all the undergraduate research assistants I've had over the years at the University of California, Berkeley, as well as Martha Olney, both for all her work in setting up a platform for undergraduate researchs and for her support and guidance. Although I am grateful to all of them, I owe a particular thanks to Avani Yadav, Shrey Vasavada and Qiping Zhang, whose heroic efforts at geocoding and data cleaning made the second chapter of this thesis possible.

This dissertation has also benefited from the generous financial support of several sponsors. I thank the Center for Effective Global Action, the Abdul Latif Jameel Poverty Action Lab's Governance Initiative, and the International Growth Centre for their support and generosity.

I also want to acknowledge the many organizations who put together the data which made this thesis possible. This includes not only the Uganda Revenue Authority, but also the Ugandan Ministry of Education and Sports, the Ugandan Ministry of Health, the Ugandan

Bureau of Statistics, the Ugandan Justice, Law and Order Sector, the Ugandan Ministry of Finance, Planning and Economic Policy, Afrobarometer, and the World Bank; I thank all who provided me with data and information at each of these institutions. I also want to thank the AidData Research Lab from the College of William & Mary, not only for providing me with the wonderful opportunity to plunge in to research during my undergraduate days, but also for the wealth of fascinating data they have made public then and since.

These acknowledgements would not be complete without the recognition of those who supported and encouraged me during my undergraduate and master's level studies at the College of William & Mary, including Arnab Basu, Dan Cristol, David Feldman, Sarah Stafford, Michael Tierney, and many others.

Last, and rather the opposite of least, I want to thank my family for their constant, unceasing support, which is the foundation on which all my success, academic or personal, rests. I thank my father, Paul Cohen, and my mother, Nancy Masters, who have been loving and stalwart in their belief that I could do anything I set my mind to, and my sister, Amelia, my lifelong friend. I also want to thank my grandmother, Marna Cohen, who passed away before I had even begun my Ph.D., but who encouraged my creativity and intellect from before I can even remember, and who showed me what it was to live a life of strength. Finally, I want to thank my husband, Doğan Gidon, who has been the most supportive partner one could ask for in this massive endeavor; his belief in me when I found it hard to believe in myself has sustained me, and I am more grateful to him than words can adequately express.

Chapter 1

Crowd-In or Crowd-Out?: The Subnational Fiscal Response to Aid

1.1 Introduction

Why do we find large effects of individual aid programs and projects, but at best can discern only small macro-level growth effects of aid? After long debate, the literature seems to have attributed this paradox to poor data quality and the difficulty of detecting aid volumes at the national level. I argue that this conclusion to the long-standing micro-macro aid debate is unsatisfactory, and that the subnational dynamics of aid allocation are a crucial and underexplored mechanism towards resolving this puzzle.

These dynamics are particularly important when considering that aid is not monolithic, but rather given by a number of different donors, whose coordination may be imperfect. Crowd-in - whereby one donor follows another geographically in its allocations - could potentially lead to the "gap" between micro effects and macro null results. The recent availability of subnational data on aid allocation makes it newly possible to rigorously study these questions.

I explore this theory in the context of Uganda, which not only has available the necessary data, but is highly suitable given its high volume of international aid. In 2010 alone, Uganda received roughly $2 billion in bilateral and multilateral development assistance, more than a tenth of its domestic product in the same year. Although hardly the only donor, the World Bank is one of Uganda's larger and more influential donors. In the last decade, it has implemented projects targeting education, health, transportation, water quality, administative reform, and more.

In this paper, I use novel subnational datasets from Uganda to assess not only subnational aid patterns, but the response of other key players to the targeting of aid. Using an instrumental variables approach, I exploit variation in the World Bank's annual funding position to tease out the response of the government and other donors to an increase in World Bank funding, both in the overall budget and specifically in the education, health, transportation and agriculture sectors.

I conclude that, in fact, crowd-in may represent a substantial threat to national-level aid effectiveness, even in the presence of individually effective projects. The marginal recipient of World Bank aid likely benefits from the increase in aid; after an increase in World Bank aid, such a district receives weakly less funding from the central government, but considerably more funding from other donors, suggesting that some non-World Bank donors may follow the Bank's lead when thinking about how to allocate aid subnationally.

I start with a brief exploration of relevant background, then discuss the context of Uganda. I then present a simple model which lays out how effective projects and crowd-in may lead to a globally inefficient allocation. Next, I explain my analytical strategy and the data I use for this analysis. I then present my empirical results, focusing primarily on the instrumental variables analysis.

1.2 Background

In the subsequent section, I start by briefly summarizing the current state of the literature on aid and growth. I next address the literatures on fiscal responses to aid and non-governmental organizations. Finally, I briefly summarize the recent literature using subnational data.

Aid and Growth

There is a long tradition of research which tries to assess the impact of aid on economic growth. Despite evidence suggesting large and positive impacts of individual projects, there are clear and unsatisfactory disagreements in the body of evidence on macroeconomic effects of aid.

For a survey of the earlier literature, see Hansen and Tarp, 2000, which summarizes much of the pre-2000 work in the literature and concludes that aid has been shown to have small, positive effects on growth. In an influential study, however, Burnside and Dollar (2000) argued that aid worked only when the policy environment was good. The next several years of papers largely centered around critiques of underlying theory and modeling, such as the inclusion of squared aid terms by Hansen and Tarp (2001), which led to the aid-policy interaction term from Burnside and Dollar (2000) losing significance. Easterly, Levine, and Roodman (2004) largely put this strain of the literature to rest by directly extending the original regression model, and finding a null (or even slightly negative) relationship. Shortly thereafter, studies like Rajan and Subramanian (2008) argued that cross-country studies showed little robust evidence of a relationship between aid and economics growth.

More recent work suggests a cautiously optimistic relationship. Arndt, Jones, and Tarp (2016) summarizes the set of aid-growth studies published since 2008 and, via meta-analysis, suggests that aid leads to a small and hard-to-detect long-run increase in growth rates. They attribute the small long-run effects to a combination of small aid volumes, relatively moderate contribution, and inherent noise in the measurement of both aid and growth.

Fiscal Response to Aid

There is also an extensive literature on the fiscal response to aid, or how a country's public expenditure and revenue respond to aid. These works encompass the narrower literature on fungibility, which is focused more on spending. The fundamental thrust of this literature is that the broader aid-growth literature tends to ignore the role of the government in mediating the effect of aid, particularly aid which goes through the government. The fiscal response of the government – such as increases or decreases in spending or tax revenue – matter crucially for the ultimate impact of budget-supporting aid.

The bulk of this literature takes time series data and estimates the long-term relationship between aid inflows and some combination of tax revenue, investment, government spending and borrowing (Heller, 1975; Iqbal, 1997; Franco-Rodriguez and Morrissey, 1998; Franco-Rodriguez, 2000; Fagernäs and Roberts, 2004a; Fagernäs and Schurich, 2004; Fagernäs and Roberts, 2004b; Fagernäs and Roberts, 2004c; Osei, Morrissey, and Lloyd, 2005; Mavrotas and Ouattara, 2006; Lloyd et al., 2009; Martins, 2010; Mascagani and Timmis, 2014; Bwire, Lloyd, and Morrissey, 2017). A review of the foundations of the literature can be found in McGillivray and Morrissey (2001). A summary of evidence from 2005 through 2015 is available in Morrissey (2015); he concludes that aid finances national government spending, is much less fungible than generally believed, and cannot be conclusively said to increase or decrease tax revenue, findings also consistent with previous decades of work.

Geographic Fungibility

One aspect of fungibility that this literature is unable to address is the potential geographic reallocation of government or other resources in response to aid, which I will refer to as geographic fungibility. This omission is easy to understand, as the vast majority of literature makes use of either cross-national regressons (such as Remmer (2004), Gomanee et al. (2005), and others) or vector autoregressive analysis from one country over a long period of years (such as Fagernäs and Roberts (2004a), Osei, Morrissey, and Lloyd (2005), Mascagani and Timmis (2014), Bwire, Lloyd, and Morrissey (2017) and others).

However, two studies have examined some questions of subnational fungibility, specifically Walle and Mu (2007) and Wagstaff (2011). Walle and Mu (2007) looks at a World Bank project which financed road construction in Vietnam; they find that while spending was generally in the road sector, indicating limited sectoral fungibility, and there were more roads built in project areas than non-project areas, spending did still leak into non-project areas. Wagstaff (2011) approaches the problem from a different perspective, examining how potential fungibility of project funds can affect efforts to assess the impact of development programs.

Overall, the question of whether aid is viewed by governments as a complement or substitute for domestic budgetary allocations remains largely unanswered, particularly at the subnational level. The national-level literature, as summarized above, suggests that aid channeled through the government increases overall government spending. We know much less about how this additional spending is targeted, including even the basic question of whether it flows towards or away from donor-targeted regions.

Non-Governmental Organizations

I next address existing evidence on how non-governmental organizations (NGOs) target aid. Much like the fiscal response literature, existing studies have tended to focus – likely by necessity – on either the national level, or very narrowly on behavior in a single province or project.

The bulk of national-level evidence focuses on the factors that NGOs take into consideration when allocating funding (i.e Alesina and Dollar, 2000; Neumayer, 2003a; Neumayer, 2003b; Svensson, 2003; Berthélemy and Tichit, 2004; Gates and Hoeffler, 2004; Canavire et al., 2005). Much of this literature is motivated by the clustering of aid and its potential implications for development, and, correspondingly, many papers uncover patterns of giving shared by groups of donors, such as a focus on political rights by some donors Alesina and Dollar (2000) or on good economic policies (Berthélemy and Tichit, 2004). A subset of the literature explicitly concludes that NGOs cluster internationally (Isopi and Mavrotas, 2009; Koch, 2009). Across the board, the literature tend to focus on cross-sectional patterns rather than dynamic ones.

Similarly, other research suggests subnational clustering of aid funds and NGO locations (i.e. Easterly, 2002; Bielefeld and Murdoch, 2004; Barr and Fafchamps, 2006). The paper with the most data-driven approach to this question is Barr and Fafchamps (2006), which looks at client surveys of NGO effectiveness, and estimates that far too many NGOs are serving the same areas from an efficieny standpoint. Overall, few studies have estimated how donors respond to one another in an allocation sense.

Subnational Aid

Data on subnational allocations of aid funds, made available on a large scale only in the last few years, represent a step forward in our ability to understand the effects of aid.

The question of subnational allocations also requires a clarification about aid types. The aid disaggregation approach distinguishies between aid types, including project aid, program aid, technical assistance, and food aid. Mavrotas and Ouattara (2006) argue that different types of aid may induce different fiscal responses, and find supporting evidence by examining time-series evidence for Cote d'Ivoire. The bulk of the fiscal response literature, at least, is largely interested in program aid and other centrally allocated funds. However, such budget support aid is allocated by the government, and its allocations often cannot be immediately distinguished from other central government funds. As such, this paper focuses on the types of aid which are allocated subnationally by donors, chiefly project aid.

The majority of evidence to date using subnational aid tends to take the same sort of specifications used to assess national aid and apply it to subnational data. These studies nonetheless represent a significant improvement in their ability to assess impact, as evidenced by positive and substantial effects on nighttime luminosity (Dreher and Lohmann, 2015) and even in GDP (Dreher et al., 2017). Other studies have looked subnationally within a country, and found significant results in aid-on-growth (Dreher and Lohmann, 2015), or used the data to look at the impact of aid on conflict cross-nationally (Gehring, Kaplan, and Wong, 2017). In other words, subnational-level analysis seems to find more substantive impacts of aid more easily than national-level analysis.

To date, I am not aware of any other studies which have used subnational data to look at the geographic fungibility of aid or other newly-exporable geography-based questions. This paper therefore represents a contribution to both the fiscal response literature and the aid effectiveness literature, utilizing the availability of more granular datasets to look at a potential mechanism through which local effectiveness may fail to translate to national growth.

1.3 Context

The activities of the World Bank and other donors in Uganda provide an excellent context in which to utilize subnational data to test theories of crowd-in and crowd-out. Uganda receives a substantial amount of donor resources, in both direct governmental budget support and extra-governmental spending. In 2018-19, 41% of Uganda's budget came from external sources, a combination of loans and grants (MoFPED, 2018).

Since 2010, Uganda has received at least $40 per capita in external aid, known as overseas development assistance (ODA) (The World Bank, World Development Indicators, 2021a). Since 2015, Uganda has received an annual total of between 40% and 50% of the same year's budget in net ODA (The World Bank, World Development Indicators, 2021b). Back of the envelope calculations suggest that most of this is spent directly by donor agencies, rather than being routed through the government's budget.

Aid from the International Development Association (IDA), the branch of the World Bank which provides loans and credits for projects and programs intended to boost economic growth, provides a valuable example of how to think about the spending of such aid. For example, IDA funded the Uganda Support for Municipal Development Project (USMID, which provided municipalities with funding to spend on infrastructure projects of their choice, with a total investment of USD $ 150 million in 2013 and an additional USD $335 million in 2018 (International Development Association, 2018). Other projects may be more direct, providing food and nutrition-related education, or traning intended to increase the skill level of workers (World Bank, 2021a; World Bank, 2021b).

In the next section, I lay out a simple, one-period model which motivates how crowd-in could help reconcile micro-impacts with weak macro-level evidence.

1.4 A Simple Model of Donor Response

Imagine a country which consists of two regions, A and B, in a simple, one-period framework. Let region A have GDP equal to Y_a and region B have GDP equal to Y_b, such that for the whole country, GDP $= Y_a + Y_b$.

Let there be some amount of aid z which is converted to GDP through a process with diminishing marginal returns. Let the allocation that maximizes the contribution of aid to GDP $f(z_a) + g(z_b)$ be (z_a^*, z_b^*) where $z_a^* + z_b^* = z$, and the total amount of funds z is fixed.

Through this simple lens, the negative effects of donor crowd-in become clear. Define crowd-in as $z_a > z_a^*$, such that region A receives more aid than optimal. Therefore, given that $z_a + z_b = z$, $z_b < z_b^*$. Under the assumption of diminishing marginal returns, $f(z_a) + g(z_b) < f(z_a^*) + g(z_b^*)$. However, $f(z_a) > f(z_a^*)$.

This simple model demonstrates why subnational fiscal response – crowd-in – could lead to a world in which the effect of z_a in region A produces detectable effects in Y_a, but the effect of z in the country is less than one would expect, heightening the difficulty of linking total aid to country-level growth in the short run. The precise magnitudes of the effect would depend on the slopes of f and g.

Note that in this simple model, I leave unexplored the mechanisms by which crowd-in results; one can think of it as a sort of agenda-setting effect, in which smaller donors observe the behavior of a larger donor – such as the World Bank – and follow their lead in prioritizing certain regions. Other mechanisms, however, are also possible; distinguishing between them is outside the scope of this paper.

1.5 Data and Analytical Strategy

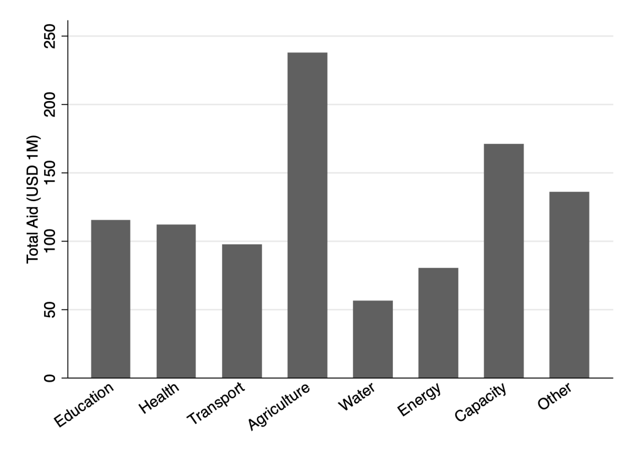
Data on World Bank projects comes from AidData (AidData, 2017). District-level locations and amount of funds comitted are available for 100% of projects implemented at the district-level from 2000 to 2014.[1] Figure 1.1 contains information on estimated total funding in millions of USD to each sector from 2010 to 2014, the period on which this analysis will focus.

The project area with the largest amount of funding from the Wrold Bank was agriculture, followed by governmental capacity; however, considerable amounts of funding went to a variety of project types, including education, health, transport, water, and energy. I confirm transfer amounts using records from Ministry of Finance, Economic Planning and Development (MOFPED)'s Office of the Auditor General.

World Bank project aid is also fairly well distributed subnationally, with most districts receiving at least some funding between 2010 and 2014. A breakdown of funding by district is in Figure 1.2, with districts indicated in darker colors receiving a higher proportion of

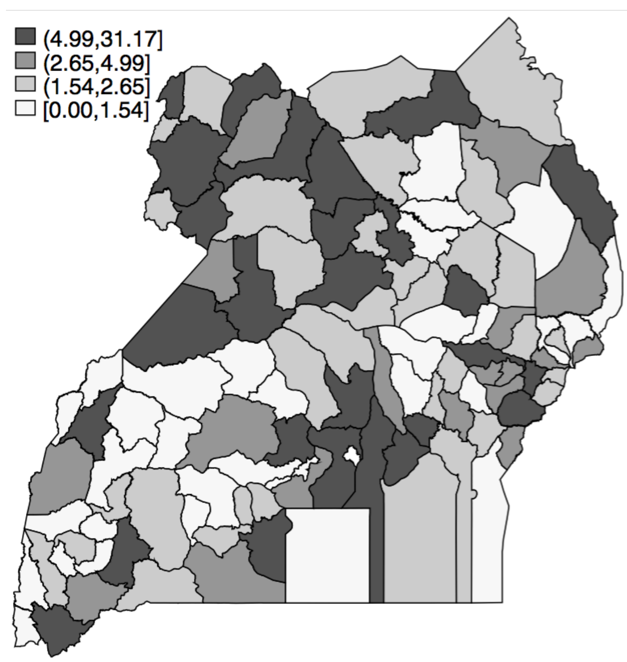
[1] These projects represent roughly 80% of the total funded projects; those with "missing" locations are either implemented in areas which do not belong neatly to any district – such as national parks – or implemented country-wide, meaning funds go to the central government.

Figure 1.1: Breakdown of Project Funding (2010 to 2014)



funding. There are some concentrations of high-volume recipients, such as in the northwest region of the country (one of its poorest, bordering South Sudan to the norh and the Democratic Republic of the Congo to the west), but there are other high-recipient areas in the central region of the country, as well as scattered throughout the south and east.

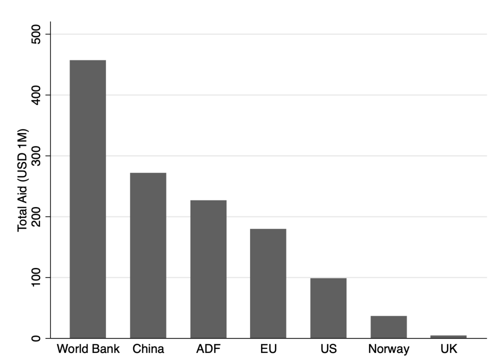
Figure 1.2: Aid Received from 2010 to 2014 (USD 1M)



AidData also provides detailed data on the activities and funding from other donors in Uganda for the same time period, 2000 to 2014 (AidData, 2016), as well as detailed data on activities and funding from China in Uganda for the same time period (Bluhm et al., 2018). Within Chinese aid, the database distinguishies between ODA-like aid, which has been categorized as developmental in nature, and contains a grant component of at least 25%, and other aid, which does not meet these requirements.

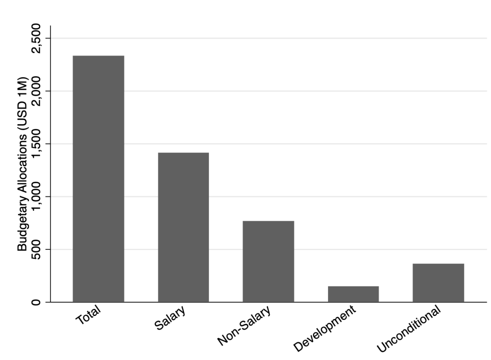
The breakdown of total volume of aid by donor can be seen in Figure 1.3. For the period of analysis, 2010 to 2014, the World Bank contributed by far the largest volume of project-based aid. The next largest donor was China, followed by the African Development Fund (ADF), the European Union (EU), the United States (US), Norway, and the United Kingdom (UK).

Figure 1.3: Aid Volume by Donor (2010 to 2014)



Other data used includes all transfers from the central government to each of Uganda's district), available for all districts from 2010 to 2015 (MoFPED, 2020). Thee transfers are the main way which district-level activities are funded, and comprise 85% or more of the average district's annual budget during the period in question. The total budgetary disbursements and a breakdown of them by type across the period of analysis is shown in Figure 1.4.

Figure 1.4: Budget Disbursement by Type (2010 to 2014)

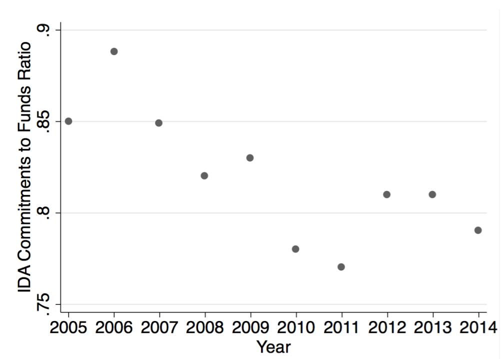


Salary, Non-Salary and Development budget items are all earmarked; for example, teacher salary funds can be spent only on teacher salaries, and not used for health worker or district administration salaries. Salary is the largest component of the budget, and includes administrative workers, teachers, health workers, and other district employees. Non-salary items focus largely on recurrent expenditures, such as medication or school lunches. Development items are one-off spending, typically on infrastructure, such as the construction of new health centers. Unconditional funds are unearmarked funds, meaning that the district controls how they are spent.

Ideally, I would regress the amount of aid from other donors and the amount of central budgetary funding on the amount of aid from the World Bank to a given district. This approach, however, is flawed, chiefly due to concerns about endogeneity. For example, sudden need in a particular district – due to natural disasters, refugee inflows, etc. – would naturally prompt funding responses from a multitude of donors and the government. Political favoritism, if it is able to redirect both governmental funding and donor funding, might also lead to spurious findings.

Because of potential endogeneity between outcomes of interest and aid, I use an instrumental variables strategy which exploits constraints in the liquidity held by the IDA. The IDA's Funding Position is calculated on an annual basis as current available funding as a percentage of what has been promised. This value is calculated for the entire IDA, rather than separately by countries, and has been made publicly available since 2008; for earlier years, I reconstruct it from data on committments and investments. Figure 1.5 displays the time series from 2005 to 2014.

Figure 1.5: IDA Funding Position



Since there is only annual variation in this instrument, I interact it with a pre-period measure of the intensity of aid received by a district. Specifically, I construct the fraction of total IDA aid received by that district in the five years prior to the first year of analysis, 2005 to 2009; I normalize that measure, so it has a mean of zero and a standard deviation of one.

The first (1.1) and second stage (1.2) equations are as follows:

$$\text{Aid Amount}_{it} = \beta \text{Prop Pre-Period Aid}_i \times \text{IDA Funding Position}_t + \psi_i + \lambda_t + \nu_{it} \quad (1.1)$$

$$Y_{it} = \beta \text{Aid Amount}_{it} + \psi_i + \lambda_t + \varepsilon_{it} \quad (1.2)$$

ψ_i is a district fixed effect, and λ_t is a year fixed effect. Standard errors are clustered at the district level. The man measure of aid is the proportion of aid in the pre-period, as below:

$$\text{Prop Pre-Period Aid}_i = \frac{\sum_{t=2005}^{2009} Aid_{it}}{\sum_{i \in I} \sum_{t=2005}^{2009} Aid_{it}}$$

Conceptually, this instrumental variables approach is analogous to a continuous differences-in-differences estimation strategy. The first stage estimates compare aid receipt in districts that received a high proportion of pre-period IDA aid to districts that received a low proportion of pre-period IDA aid, in years with high IDA liquidity to years with low liquidity. Since liquidity is calculated based on the Bank's global portfolio, it is plausibly exogenous to conditions in Ugandan districts.

This strategy is related to the approach used by Nunn and Qian (2014). The primary difference is that Nunn and Qian (2014) use contemporaneous propensity to receive aid whereas I use preperiod data, and measure propensity using proportion of amount, rather than proportion of years. The intuition is that the IDA takes into consideration which districts have been given aid in previous years when deciding where to allocate aid. This can take the form of committment effects, where districts which have been receiving aid (or a certain type of aid) get more aid, or substitution effects, in which targeted districts change over time to even out spending. I hypothesize that the second of these will be a stronger force intratemporally conditional on project timing, which is staggered across years.

In the regressions above, Y_{it} might take the form of either governmental funding, or aid from non-World Bank donors. In either case, $\beta > 0$ would represent crowd-in, where the government or other donors spend more in the district because the World Bank has funded projects there, and $\beta < 0$ would represent crowd-out, where the government or other donors avoid areas where the World Bank has funded projects.

1.6 Results

For the first stage, I display two different ways of calculating Aid Amounts due to the challenge of imputing aid amounts. For each project, I know the amount of total funding annually and the locations in which the project was active during its lifetime. I therefore impute the project-year amount either by dividing total funding evenly across locations (location), or by allocating it in accordance to the subunit's population (population).

Specifically, the specifications are as follows.

The location-based imputation divides a given project's total spending amount in a given year across all locations:

$$\text{IDA Aid Amount}_{it} = \sum_{p \in P_{it}} \frac{\text{Project Total}_p}{\text{Nr Districts}_p}$$

The population-based imputation allocates proportions of aid for a given project in a given year acros districts based on their population share:

$$\text{IDA Aid Amount}_{it} = \sum_{p \in P_{it}} \text{Project Total}_{pt} \frac{\text{Pop}_{it}}{\text{Pop}_{pt}}$$

Overall, magnitude and sign are consistent across both specifications, and either is potentially suitable for an instrumental variables approach. I proceed with column (1), which I judge to be the more realistic specification.

Table 1.1: First Stage: IDA Aid Amount (US 1M)

	(1) IDA Aid	(2) IDA Aid
Prop Aid X Funding	-1313.0***	-1667.6***
	(238.7)	(320.8)
Imputation Method	Location	Population
Aid Years	2010-2014	2010-2014
N. of obs.	555	555
F-Stat	30.24	27.02

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

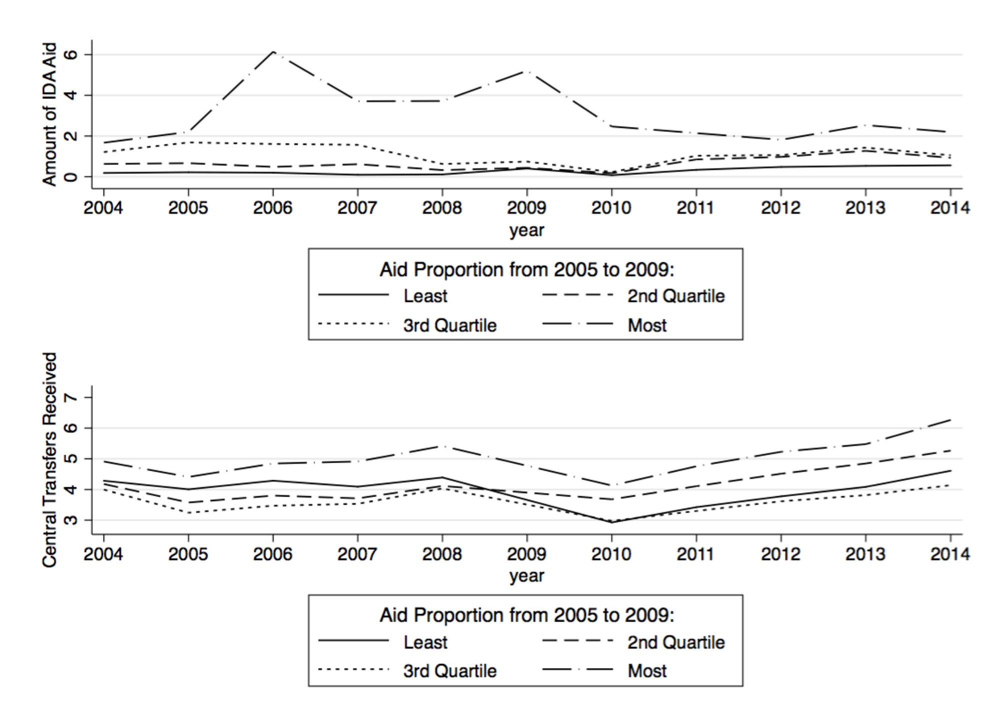
The negative coefficient suggests that in a year where IDA liquidity exogenously increases, aid receipts partially equalize between districts with high and low proportions of pre-period IDA aid. In other words, "generosity" accrues not to districts which historically received a high proportion of aid, but rather to districts which historically received a lower proportion of aid. In years where the funding position is stronger, the IDA spreads aid funding more widely; in years where the funding position is weaker, the IDA focuses on districts where it has historically spent more.

In other words, the negative coefficient suggests that the gap between historically favored and non-favored districts narrows when the funding position exogenously improves.

To interpet the magnitude, recall that the pre-period measure of IDA aid proportion is normalized. As the funding position increases, the relative gap between districts one standard deviation apart in proportion of aid received in the pre-period decreases. With a increase in the funding position of .05, from 80% to 85%, the difference betwee two districts one standard deviation apart in the proportion of pre-period aid they received would decrease by roughly $1.1 million USD.

In Figure 1.6, I present the time trends for aid receipts and central government transfers, split up by the proportion of aid received in the pre-period, from 2005 to 2009. Unlike a typical difference-in-difference, strict parallel trends is not necessary; what is key is that there are not potentially-spurious similarities in the trends over time.[2] This figure also demonstrates the broader patterns of aid allocation across districts. These patterns are fairly consistent over time, with the districts which received the most aid from 2005 to 2009 also receiving the most said from 2010 to 2014. That said, there is also evidence of shifting in aid patterns, broadly consistent with the negative coefficient in the table. For example, districts in the lowest quartile of aid distribution from 2005 to 2009, for example, overtake districts in the third quartile in average allocation during the second half of the time series.

Figure 1.6: Time Trends in Aid and Budget Transfers



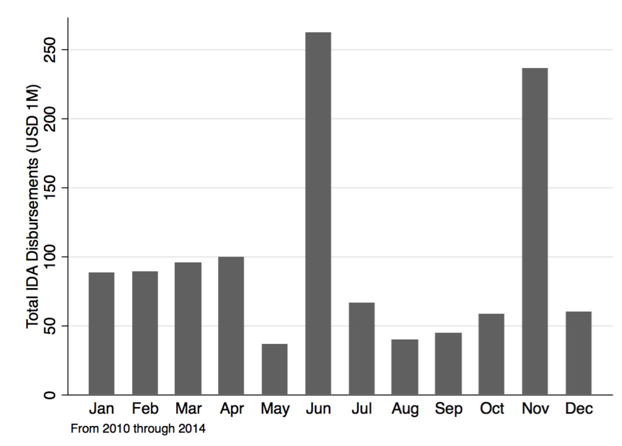
In the second stage tables that follow, each coefficient represents a separate regression of the outcome of interest on the instrumented aid amount variable. I also report two p-values: the left is uncorrected for multiple inference and the right is a q-value constructed using the Benjamini-Hochberg procedure to control for false discovery rates in each column, following Anderson (2008).

One caveat, however, is that a same year approach may not be the most accurate representation of the behavior of the relevant players. Both the IDA and the Ugandan government operate on a July-to-July schedule. It is well-known, however, that the IDA tends to release large tranches of funds in June, at the end of the fiscal year, as can be seen in Figure 1.7,

[2]The main critique that Christian and Barrett (2017) make of the Nunn and Qian (2014) approach is that wheat production and conflict only in countries with the highest propensity to receive aid both happen to both peak during the same decades, creating an unusual similarity in nonlinear trends visible in a standard difference-in-differences style graph. No analogous similarity between trends is detectable in Figure 1.6.

which shows monthly disbursements from 2010 to 2014. Correspondingly, my primary specification uses the aid amount for the previous year, instrumented for with the previous year's IDA position, in order to look at current year fiscal responses.

Figure 1.7: Monthly IDA Aid Disbursements



Specifically, for each of the outcomes of interest, I also estimate the following from 2011 to 2015:

$$Y_{it} = \beta \text{Aid Amount}_{i,t-1} + \nu_i + \lambda_t + \varepsilon_{it}$$

where all the rest is described as above; the fundamentals of the IV approach are unchanged.

As large amounts of funding are also released in November, I present both the current and lagged specifications in the tables that follow.

Government Crowd-Out

These tables focus on the governmental response to an exogenous increase in World Bank aid. Table 1.2 breaks down governmental funding by types of allocations, and Table 1.3 differentiates between sectors.[3]

Overall, the results suggest evidence of government crowd out: when a district receives more IDA funds, the amount they receive from the government decreases slightly. When the IDA allocates an additional \$1 million, there is a decrease in the same year of \$87,000 in government transfers, largely from decreased expenditure on salary. In the subsequent year, the evidence shows a decrease of \$114,000 in government transfers. In terms of sectors, the results are driven by changes in funding in education and health; there are, however, weakly insignificant increases in agricultural sector funding. The results which survive correction

[3]Appendix Table A.1 contains the OLS estimations for Table 1.2, and Appendix Table A.5 contains the reduced form estimations. For Table 1.3, the OLS estimations can be found in Appendix Table A.2 and the reduced form estimations in Appendix Table A.6.

Table 1.2: Revenue Impacts of Aid (1M USD)

	Current				Lagged			
Outcome	Coef	SE	p/q	Obs	Coef	SE	p/q	Obs
Total	-0.087	0.031	0.01/0.03	555	-0.114	0.046	0.01/0.04	555
Salary	-0.074	0.029	0.01/0.03	555	-0.128	0.044	0.00/0.02	555
Non-Salary	-0.005	0.008	0.51/0.72	555	0.006	0.013	0.64/0.81	555
Development	-0.006	0.018	0.72/0.73	555	0.002	0.008	0.78/0.81	555
Unconditional	-0.003	0.005	0.58/0.72	555	-0.001	0.003	0.80/0.81	555

Note: 2010-2014 aid sample, district and year fixed effects. Total is the sum of Salary, Non-Salary and Development; Unconditional contains both Salary and Non-Salary items. Q-values are constructed using the Benjamini-Hochberg procedure to control for false discovery rates.

Table 1.3: Revenue Impacts of Aid (1M USD)

	Current				Lagged			
Outcome	Coef	SE	p/q	Obs	Coef	SE	p/q	Obs
Education	-0.091	0.027	0.00/0.01	555	-0.110	0.045	0.02/0.08	555
Health	-0.010	0.008	0.20/0.34	555	-0.024	0.012	0.04/0.10	555
Agriculture	0.005	0.004	0.20/0.34	555	0.005	0.003	0.12/0.19	555
Transport	0.004	0.006	0.49/0.62	555	0.004	0.005	0.50/0.62	555
Water	-0.000	0.001	0.75/0.75	555	0.000	0.001	0.72/0.73	555

Note: 2010-2014 aid sample, district and year fixed effects. Q-values are constructed using the Benjamini-Hochberg procedure to control for false discovery rates.

for multiple inference are those on total budgetary spending, salary spending, and eduction spending, across both the current and lagged specifications.

These results are largely consistent with the broader literature, which views aid funds as substitutes for government spending. However, they are inconsistent with other subnational findings, which do not find evidence of governments shifting funds between geographies to balance out donor spending. Interestingly, the fact that the results are driven by education and health - which receive relatively small proportions of World Bank funds relative to sectors like agriculture - additionally suggest some potential sectoral fungibility.

Donor Crowd-In

Next, I explore the response of other donors to World Bank aid. I focus on non-Chinese donors in Table 1.4 and on aid from China in Table 1.5.[4] In both tables, I present both the current and lagged specifications, as described previously.

Table 1.4: Other Donor Impacts of Aid (1M USD)

	Current				**Lagged**			
Outcome	Coef	SE	p/q	Obs	Coef	SE	p/q	Obs
All Donors	1.387	0.431	0.00/0.01	555	0.403	0.520	0.44/0.52	444
Japan	0.410	0.396	0.30/0.35	555	0.090	0.372	0.81/0.81	444
Norway	0.015	0.012	0.20/0.29	555	0.023	0.030	0.45/0.52	444
ADF	1.316	0.170	0.00/0.00	555	0.211	0.054	0.00/0.00	444
US	-0.015	0.016	0.35/0.35	555	-0.024	0.027	0.37/0.52	444
UK	0.002	0.001	0.10/0.18	555	-0.002	0.001	0.10/0.25	444
EU	-0.115	0.057	0.05/0.11	555	-0.098	0.058	0.10/0.25	444

Note: 2010-2014 aid sample, district and year fixed effects. Q-values are constructed using the Benjamini-Hochberg procedure to control for false discovery rates.

Although it is difficult to detect the responses of individual donors, aid amounts increase across almost all donors in the current year, and increase highly significantly in the case of combined donors and the African Development Foundation (ADF). The coefficient on aid from all donors is large and significant, suggesting that an increase of $1 million from the World Bank leads to an increase of $1.4 million from other donors. In the next year, the results are consistently positive, with a significant though smaller in magnitude coefficient on funding from the ADF.

Overall, the evidence is consistent with fairly strong crowd-in from other donors; an additional $1 million USD from IDA might be as much as doubled thanks to contributions from other donors. These fundings suggest that within a given year, other donors may know something about where the IDA intends to spend funding, and also allocate funding to similar districts; this sort of crowd-in could be indicative of a follow-the-leader approach to aid, where the smaller donors take their funding allocation priorities from a larger one.

Interestingly, there are no meaningful results of an exogenous increase in World Bank funding on Chinese project spending, whether traditional or non-traditional forms of aid. Given that Chinese aid is in many senses separate from aid given by the World Bank and

[4]For Table 1.4, the OLS estimations can be found in Appendix Table A.3 and the reduced form estimations in Appendix Table A.7. For Table 1.5, the OLS estimations can be found in Appendix Table A.4 and the reduced form estimations in Appendix Table A.8.

Table 1.5: Chinese Donation Impacts of Aid (1M USD)

	Current				**Lagged**			
Outcome	Coef	SE	p/q	Obs	Coef	SE	p/q	Obs
Other Aid	0.500	0.503	0.32/0.44	555	0.276	0.278	0.32/0.93	444
All Aid	0.598	0.532	0.26/0.44	555	0.358	0.894	0.69/0.93	444
ODA-like Aid	0.098	0.126	0.44/0.44	555	0.082	0.893	0.93/0.93	444

Note: 2010-2014 aid sample, district and year fixed effects. Q-values are constructed using the Benjamini-Hochberg procedure to control for false discovery rates.

other "Western" countries, these results make sense. They suggest that while more Western donors may follow the lead of the World Bank in determining where to allocate aid, China does not do so; subsequently, there is no particular subnational fiscal response to World Bank spending in China's allocations.

1.7 Conclusion

In conclusion, despite some evidence of government crowd-out, I overall find evidence of considerable crowd-in of aid. The data suggests that when a district receives an additional \$1 million USD from the World Bank that it receives \$100,000 less from the government, but \$1.3 million USD more from other donors in the same year, suggesting a net increase of \$1.2 million USD to the district, more than double the original spending. These results are driven primarily by the African Development Fund, although Norway, Japan and the United Kingdom consistently move in tandem with te World Bank.

The magnitudes of these results suggest that crowd-in may play an important role in understanding the gap between micro-level impacts of aid, and macro-level null results. Consider a scenario in which other donors follow the World Bank's lead in deciding where to allocate project funds, but the World Bank imperfectly assesses how likely other donors are to do so when making their own allocations. In that case, even in a world in which the World Bank allocates funds strategically towards the goal of nation-wide development, imperfect predictions of crowd-in by other donors could lead to a scenario like the one outlined by the model, where project areas have exceptional results, but the country as a whole does not. Imperfect targeting of project funds by the World Bank (relative to a goal of maximizing national development) would make such results even more likely.

The results of this analysis suggest that the distribution of aid funds subnationally, and the levels of coordination between donors and the recipient government, may be very important to reconciling the micro-macro aid puzzle. Further analysis of these issues may shed additional light on the relevant dynamics.

Chapter 2

Low-Cost Tax Capacity: A Randomized Evaluation on Tax Compliance with the Uganda Revenue Authority

2.1 Introduction

The power to tax "lies at the heart of state development" (Besley and Persson, 2013). Perhaps unsurprisingly, many of the poorest countries in the world have low levels of tax collection, even relative to their own gross domestic product (GDP); existing evidence suggests that this is at least in part due to lower levels of tax compliance in such countries.

The traditional view of tax compliance argues that the decision to comply is based on rational cost-benefit calculation, with the cost of paying coming from the threat of detection in audit or other enforcement methods and from potentially high fines for non-compliance. However, in low-capacity countries, traditional enforcement is often minimal or non-existent. Although taxes are paid at a lower rate than in high-capacity countries, there is nonetheless significant tax compliance even among populations functionally without enforcement. This fact begs the question of whether compliance may be rooted in factors beyond merely enforcement, and invites consideration as to what sort of interventions may be effective in the absence of direct enforcement.

This paper focuses on tax reminder messages, which are a low cost, high feasibility intervention. Evidence suggests that such reminders work quite well to improve tax compliance in a variety of contexts: not just Western Europe and the United States, but also countries in Latin America. I join a nascent but growing body of literature testing them in even lower capacity countries – in this case, Uganda – to assess whether they work in contexts which may not be able to implement more intensive measures. The results suggest that not only is the intervention in question effective, but that it works best in the lowest capacity parts of the country, and particularly in areas where recent investments in public goods have been made. This evidence suggests that the fiscal exchange model of tax compliance, in which taxpayers view themselves as paying taxes in exchange for public goods, may play an important role in understanding tax payment behavior.

To study these questions, I undertake an experiment in conjunction with the Uganda Revenue Authority (URA) to test whether low cost tax nudges work to improve compliance among tax payers in Uganda. I focus specifically on small business owners who pay (one of) Uganda's individual taxes, which jointly comprise about 15% of total tax revenue. Although all individuals in the sample have paid taxes in some recent year, there is room for compliance improvements; roughly 50% paid any tax in the fiscal year prior to the treatment, and only 30% did so for the fiscal year in which the treatment is implemented. I also randomize message content, comparing information-focused messages, encouragement-focused messages, and enforcement-focused messages to assess the differences in efficacy between different types of messages.

I first find that messaging works, particularly by inducing individuals to pay whom otherwise would not have in the year of the experiment. In the control group, I find that roughly 5.4% of the sample pays during the short-run period post-treatment. The enforcement-focused message increases compliance by 14%, and the information-focused message increases compliance by 9%; I find no statistically discernible effect of the encouragement message. These results are largely driven by changes in compliance for the individual tax that is more focused on small business owners, called the presumptive tax.

In order to assess the general effect of messaging, I pool together all three treatments. I estimate the results on overall compliance and treatment amount using this approach, and find that sending a message resulted in a .3 percentage point increase in tax compliance, or a 6% increase relative to the control group, and raised roughly $0.20 USD (20 cents) per message. With a highly conservative estimate of .035 cents per text in cost, back of the envelope calculations suggest a 6x rate of return on average, and a 13x rate of return for the enforce treatment.

Although these are short-run results, focusing on the period from the date of the treatment (June 28, 2019) through early September of the same year, they persist strongly over that period, with no evidence of the gap narrowing over time.[1]

Next, I explore why the treatment works, and for whom. By using URA's administrative data on taxpayers and past compliance, I look at individual heterogeneity in treatment efficacy. I find that the treatment works particularly well among relatively more recently registered and lower-paying businesses, even conditional on other individual characteristics.

I next explore whether these low-cost messages are complements or substitutes with existing state capacity by building a novel, granular dataset of Uganda's public goods and looking at treatment heterogeneity accordingly. I start by geocoding the full sample of participants, nearly one hundred thousand businesses. I next build a dataset linking these

[1] I focus exclusively on the short-run, as a planned 2020 follow-up experiment has been postponed due to COVID-19 and the Uganda-wide shutdown which began in March 2020.

individuals to ten other rich administrative and secondary data sources, all geocoded down to the village or individual level.

This dataset allows me to visualize for each individual the level of public goods to which they have access, including education services, health services, police services, and court services. I can also control for economic activity in their local area, the budget of their local government, and more. I construct a broad measure of input capacity (comparable to Acemoglu, García-Jimeno, and Robinson (2015)) utilizing this data, and find that the treatment effect is generally larger in areas where existing state capacity is low. I use a model to differentiate predicted effects based on human capital and enforcement capacity, and find that the interaction effect, in this context, is driven by the human capital component of state capacity.

Rather than relying solely on input measures, I take advantage of this rich dataset to construct a variable which measures the extent to which these state inputs translate to actual tax compliance. Specifically, using the control group, I estimate tax compliance based on purely local factors using a Post-LASSO estimation procedure; I use the coefficients from this regression to construct an proximity-based prediction of tax compliance for each individual. I then take this measure of output-based state capacity and interact it with my treatment effect. I find results which are very consistent with the input-based measures. First, tax compliance is considerably higher in areas where predicted tax compliance is high. More importantly, the treatment closes more than half of the existing gap between low and high capacity areas, robust to a variety of specifications and controls.

Results from correlating the predicted tax compliance measure with Afrobarometer political opinion data rule out the alternative explanation that high capacity areas may have tax-negative attitudes. In areas with higher predicted tax compliance, individuals are more likely to report that it is not very hard to find out from the government what taxes and fees they are supposed to pay, and more likely to report that it is wrong and punishable to avoid paying the taxes they owe on their income. Similarly, they are more likely to report that the tax authority always has the right to make people pay taxes.

By combining the predicted tax compliance approach with the individual heterogeneity specifications, I am further able to establish that both predicted tax compliance and business characteristics are meaningful in determining treatment effectiveness. In other words, the treatment is more effective in lower capacity areas, but also among relatively newer businesses and relatively smaller businesses, and these characteristics are not merely proxies for one another. These effects are all consistent with a story in which low-cost treatments are able to reach potential taxpayers with whom the state otherwise struggles to engage.

I then interact predicted capacity with a measure of recent investment in public goods, which captures relatively newer local investments. I find that although predicted tax capacity remains significant, higher investment in new services leads to a considerably stronger interaction effect between predicted capacity and the treatment. In other words, although the treatment is generally more effective in all low capacity areas, it is particularly effective in low capacity areas with relatively recent invesments in public good provision.

These results highlight opportunities for low capacity states to increase revenue, and suggest potential new insights into the mechanisms of tax compliance. First, they suggest that even in low-capacity contexts, simple interventions can have dramatic results and increase revenue in a cost-effective way. Second, they suggest that low-cost treatments may potentially extend the reach of the state, working most effectively in areas and for businesses which may be ex-ante harder for the government to reach. The results suggest that these findings are driven at least in part by the extension of fiscal exchange, where taxpayers in areas with recent investments in services are most responsive to attempts by the government to induce tax compliance.

Next, I will briefly discuss the literatures to which this paper contributes, followed by the relevant background in Uganda's tax system, including overall tax capacity, the specific taxes studied here, and enforcement and tax-related attitudes generally. I will then describe the experiment which was conducted, and estimate average treatment effects across the sample. I will build a simple model to help us think intuitively about the expected results of forthcoming heterogeneity analysis, then discuss in some detail the geocoding process and its results. Finally, I will undertake heterogeneity analysis across measured and predicted state capacity.

2.2 Related Literature

This paper contributes primarily to three different literatures.

First, it adds experimental evidence to the literature on state capacity and taxation. There is an extensive literature which explores the evolution of tax systems over time, and between high and low income countries.[2] In general, high-income countries collect much more tax revenue relative to GDP than low-income ones (Besley and Persson, 2014). Increases in tax revenue are accompanied by shifts from collecting revenue via trade taxes and excises to collecting revenue via labor income and value-added taxes. Such cross-country comparisons also indicate that that tax rates do not uniquely determine tax revenue; many low-income and middle-income countries have similar statutory rates to high income countries, but still take in much less tax revenue even as a ratio to GDP, and particularly much less income tax revenue. In other words, in addition to tax systems, compliance also matters.[3]

Evidence on subnational variation in capacity is more limited, focusing on subnational variation in the development of state capacity over time (e.g., Acemoglu, Reed, and Robinson, 2014) or assessing strategic complementarities between subnational and central state capacity (e.g., Acemoglu, García-Jimeno, and Robinson, 2015). Instead, this paper explores the heterogeneity of a tax-focused intervention across the dimension of subnational state capacity, examining both causes and consequences of within-state variation of this crucial dimension of statehood.

[2]See Besley and Persson (2013) for a comprehensive review of the state of this literature.

[3]See Pomeranz and Vila-Belda (2019) for a review of experimental evidence on state capacity, broadly defined; this review also overlaps substantially with the literature on tax compliance interventions.

This paper also contributes to literatures which explore the behavior underlying tax compliance, and specifically those which test the efficacy of interventions intended at improving tax compliance. This literature dates back a number of years, and is particularly well known for experiments sending letters or equivalent to taxpayers in order to encourage compliance.

The theoretical underpinnings of this literature focus on the broad question of why individuals pay taxes. Theories along these lines can be subdivided into five distinct but interconnected categories: (1) economic deterrence; (2) fiscal exchange; (3) social influences; (4) comparative treatment; and (5) political legitimacy.

Economic deterrence is the most well-known of these theories, suggesting individuals decide on tax compliance based on rational cost-benefit calculations; the foundational model of this type is Allingham and Sandmo (1972), and the state's levers, accordingly, are likelihood of evasion detection and the magnitude of the punishment for evasion. The fiscal exchange theory, on the other hand, suggests that taxation and the provision of public goods are viewed as a contractual relationship between taxpayers and the government; under this theory, taxpayers may comply without direct coercion (Bodea and LeBas, 2013). Social influence theory suggests tax compliance behavior is set by social norms held in an individual's reference group, which it has been argued applies to taxation as elsewhere (Snavely, 1990). Comparative treatment suggests that an individual's evaluation of how they and their identity group are perceived by the government affects tax compliance, with favored groups more likely to comply (McKerchar and Evans, 2009). Political legitimacy suggests that the more individuals trust their government, the likelier they are to comply with tax burdens (e.g. Kirchler, Hoelzl, and Wahl, 2008).[4]

These theories are not necessarily believed to be mutually exclusive, and there is no reason to posit that proof of one is equivalent to a rejection of any of the others. In practice, many tax letter RCTs are framed as assessing support of such theories, with different messages designed to target various levers: audit-focused letters, for example, can be thought of as targeting economic deterrence (e.g., Slemrod, Blumenthal, and Christian, 2001; Kleven et al., 2011; Pomeranz, Marshall, and Castellon, 2014). Letters which include a component of information about similar businesses can be viewed as assessing the effects of social influence theory, and letters which include information about the value of services are often perceived as affecting fiscal exchange (e.g., Blumenthal, Christian, and Slemrod, 2001; Ariel, 2012; Castro and Scartascini, 2013; Ortega, Ronconi, and Sanguinetti, 2016). One can also think of reminders as reducing the hassle costs of paying, as evidence suggests in other contexts that taxpayers forego tax savings due to the effort of complying (Benzarti, 2015).

In general, such letters, particularly when focused on deterrence, seem to work fairly well in inducing compliance among populations where non-compliance is feasible, though precise results for various messages vary based on context and content both. All such experiments, however, also serve another purpose: they are very low-cost interventions, particularly when

[4] In practice, the term tax morale is often used to mean some combination of theories of compliance other than economic deterrence; a recent review by Luttmer and Singhal (2014) unpacks the main mechanisms of this term.

text- or email-based, and they require little more than a knowledge of whom potential taxpayers are, and a means of contacting them.

Correspondingly, in recent years this literature has also been extended to look at the developing world. Experiments in Chile, Colombia and Guatemala, for example, show that such low-cost interventions can work outside of the developed world, particularly punishment-based messages (Pomeranz, 2015; Kettle et al., 2016; Ortega, Ronconi, and Sanguinetti, 2016). More intensive interventions, such as going door to door for tax collection, have also been shown to work even in some very low capacity contexts (Weigel, 2020). As well, projects which assess more substantive reforms, such as changing the incentive structures of tax collectors or introducing electronic billing machines, also find the potential for high returns, even in lower capacity contexts (Eissa and Zeitlin, 2014; Khan, Khwaja, and Olken, 2016; Khan, Khwaja, and Olken, 2019).[5] Such interventions, however, are usually either costly, require existing levels of bureaucracy for implementation and monitoring, or potentially both, meaning they may be infeasible to implement in contexts where resources and infrastructure are sufficiently scarce.

The tax enforcement environment in a country like Uganda differs meaningfully not just from Western Europe and the United States, but even from Latin America. As such, though in recent years more experiments have been conducted in low-capacity environments, there is still a dearth of published evidence in such contexts on the levers of tax compliance and the efficacy of low-cost solutions.[6] One published exception is Mascagni, Nell, and Monkam (2017), which finds that low-cost messaging, including tax morale and deterrence-focused text messages, worked well in Rwanda for a business profits tax.[7] This experiment furthers that branch of the literature, and specifically tries to understand how Uganda's state capacity or lack thereof intersects with treatment effectiveness, as well as to provide some potential insight on the relevant theoretical mechanisms through both message differentiation and heterogeneity analysis.

Last, the use of geocoded administrative, secondary and satellite-based datasets adds this paper to a growing list of works which supplement randomized control trials and other causal estimation techniques with already existing data (see Donaldson and Storeygard, 2016; Bouguen et al., 2018). The use of this sort of data allows me to exploit the nation-wide sample that I have in my experiment without needing to rely exclusively on URA's administrative data or conduct any kind of expensive nation-wide household survey.

[5]For two excellent reviews of this literature, see Slemrod (2018) overall and Mascagni (2018) for a look at low-income countries in particular.

[6]There is a small but relevant literature which uses opinion surveys to deconstruct tax-related attitudes in sub-Saharan Africa, though the results are relatively inconclusive e.g., Ali, Fjeldstad, and Sjursen, 2014.

[7]Another related study of note in Nigeria, currently unpublished was part of EGAP II: Taxation Metaketa (Gottlieb, LeBas, and Obikili, 2018).

2.3 Background

Even within the context of relatively low-capacity countries, Uganda has low tax capacity. In 2018, one of the latest years for which the IMF has data, Uganda collected roughly 14% of its GDP in tax revenue, as compared to 22% in Rwanda, and 17% in Kenya (International Monetary Fund, 2020).[8] Uganda falls relatively more behind when one considers income tax rather than all sources of taxation. In 2018, Uganda collected roughly 2.3% of GDP (16% of total tax revenue) in income tax revenue payable by individuals.[9] Rwanda collected roughly 4.4% of GDP (20% of tax revenue) in income tax revenue payable by individuals; Kenya collected roughly 4.9% of GDP (29% of tax revenue) in income tax revenue payable by individuals.[10]

Uganda has two individual-payable income taxes, the presumptive tax and the personal income tax (PIT). Businesses with total sales volumes between $2,600 USD (UGX 10 million) and $40,000 USD (UGX 150 million) annually are required to pay the presumptive tax, a step-function tax with amounts increasing in revenue.[11] Payments for businesses with sales volumes less than $13,000 USD (UGX 50 million) annually vary based on location and industry, with more remote businesses paying less at each step.

For these sorts of small businesses, no business tax registration is required; rather, an individual pays using their personal taxpayer identification number (TIN), and does not need to declare to the government that they have a business in order to do so. In practice, relatively few taxpayers file as well as pay, although doing both is mandated by law.[12]

In general, one should think here of small businesses with somewhere between no employees and perhaps as many as ten or so (though not all may necessarily be full-time). These businesses operate in sectors including general trade, carpentry, metal workshops, motor vehicle repair, hair and beauty salons, restaurants and bars, drug shops, and even clinics. Examples might include a small shop selling grocery items in a rural town, or a larger private garage in a more urban setting. In general, microenterprises – such as a stall selling homemade food at night, or a woman selling charcoal from a blanket on the roadside – would generally fall beneath the minimum turnover threshold, and therefore be tax-exempt.

The personal income tax applies to all individuals who generate income via self-employment

[8] In the same year, by way of reference, the U.K. collected 36% of its GDP in tax revenue, and Australia collected 26% of its GDP in tax revenue.

[9] Standard measures of income tax include income tax payable by individuals and income tax payable by corporations and other enterprises; I focus here on income tax payable by individuals, but the comparisons are not dissimilar when including corporate-payable income taxes.

[10] By way of comparison, the U.K. collected 9.2% of GDP (24% of tax revenue) in income tax payable by individuals; Australia collected 11.2% of GDP (43% of tax revenue) in income tax payable by individuals.

[11] Larger businesses have their own sets of taxes and requirements, including the Pay-As-You-Earn Tax for employees (PAYE) and Uganda's Value Added Tax (VAT). For more information, see Uganda's Taxation Handbook.

[12] In practice, a business may pay without filing by entering their self-calculated tax burden on the URA website, printing out a bank slip, and taking this slip to a bank to make payment. For more information, see Uganda's Small Business Taxpayers guide.

above a certain threshold. Individuals who earn more than $750 USD (UGX 2.82 million) annually are required to pay this tax, which has increasing marginal rates capping out at 30%, with no upper limit. Note that this tax applies exclusively to self-employment income; employed individuals do pay income tax, but these taxes are paid by corporations or other employers, not by the individual him or herself. The proportion of individuals required to pay the personal income tax is relatively small; most, though not all, have their own business, as per the law individuals who earn money as a consultant are intended to be covered under business-focused taxation.

Note that the structure of these two taxes means that their populations may overlap in theory. In practice, though some individuals pay both taxes, many of those who pay the presumptive tax may fall below the lower threshold for the personal income tax, and many of those who pay the personal income tax may earn income from sources other than a small business.[13]

In practice, direct enforcement actions for these taxes by the Uganda Revenue Authority are limited. Although paying taxes is required in Uganda, and there are legally mandated penalties for evading taxes or making late payments, URA reserves its audits (generally) for larger taxpayers. Lower level governmental organizations, such as the Kampala Capital City Authority (KCCA) and district goverments across Uganda, do sometimes take enforcement actions against small business owners to encourage payment of local taxes and fees. If enforcement actions take place, whether mandated by the central tax authority or local governments, enforcement is done in practice by the Ugandan police, at least among this set of (smaller) businesses.[14]

URA's administrative data reveals evidence of potential extensive and intensive evasion. Although TIN registration has increased widely across Uganda over the last five or six years, many individuals who report having a business at the time of registration never pay either of Uganda's individual taxes. Payments are inconsistent, with some individuals paying the presumptive or personal income tax in a given year, skipping the next year, and paying again in the following year. It is, naturally, difficult to distinguish in administrative data between an individual whose business has failed and one who has evaded taxes; however, the volume of non-compliance is indicative.

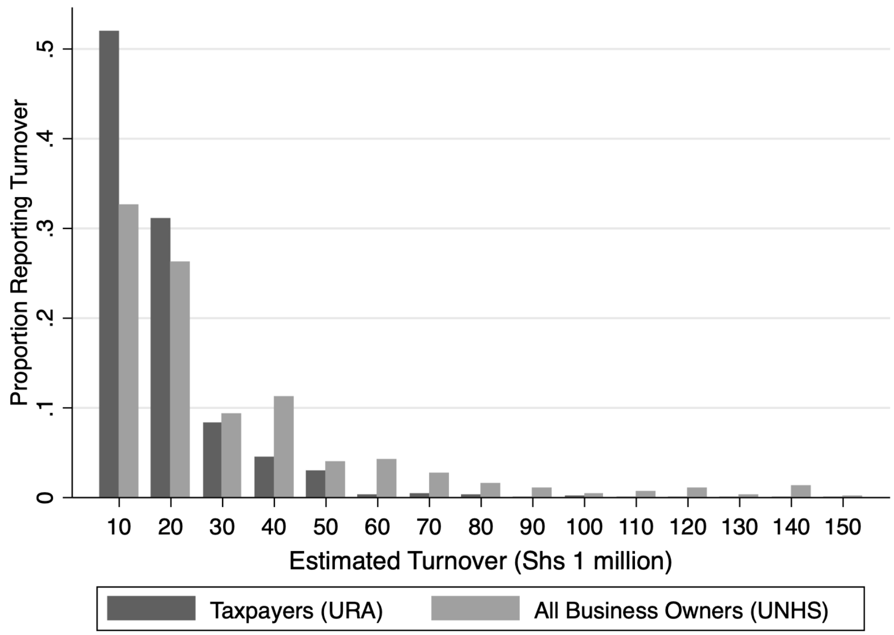
Comparison of URA's administrative data with other sources also yields suggestive evidence of pervasive non-compliance. For the subset of individuals who reported their annual sales, it is possible to compare the histogram of their reporting with the Uganda National Household Survey (UNHS) which asks a similar question of small business owners.

In Figure 2.1, I graph the relative percentage in each ten million UGX bin between zero and one hundred and fifty million UGX. This comparison shows that a high fraction of business owners reporting to URA - roughly 50% - report themselves at the minimum

[13]It is easy to see how this might be the case; a small business with an annual sales volume of $2,600 very plausibly might not result in $750 worth of income annually for its owner.

[14]During qualitative interviews with tax officials in lower level governments, I was told that in practice it often costs district governments more than they receive in payment to send the police to enforce tax compliance, due to the fees they must pay to the police.

Figure 2.1: Intensive Margin Misreporting



threshold for eligibility, at or below ten million UGX; however, only 30% of business owners report themselves to UNHS as at this threshold. There is also a significant percentage of business owners who, according to the UNHS distribution, should be reporting turnover in the realm of UGX fifty million to UGX one hundred and fifty million; however, there are almost no business owners who do so in the URA data.[15]

Given these facts, it is interesting as well to briefly explore the opinions of Ugandans regarding taxation. The Afrobarometer surveys are a pan-African series of national public attitude surveys on democracy, governance and society. They have conducted a number of rounds in Uganda, including a general opinion survey in 2014 (Afrobarometer Data, 2014). In this survey, roughly 73% of Ugandans agree that the tax authority has the right to tax citizens, suggestive of relatively high political legitimacy. Roughly 85% of Ugandans say it is wrong to evade taxes, split between approximately 46% who say it is wrong and punishable to evade taxes, and approximately 39% who say it is wrong but understandable to evade taxes. Interestingly, just over 47% of Ugandans say that it is very hard to know their tax burden.

As part of this broader study, I conducted a phone survey of roughly 5,100 study participants, who comprised a representative sample of Kampala-based study participants.[16] Their

[15]This missing mass can be explained either by underreporting of profits during filing by larger businesses, or if the likelihood of a business owner failing to file is increasing in business size; either way, these patterns are plausibly indicative of non-compliance.

[16]The phone numbers used for this opinion survey were the same as those used in the experiment; although

responses reveal similar attitudes.

In general, phone survey respondents reported that roughly half of other Ugandans who are meant to pay either of these individual taxes do so. When asked for the main reason why a Ugandan might not wish to pay their taxes, approximately 16% of Ugandans reported that they did not believe the government should tax; this is quantitatively close to the magnitude of Afrobarometer respondents who believe tax evasion is not wrong. The primary reason given for non-compliance was overtaxation, as per 37% of respondents; only 8% of respondents said that knowledge was a barrier for Ugandans.[17] Notably, however, only 17% of respondents could correctly answer a question on the minimum tax obligation at which an individual is required to pay the presumptive tax; 28% answered too low a figure, and 55% answered too high a figure.[18]

Finally, roughly half of respondents agreed or strongly agreed with a statement that the government of Uganda uses the money paid by taxpayers in a way which helps Uganda's citizens, which could be interpreted as limited support for the fiscal exchange model of tax compliance.

In the next section, I discuss the experimental design of the randomized control trial and sample balance, and present average treatment effects on tax payment incidence and amount, both disaggregated by treatment and pooled together. I also present individual heterogeneity results using URA's administrative datasets.

2.4 Experimental Evidence on Tax Compliance

Experimental design

In conjunction with the Uganda Revenue Authority (URA), I designed an experiment to test the efficacy of tax payment reminders in Uganda. Our sample consisted of the 98,137 still active taxpayers who had paid any of Uganda's individual payment taxes (the presumptive tax or the personal income tax) in any financial year since 2014-2015; regression results will be estimated on the 85,000 of these who could be geocoded and matched to public good data, as will be discussed in detail in Section 2.5.[19]

These individuals were subdivided into four groups, a control group and three treatment groups, each of which was sent a different message. The treatment groups received a message on June 28, two days prior to the end of the fiscal year on June 30, June 30 is also the deadline to file and pay individual taxes, though late payments are routinely accepted.

The breakdown of treatment groups was as follows:

not all of those contacted answered the survey, we were able to reach roughly 90% of those targeted.

[17] As this was a sample who had previously paid at least one individual tax, this number is likely an underestimate for the population as a whole

[18] This survey question was not incentivized in any way; as such, this is likely a lower bound. The true answer was 10 million UGX; if I look at those who answered between 5 million UGX and 20 million UGX, a very generous bound, I still find that only 50% of the sample gave a correct answer.

[19] Results are robust to estimation on the full sample (unreported).

- **Control**: No text sent
- **Inform**: Dear esteemed client, please file your income tax return and pay the tax due by 30th June 2019. URA
- **Encourage**: Dear esteemed client, by paying your taxes you make it possible to educate our children, fund our healthcare, and keep us safe. URA
- **Enforce**: Dear esteemed client, file your income tax return and pay the tax to avoid unnecessary payment of interest, penalties, and possible enforcement action like closure of business. URA

Group means and balance are presented along a variety of dimensions based on the full suite of administrative data available from URA, as can be seen in Table 2.1.

Table 2.1: Balance Table

Variable	Control	Inform	Encourage	Enforce	p (All)	p (T v C)
Taxpayer Age (years)	42.5	42.6	42.4	42.5	0.218	0.831
Years Since Registration	5.55	5.53	5.55	5.55	0.887	0.942
Taxpayer is Male (d)	.704	.708	.71	.705	0.532	0.318
Located in Kampala (d)	.489	.477	.484	.482	0.386	0.043
Registered Business (d)	.702	.695	.701	.699	0.360	0.330
Paid Tax 2017-18 (d)	.497	.502	.492	.497	0.086	0.919
Unpaid by June 28 (d)	.699	.701	.706	.703	0.497	0.188
Last Payment (1k USD)	.202	.205	.227	.251	0.408	0.347
N=84,955						.

The first four columns represent the group-specific mean for the variables listed on the lefthand side, based on the URA administrative database. The sample size is based on the Google Maps method of geocoding. p(All) contains the p-value from a test of joint equality for coefficients on inform, encourage and enforce; p(T v C) contains the p-value for the coefficient on dummy variable indicating the respondent was assigned to any treatment group.

The variables in registration and past payment data include age, with the average participant around 40 years of age; gender, with 70% of participants male; and location, with just under half of the sample located in or around Kampala, the center of the Ugandan economy. Roughly half of the sample paid any individual tax in the 2017-18 financial year.[20] As of July 28, only about 30% of the sample had paid either individual tax in 2019, and the average tax payment in the most recent payment year prior to the 2018-19 fiscal year was

[20]The same is true for the 2016-17 financial year, although the sample is slightly smaller, as some individuals in the same did not register until 2016-17.

roughly \$230 USD, including both the presumptive and the personal income tax. The data suggests that a little over a third of respondents paid the minimum of the presumptive tax during their last payment.

I report the p-value for the joint equality of all coefficients in the first column (p (All)), and the p-value for a comparison of the pooled treatment compared to control in the second (p (T v C)); I show both because I will present results with both disaggregated and aggregated treatments. In general, the results are balanced; however, there are two exceptions worth noting. First, in the disaggregated treatment comparison, the coefficient on whether individuals paid tax in 2017-18 is significant at the 10% level, though the magnitude is fairly small. Accordingly, I include this covariate as a control in my preferred specification. Second, the coefficient on whether an individual was located in Kampala was significant in the test of pooled treatment against control at the 5% level, although just barely, with control participants very slightly more likely to live in Kampala than treated participants. To deal with this, my preferred specification controls for district fixed effects, which include a control for whether or not an individual lives in Kampala.

Experimental results

First, I examine the average effects of the treatment on the study population on tax payment incidence in Table 2.2, and on tax payment amount in Table 2.3.

In Panel A of each table, I estimate the following regression:

$$Y_{id} = \alpha + \beta_1 \text{Inform}_i + \beta_2 \text{Encourage}_i + \beta_3 \text{Enforce}_i + X_i + \delta_d + \epsilon_{id}$$

for individual i in district d, where Y_i measures either tax incidence or tax payment amounts. I estimate this regression with robust standard errors, where X_i includes randomization strata fixed effects and other individual controls in some specifications, and δ_d are district fixed effects.

In Panel B of each table, I estimate the following regression:

$$Y_{id} = \alpha + \beta \text{Any Treatment}_i + X_i + \delta_d + \epsilon_{id}$$

for individual i in district d, where Y_i measures either tax incidence or tax payment amounts. I estimate this regression with robust standard errors, where X_i includes randomization strata fixed effects and other individual controls in some specifications, and δ_d are district fixed effects. In this specification, I pool together all treatments, and estimate their joint effect.

In both tables, results are disaggregated by tax type, with outcomes (in order) indicating payment for any tax, for the presumptive tax, and for the personal income tax (PIT). Note that since it is not known ex-ante which respondents are meant to pay which tax (and in some cases, respondents may need to pay both taxes), all regressions are estimated on the full sample, regardless of which outcome is specified.

Table 2.2: Treatment Effects

Panel A. Separate Treatments						
	Paid Tax		Presumptive		PIT	
Inform	0.481**	0.429*	0.277*	0.250*	0.214	0.188
	(0.032)	(0.051)	(0.054)	(0.078)	(0.228)	(0.282)
Encourage	-0.118	-0.094	-0.087	-0.073	-0.031	-0.021
	(0.588)	(0.661)	(0.529)	(0.594)	(0.857)	(0.904)
Enforce	0.736***	0.727***	0.725***	0.729***	0.049	0.034
	(0.001)	(0.001)	(0.000)	(0.000)	(0.781)	(0.843)
Inform vs. Enforce	0.270	0.189	0.004	0.002	0.354	0.382
Inform vs. Encourage	0.007	0.017	0.011	0.022	0.166	0.232
Encourage vs. Enforce	0.000	0.000	0.000	0.000	0.647	0.750
Panel B. Pooled Treatment						
	Paid Tax		Presumptive		PIT	
Any Treatment	0.366**	0.353**	0.304***	0.302***	0.077	0.067
	(0.043)	(0.047)	(0.008)	(0.008)	(0.591)	(0.635)
Individual Controls	No	Yes	No	Yes	No	Yes
Control Mean	5.414	5.414	2.100	2.100	3.376	3.376
Observations	84955	84955	84955	84955	84955	84955

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Outcomes consist of indicators for whether any tax, the presumptive, and the personal income tax, respectively, were paid between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. All regressions include strata and district fixed effects and are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table 2.2 shows that the treatments worked well in improving compliance. These results indicate that from June 28, the date of the message, to September 1, there was a statistically significant 8.7% increase (.48 percentage points) in tax compliance due to the inform message and a statistically significant 13.7% increase (.74 percentage points) in tax compliance due to the enforce message, both relative to a control mean of 5.4% of the sample paying their taxes during this time period; note that we cannot distinguish between the effects of the two treatments when looking at overall tax compliance. The encourage treatment does not seem to have meaningfully affected the compliance rate.

These results were driven largely by the presumptive tax, particularly the results for the enforce treatment. Given the specific content of the message, which focuses on punishments (including the closure of business), this difference between the treatments seems to make intuitive sense. I also pool together all three treatments and estimate their combined effect on the sample as a whole; the results suggest an 6% average increase in tax compliance.

Table 2.3: Treatment Effects

Panel A. Separate Treatments						
	Tax Amount (USD)		Presumptive		PIT	
Inform	0.185	0.168	0.255*	0.236	0.114	0.104
	(0.277)	(0.318)	(0.089)	(0.113)	(0.488)	(0.524)
Encourage	0.038	0.052	-0.132	-0.116	0.075	0.088
	(0.824)	(0.758)	(0.321)	(0.381)	(0.650)	(0.587)
Enforce	0.150	0.139	0.464***	0.473***	0.006	0.000
	(0.385)	(0.414)	(0.002)	(0.001)	(0.969)	(0.999)
Inform vs. Enforce	0.837	0.865	0.177	0.123	0.511	0.524
Inform vs. Encourage	0.387	0.490	0.006	0.012	0.813	0.927
Encourage vs. Enforce	0.515	0.608	0.000	0.000	0.676	0.587
Panel B. Pooled Treatment						
	Tax Amount (USD)		Presumptive		PIT	
Any Treatment	0.124	0.120	0.195*	0.197*	0.065	0.064
	(0.371)	(0.383)	(0.095)	(0.089)	(0.627)	(0.627)
Individual Controls	No	Yes	No	Yes	No	Yes
Control Mean	2.948	2.948	1.546	1.546	2.767	2.767
Observations	84955	84955	84955	84955	84955	84955

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Outcomes consist of the amount paid for any tax, the presumptive, and the personal income tax, respectively, between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. All regressions include strata and district fixed effects and are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table 2.3 shows the efficacy of the treatments in increasing tax payments. I find that the enforce treatment was effective in increasing the amount of tax paid, particularly for the presumptive tax (although note that I cannot distinguish statistically between the results of the inform and enforce treatments). When I pool together all treatments, I find that the average presumptive tax payment increased by a statistically significant \$0.20 USD (20 cents) across the sample, relative to an average per-person payment of \$1.55 in the control group. With 63,700 treatment messages sent across the sample, this is equivalent to just shy of

$12,750 USD in revenue raised for the Uganda Revenue Authority. Assuming a conservative per-text cost of .035 cents (130 UGX, the maximum charged by any network per text, with no bulk or government discount), back of the envelope calculations suggest over a 6x rate of return on the pooled treatment.[21]

I find that the payment amount results seem to be largely driven by increases in payment incidence; there is no evidence to suggest that individuals are paying more as a result of the treatment.[22]

Individual heterogeneity

In Table 2.4, I use administrative data from URA to explore which types of respondents were more responsive to the treatment. Specifically, I look at heterogeneity along the years since TIN registration, and the size of most recent tax payment. Note that from this point forward, I pool together all three treatments and estimate a combined treatment effect across them. For this table, as for all subsequent tables, I report disaggregated treatment effects by treatment in Appendix B.2.[23]

I find that the effects of the treatment are (weakly) decreasing with years since registration. These findings suggest that, generally, the effect of the treatment was greater among relatively newer businesses in the pooled tax compliance sample. As well, I find that the effect of the treatment decreases with the size of the latest tax payment, suggesting that the treatment was also relatively more effective among lower-paying businesses.

On the whole, these results are indicative of the treatment being most effective among newer and lower-paying businesses. The results on businesses paying less are most difficult to interpret; recall that a business paying less might be a result of the business being smaller, the business evading more, or the business being located in a more rural area (as the presumptive tax has a lower tax burden for more remote businesses). Under this specification, as such, it is difficult to determine whether it is the size of the business or where it is located which drives the results.

It is customary in this strand of the literature to interpret the differences between the treatments as evidence for one or another theories of tax compliance. In this case, that would lead me to conclude that the treatment works by decreasing the hassle costs of compliance (as in the inform treatment) or else by changing expectations or salience of punishment (as in the enforce treatment). The lack of results from the encourage treatment would suggest little in the way of support for the fiscal exchange hypothesis. Such interpretations would – and should – be caveated with a mention that other phrasings may yield different results.

In addition to this analysis, however, in the following parts of the paper I use an extensive, granular dataset to examine heterogeneity effects by local levels of state capacity. Doing

[21]Note that looking exclusively at the enforce treatment suggests a rate of return of roughly 13x; looking at the pooled inform and enforce treatments suggests a rate of return of roughly 10x.

[22]See Appendix Table A.1 for the relevant results; there is no indication of a treatment effect of repayment amount among those who paid either or each tax.

[23]Results for Table 2.4 disaggregated by treatment can be found in Appendix Table B.4.

Table 2.4: Individual Heterogeneity

	Paid Tax		
	(1)	(2)	(3)
Treatment	0.409**	0.358**	0.406**
	(0.022)	(0.044)	(0.023)
Last Tax Payment (1k USD)	0.413***		0.377***
	(0.004)		(0.008)
Treatment x Last Tax Payment (1k USD)	-0.275*		-0.258*
	(0.081)		(0.100)
Years Since Registration		0.330***	0.304***
		(0.000)	(0.000)
Treatment x Years Since Registration		-0.100*	-0.089
		(0.069)	(0.105)
Observations	84955	84955	84955
Individual Controls	Yes	Yes	Yes
District FE	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. All regressions include strata and district fixed effects as well as individual controls and are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

so allows me to utilize another way to examine the potential channels and mechanisms underlying the study results; specifically, I explore whether treatment efficacy varies with state capacity, which one would likely not expect under economic deterrence or hassle costs. In the next section, I discuss the process by which I geocode the study participants and link them geographically to a number of other rich datasets, allowing me to measure state capacity at the business level; in subsequent sections.

2.5 Geocoding and Dataset Creation

In order to move beyond the heterogeneity analysis made possible using the Uganda Revenue Authority's administrative data and estimate versions of the relationships in the model above, I undertook an extensive geocoding process with the goal of linking as many businesses as possible in the sample to the public goods (health, education, legal and safety services) provided in their vicinity, utilizing a variety of different data sources and approaches to allay potential concerns about data quality.

The Ugandan central government provided roughly 94% of an average district or municipality's annual budget in the 2018-19 fiscal year. Close to 80% of this central government funding is provided via conditional or other government transfers, with other 14% coming in the form of discretionary central government transfers. This funding pays for salaries, facilities maintenance, and other such factors which contribute to public good provision, with the central government often constraining how such funds can be spent. Local governments do have discretion over some aspects of allocation decisions, though generally not the bulk of the budget.[24] Given this funding structure, I argue that these local public goods may be construed as a measure of national state capacity.

Business geocoding

I started with physical addresses provided to URA by individuals at the time of their TIN. In the vast majority of cases, although not always, such addresses contain detailed location information including village name, subcounty name, and district name. In many cases, street names were also included; in some cases, street numbers or building names were available. I used two different methods, Google Maps API (via Geocode by Awesome Table) and ArcGIS, to transform these physical addresses into latitudes and longitudes via a search of the underlying address locator. The only restriction I placed on either search was to require that the addresses found be located in Uganda. Each program was able to return a physical address within Uganda for 99% of businesses.

Based on these addresses, I compile a large dataset linking each business to other datasets, the full list of which can be seen in the Secondary Data Sources table.

For education, health, police posts and local courts, I construct a measure averaging the inputs into public goods provided in a five kilometer radius j around each business i; more specific details on measures of inputs and the construction of the average measure are discussed below.

The decision to use a five kilometer radius, rather than focusing on the nearest health center or school, is in part due to concerns over data quality issues; there is good reason to believe these geocoding processes are less accurate in a country like Uganda where map quality is poor, and addresses are much less standardized, including both transliteration errors and a lack of street names and numbers. As well, the "nearest" health center or school may not be the one which the business owner uses, particularly in an urban or peri-urban setting, suggesting that averaging what is available in the business's vicinity may represent more realistically the public goods to which they have access. Cross-geocoding

[24] Teachers demonstrate how such decisions are allocated. Teacher salaries are paid by designated central government funds; the central government regulates the minimal number of teachers per school and per student. The final decision, however, about how many and which teachers to place at a given school is made by the local government.

Table 2.5: Secondary Data Sources

Data Source	Citation
Uganda Ministry of Education	MoES, 2016
Uganda Ministry of Health	MoH, 2016
Uganda Justice, Law and Order Sector	JLoS (2013)
GeoQuery Database	Goodman et al. (2019)
VIIRS nighttime lights	Elvidge et al. (2017)
Accessibility to cities	Weiss et al. (2018)
Population density	CIESIN (2016)
Health Spending	AidData (2018)
Uganda Ministry of Finance	MoFPED (2020)
Uganda National Household Survey	UBOS (2017)
Afrobaromoter	Afrobarometer Data (2014), BenYishay et al. (2017)

method robustness suggests that overall results are not driven by geocoding choices; see Section 2.5 for more detail.

Index construction

Appendix Figure A.1 provides a visual overview of the geocoding process. Each business i is first placed in a district d and village v, which are associated with data on budget priorities and village characteristics, particularly as relate to urbanness and economic activity. The same business i is also linked to public good inputs averages within its five kilometer radius j, including Education Inputs$_j$, Health Inputs$_j$, Police Inputs$_j$ and Court Inputs$_j$; the specific definition of each index follows in subsequent sections. Finally, business i is mapped to nearby Afrobarometer and Uganda National Household Survey Villages, also within its given j radius.

Using these data linkages, I build the following indices:

Input Capacity$_{jv}$: The input capacity index combines input-focused measures from the education, health care, courts and police datasets, taking normalized five kilometer averages of the relevant measures from the EMIS, HMIS, and JLoS datasets, respectively, including measues of personnel, infrastructure, equipment, and supplies. This includes subsets of Education Inputs$_j$ and Health Care Inputs$_j$, as well as Court Inputs$_j$ and Police Inputs$_j$ in full, and a measure of Health Spending$_v$ from the village-level data.

Urban$_v$: The urbanness index combines normalized measures of urbanness from the GeoQuery database, including nighttime lights, population density, and accessibility to cities.

Human Capital Capacity$_j$: The human capital index combines normalized versions of the Education Inputs$_j$, Health Care Inputs$_j$, and Court Inputs$_j$ indices, five kilometer averages of input indices from the EMIS, HMIS and JLoS datasets, respectively, incorporating normalized measures of personnel, infrastructure, equipment, and supplies.

Enforcement Capacity$_j$: The enforcement index is a five kilometer average of the police input index from JLoS, Police Inputs$_j$, incorporating normalized measures of personnel and infrastructure.[25]

New Services$_j$: The new services index combines measures of service investment from the education, courts and police datasets, all the datasets for which I am reliably able to identify recent investment. Specifically, it combines measures of the proportion of recently constructed schools, courts, and police posts in the vicinity of a given business.[26]

For a precise list of the variables in each index, see Appendix Section B.5. Note that observations are in general balanced at the 5% level across these indices using the Google Maps data, as can be seen in Appendix Table A.2.[27]

District and village-level data

Using a combination of the physical addresses and latitudes and longitudes, I was able to match each business to its relevant contemporary district d, Uganda's primary subnational administrative unit; this enables me to match each business to its relevant local governmental budget for the year prior to the treatment (MoFPED, 2020). For this project specifically, I use the approved district and municipality level budgets for the 2018-19 fiscal year, which contains information on local government revenue and spending. These are supplemented with the Uganda national budget for the same year, which contains information on spending in Kampala.

I am also able to match each business to a village unit v, linking it to the GeoQuery database (Goodman et al., 2019) and therefore to rich data on nighttime lights, population density, and distance to the nearest urban center (Elvidge et al., 2017; Weiss et al., 2018; CIESIN, 2016). In other words, each business i is geographically mapped to a village unit v, and therefore to measures of Nighttime Lights$_v$ from 2015, Population Density$_v$ from 2015, and Accessibility to Cities$_v$ measuring travel time to cities of 50,000 or more from 2015. These three measures are combined into an urban index, as discussed in Section 2.5. Additionally, I use the GeoQuery database to link businesses to health sector project spending in 2014-15, using a georeferenced database which maps projects to the local level for that year only, which produces the Health Spending$_v$ measure (AidData, 2018).

[25]In practice, although direct enforcement actions by URA are rare, tax enforcement in Uganda is implemented largely by the Ugandan police.

[26]Note that the date of recently constructed varies depending on the data I have access to: for schools, recent is defined as 2011 to 2016; for courts and police posts, recent is defined as 2011 to 2013. In each case, the last year of recent is the latest year for which I have data. The index and all analytical results are robust to the use of other 'start' years in the definition of recent.

[27]The only index for which the results are significant at below the 5% level in the test of joint equality for all coefficients is the court index, one subcomponent of the human capital index; however, the full index is balanced. The urban index is also significant at the 10% level; this variable is used as a control in most specifications. There are no significant differences across the indices when looking at the test pooling the treatments and comparing them to the control group.

Health center and schools data

Next, I utilized data from the Uganda Ministry of Health Information Management System (HMIS) (MoH, 2016) and the Uganda Ministry of Education Information Management System (EMIS) (MoES, 2016). HMIS provided me with rich (if incomplete) data on each health center in Uganda, including data on procedures, supplies stockouts, and infrastructure.[28] EMIS provided me with detailed data on each school in Uganda, including pupil-teacher ratios and infrastructure. In both cases, I had physical addresses with varying degrees of accuracy; these, too, I ran through the Google Maps and ArcGIS geocoding processes to transform them into latitudes and longitudes. This geocoding allows me to map health centers and schools to businesses, and specifically to find, for each business i, all health centers and schools within a five kilometer radius j.

For each school s and health center h, I construct the following measure:

$$\text{School Input}_s = \sum_{m=1}^{n} \frac{X_m^s - \mu(X_m))}{\sigma(X_m)}$$
$$\text{Health Center Input}_h = \sum_{m=1}^{n} \frac{X_m^h - \mu(X_m))}{\sigma(X_m)}$$

with normalized measures of input X specific to the HMIS and EMIS datasets. See Appendix B.5 for thte specific list of inputs included in each measure.

For each business i with five kilometer radius j, I construct:

$$\text{Education Inputs}_j = \sum_{s=1}^{n} \text{School Input}_s$$
$$\text{Health Inputs}_j = \sum_{h=1}^{n} \text{Health Center Input}_h$$

where n is determined by the number of schools or health centers in the business's specific five kilometer radius.

In other words, I construct a measure of health center or school inputs for each unit, and then create a business-radius-specific input measure based on averaging the unit input measures in the businesses's vicinity.

Police and local courts data

As well, for each participant, I am able to map them to every police post or local court within their five kilometer radius j, using a geocoded dataset from the Uganda Justice, Law and Order Sector (JLoS) which contains data on not only the location of police posts and

[28] In particular, I utilize maternal and child health scorecard measures, because they provide a wide variety of outcomes and have relatively less missing data.

local courts, but also their levels of personnel, equipment and infrastructure (JLoS, 2013). I construct a input index for each police post and local court, and for each business construct an average input index based on averaging across each unit in the match radius.

For each police post p and local court c, I construct the following measure:

$$\text{Police Post Input}_p = \sum_{m=1}^{n} \frac{X_m^p - \mu(X_m))}{\sigma(X_m)}$$
$$\text{Local Court Input}_c = \sum_{m=1}^{n} \frac{X_m^c - \mu(X_m))}{\sigma(X_m)}$$

with normalized measures of input X specific to the police and local court databases. See Appendix B.5 for thte specific list of inputs included in each measure.

For each business i with five kilometer radius j, I construct:

$$\text{Police Inputs}_j = \sum_{s=1}^{n} \text{Police Post Input}_p$$
$$\text{Court Inputs}_j = \sum_{h=1}^{n} \text{Local Court Input}_c$$

where n is determined by the number of police posts or local courts in the business's specific five kilometer radius.

Household survey data

I also link each participant to all nearby villages in the 2017 Uganda National Household Survey, a nationally representative survey of households and communities containing a variety of rich data on income, public goods, and other similar measures (UBOS, 2017). Specifically, I use the community surveys which ask a village official or equivalent about the availability and quality of health, education and police services in the enumeration area and its immediate vicinity. Using data on enumeration area addresses, I geocode each enumeration area separately using Google Maps and ArcGIS and link each business to all community surveys within a five kilometer radius.

Last, I link each participant to all nearby village in the 2015 geocoded Afrobarometer Opinion Survey, a representative survey which asks Ugandans (and citizens in many other African countries) a variety of questions about their political beliefs (Afrobarometer Data, 2015). I average the Afrobarometer data to the community level using relevant weights, and construct community-level opinion measures for each village.[29] I link every business to all Afrobarometer opinion villages within five kilometers, and average the results. In a handful

[29]As Afrobarometer does random sampling within its selected villages, this method of aggregation leads to a belief measure representative at the village level.

of cases, where no village is available within five kilometers, I use the closest Afrobarometer enumeration area within ten kilometers.[30]

For each business i with five kilometer radius j, I construct for question q:

$$\text{Average Response}_j^q = \frac{\sum_{x=1}^{n} \text{Village Response}_x^q}{n}$$

where n is the number of villages.

Cross-Method Reliability

Although I will generally refer to the Google Maps data in the following results sections, the main findings of the paper are robust to using either the Google Maps or ArcGIS-based dataset.[31] This robustness is despite their relative disagreements, which can be seen below; for example, while 75% of businesses are found within five kilometers between the two datasets, only 31% are found less than one kilometer apart between the two datasets.

Table 2.6: Google Maps and ArcGIS Geocoding

	Found (%)		GoogleMaps to ArcGIS (%)		
	GoogleMaps	ArcGIS	$\leq$5 km	$\leq$2 km	$\leq$1 km
Businesses	99%	99%	75%	47%	31%
Schools	99%	89%	85%	63%	50%
Health Centers	99%	45%	90%	77%	62%

2.6 State Capacity Heterogeneity

Using the datasets described above, I undertake a variety of methods to assess heterogeneity of the treatment to state capacity. First, I utilize the direct Input Capacity$_{jv}$ measure; next, in order to look at general public goods as opposed to enforcement-specific public goods, I instead look at interactions with Human Capital$_j$ and Enforcement$_j$. Last, I undertake a prediction exercise to focus specifically on capacity correlated to taxation, described in considerable detail below.

[30]Note that this sort of exception occurs for Afrobarometer because there are relatively few enumeration areas – 263 across all of Uganda for the 2015 round – leading to a relatively high number of business with no matches, particularly more remote businesses. I utilize the 2015 instead of the 2018 round precisely due to this issue, as only 140 villages were surveyed in 2018.

[31]Note that the sample size for Table 2.2, Table 2.3 and Table 2.4 were based on cross-matching with the Google Maps database. The results are robust to estimation on the full sample (not reported) as well as to estimation on the sample cross-matched in the ArcGIS database, as can be seen in Appendix Table C.2, Appendix Table C.3 and Appendix Table C.4.

All methods involve estimating versions of the following:

$$\text{Y}_{ijvd} = \alpha + \beta \text{Treatment}_i + \gamma \text{Index}_j + \delta \text{Treatment}_i \times \text{Index}_j + \text{Urban}_v + X_{ijv} + \delta_d + \varepsilon_{ijvd}$$

for individual i located in five kilometer radius j, village v and district d. Y_{ijvd} indicates whether or not that individual paid their tax, and X_{ijv} is a vector of individual, local and village covariates which vary across the specifications, but always include strata fixed effects.

Index_j may represent Input Capacity_{jv}, Human Capital_j or Enforcement_j depending on the specification. Specifications include a variety of controls, potentially including district fixed effects δ_d, the Urban_v index, and individual covariates X_i beyond what is specified above.

Heterogeneous effects

Table 2.7: Input Capacity Heterogeneity

	Paid Tax			
	(1)	(2)	(3)	(4)
Treatment	0.368**	0.352**	0.352**	0.352**
	(0.041)	(0.047)	(0.047)	(0.048)
Capacity Index	-0.120	0.342**	0.342**	0.294*
	(0.459)	(0.048)	(0.048)	(0.092)
Treatment x Capacity Index	-0.388**	-0.392**	-0.392**	-0.389**
	(0.042)	(0.036)	(0.036)	(0.037)
Observations	84955	84955	84955	84955
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Controls	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. All regressions include strata fixed effects and are estimated with robust standard errors. Depending on the specification, regressions may include district fixed effects, individual controls, and an index for the urbanness of the respondent's village. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

In Table 2.7, I estimate treatment heterogeneity by input capacity, which is the measure of state capacity most consistent with the subnational approaches used, for example, by

Acemoglu, García-Jimeno, and Robinson (2015).[32] The evidence suggests that the treatment is most effective in lower capacity areas, or areas in which the state seems to invest less in personnel, infrastructure, and supplies. Colloquially, in Uganda, many districts have regions within them which are referred to as hard-to-reach; these areas may be geographically more remote (especially from main roads), have a bad reputation, or otherwise governmental officials may find it more difficult to provide services there. In practice, there is good reason to think both that there is significant variation in capacity across Uganda, and that there are areas where the presence of the state is relatively less.

What I find, crucially, is that this low-cost, easy to implement text message intervention was most effective in these lower capacity areas, or in other words, the areas in which the state has invested less in inputs. This result suggests that digital technology can allow the state to broaden its reach; note that here I focus on individuals who had paid at least once before, but in a year in which it seems like, in the absence of the treatment, some fraction would not have paid.

One question which arises is whether it is possible to differentiate between public goods which are solely final consumption, and not in and of themselves "extractive," which I refer to here as human capital investments, and extractive capacity, i.e. capacity more specifically focused on enforcement. It is not clear, from the aggregate results on input capacity, whether these heterogeneity effects are driven by differences in non-extractive or extractive capacity. As discussed in Section 2.5, I construct indices which atempt to differentiate between the two, given data limitations.

In Appendix B.4, I construct a simple, one-period model of the state's tax revenue maximization problem. The state, working from a budget constraint, spends on some combination of enforcement, sensitization, and human capital investment. I introduce a non-linearity in enforcement, which reflects the real-world need for a country to build up meaningful infrastructure before enforcement is feasible, and allow for there to be positive tax compliance even in the absence of enforcement, consistent with evidence on the non-zero rate of compliance in Uganda. One crucial assumption is that I model compliance as a function of sensitization rather than enforcement based both on the fact that compliance occurs in low enforcement environments, but also in accordance with recent evidence on the importance of salience in economic decision-making (Chetty, Looney, and Kroft, 2009), which has been suggested to be relevant to taxation by Meiselman (2018) and Bergolo et al. (2019).

I explore the implications for the complementarity of the marginal change in compliance from sensitization (for which I argue my treatment is a good proxy) with enforcement and human capital spending, and their combination, which we might typically define as capacity, under zero and non-negative levels of enforcement. I focus here on the case of zero enforcement, which is most consistent with the study context, although the case with non-negative levels of enforcement proves interesting in considering the external validity of the results.

[32]Results disaggregated by treatment can be found in Appendix Table B.6; results estimated using the ArcGIS method of geocoding can be found in Appendix Table C.6. The results are robust to both alternative methods of specification.

In the model, I find that the effects of an increase in sensitization should be larger in areas where human capital capacity is low. The basic intuition is that under scarce resources, optimally, sensitization is a substitute for human capital investment. When the state is constrained, it should seek to balance between investmentes in sensitization and human capital, rather than putting all of its resources in one or the other.

Table 2.8: Human Capital and Enforcement Heterogeneity

	Paid Tax		
	(1)	(2)	(3)
Treatment	0.353**	0.353**	0.353**
	(0.047)	(0.047)	(0.047)
Human Capital Index	0.265		0.266
	(0.114)		(0.113)
Treatment x Human Capital Index	-0.480***		-0.479***
	(0.008)		(0.008)
Enforcement Capacity		0.034	0.060
		(0.863)	(0.761)
Treatment x Enforcement Capacity		0.041	0.003
		(0.839)	(0.990)
Observations	84955	84955	84955
Individual Controls	Yes	Yes	Yes
District FE	Yes	Yes	Yes
Urban Interaction	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. All regressions include strata and district fixed effects, individual controls, and an index for the urbanness of the respondent's village. All regressions are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Consistent with the model, I find significant and quantitatively large heterogeneity across the human capital capacity index, as can be seen in Table 2.8.[33] As predicted by the model, I find little to no heterogeneity on the enforcement capacity index, though it is difficult to

[33]Results disaggregated by treatment can be found in Appendix Table B.5; results using the ArcGIS method of geocoding can be found in Appendix Table C.5. The results are robust to both alternative methods of specification. Results using separate estimations for each component of the human capital index

rule out potential countervailing effects. To address concerns about potential correlation, I estimate interactions for the human capital and enforcement capacity indices in the same regression, though the results are not meaningfully different from separate estimations.

I find that the treatment was relatively more effective in areas where measures of human capital inputs are low. These results are true even when I control for district fixed effects and the urbanness of the business's village. In general, these results are fairly large in magnitude; the average treatment effect on payment is .35%, but grows to a predicted .8% when one looks at a business in an area one standard deviation below the mean in the human capital index.

I find positive but insignificant treatment effects for the enforcement capacity index. Note that one might expect here two countervailing effects, as enforcement is measured here by police inputs. Specifically, police may serve as both a human capital investment (police protection being a potential public good) and also increase enforcement capacity. In the model, when taxes are enforced, one would expect countervailing pressures from these two effects. The results are also consistent with a functional lack of enforcement of taxation, which is empirically the case in this context.

Predicted tax compliance

The measures in the previous section focus on inputs, without being able to think about how they translate to outputs, or the efficacy of the state. In the next section, I try to tackle heterogeneous effects which differentiate by how effective the state is at translating inputs into outputs. I focus specifically on tax compliance, which I would suggest should be considered the most relevant pillar of statehood in this context. To do so, I undertake an exercise to predict tax compliance based purely on local factors, which measures how local inputs translate into tax capacity.

Among the control group, I estimate:

$$Y_{ijvd} = \theta + \zeta J_j + \tau V_v + \xi(J_j^1 \times J_j^1 \dots J_j^N \times V_v^1 \dots V_v^N \times V_v^N) + \delta_d + \epsilon_{ijvd}$$

for individual i located in five kilometer radius j, village v and district d. Y_{ijvd} indicates whether or not that individual paid their tax. $J_j^1 \dots J_j^N$ is a vector of local covariates within the respondent's five kilometer radius, and $V_v^1 \dots V_v^N$ is a vector of village covariates, including data from every source described in Section 2.5 above. All local and village covariates are interacted with themselves (quadratics) and with every other variable in both vectors. I also include district fixed effects δ_d.

The inclusion of all quadratics and interaction specifications means that the set of these variables is large; the fact that many of the variables are different measurements of the same underlying fundamentals leads to potential concerns around multicollinearity. As such, I use a machine learning technique, Least Absolute Shrinkage and Selection Operator ("LASSO")

can be found in Appendix Table A.3, with further disaggregation in Appendix Table A.4. Results using separate estimation for each component of the enforcement index can be found in Appendix Table A.5.

to optimize the set of predictors of tax compliance. Specifically, I use a Post-LASSO estimation technique, in which I use LASSO to select predictors among the large set of public good and economic condition-related variables, then use the variables chosen by this procedure to estimate the relationship between these variables and tax collection among the control group using OLS.[34]

I then predict $\hat{Y}_{jvd}$, a purely environmental measure of tax compliance; $\hat{Y}_{jvd}$ is essentially a weighted average of a subset of public good inputs and other local factors, where the selection of inputs is chosen by the LASSO model's determination of their contribution to tax compliance.

I then estimate versions of the following across the full sample:

$$\mathrm{Y}_{ijvd} = \alpha + \beta \mathrm{Treatment}_i + \gamma \hat{Y}_{jvd} + \delta \mathrm{Treatment}_i \times \hat{Y}_{jvd} + \mathrm{Urban}_v + X_{ijv} + \delta_d + \varepsilon_{ijvd}$$

where X_{ijv} represents individual, local and village covariates, depending on the specification. As elsewhere, X_{ijv} always includes strata fixed effects. To adjust the variance matrix in the second step for potential measurement error in the predicted value $\hat{Y}_{jvd}$, I bootstrap the two-step procedure for each regression in the subsequent tables.[35]

I estimate these effects in Table 2.9.[36] I find large and statistically significant treatment effects in the same direction as the input capacity measure, although relatively larger. The effects are consistent across a variety of specifications; my most preferred includes district fixed effects, individual controls, and all indices with interactions. Even when controlling for all these factors, there is a significant and large positive coefficient on predicted tax capacity, suggesting that for each standard deviation increase in this measure, there is a 1.5 percentage point increase in the proportion paying their taxes.

The treatment closes between forty and eighty percent of the capacity gap in tax compliance, depending on the specification.[37] In other words, this low cost, low capacity treatment significantly decreases the compliance gap between low and high capacity areas of Uganda, considerably increasing the likelihood of payment in areas where it was previously lower. Interestingly, as can be seen in Appendix Table B.7, there is a negative and statistically significant relationship for all three treatments on the interaction term.

[34]The use of Post-LASSO in the context of this sort of prediction problem is in keeping with a large literature suggesting the estimator enhances prediction accuracy; see Belloni, Chernozhukov, and Hansen (2011) for an example in the context of instrumental variables. The results, however, are quantitatively robust to the use of OLS, Elastic Net or Ridge regression for the prediction stage.

[35]This approach is in accordance with papers like Petrin and Train (2003) which, although focusing on discrete choice models, shows that in the context of predicted regressors, the bootstrap provides very similar standard errors to an asymptotic formula-based approach such as would be prescribed by Murphy and Topel (1985).

[36]Results disaggregated by treatment can be found in Appendix Table B.7; results estimated using the ArcGIS method of geocoding can be found in Appendix Table C.7. The results are robust to both alternative methods of specification. The directionality, relative magnitude and patterns of significance are also robust to estimating the prediction equation with OLS, as in Appendix Table A.7, elastic net regression, as in Appendix Table A.8, and ridge regression, as in Appendix Table A.9.

[37]Note that the magnitude of the interaction effect remains roughly similar; it is the magnitude of the gap between low and high capacity which decreases as various controls are added.

Table 2.9: Predicted Capacity Heterogeneity

	Paid Tax			
	(1)	(2)	(3)	(4)
Treatment	0.338*	0.351*	0.348*	0.379**
	(0.066)	(0.053)	(0.055)	(0.037)
Predicted Tax Capacity	2.854***	2.545***	1.401***	1.530***
	(0.000)	(0.000)	(0.000)	(0.000)
Treatment x Predicted Tax Capacity	-1.237***	-1.188***	-1.064***	-1.212***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	82165	82165	82165	82165
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Control	No	No	Yes	Yes
All Index Interactions	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. Predicted tax capacity is estimated in a first stage using LASSO regression. All regressions include strata fixed effects; standard errors are bootstrapped using 1000 replications of the two-step procedure. Depending on the specification, regressions may include district fixed effects, individual controls, an index for the urbanness of the respondent's village and the human capital, enforcement and urbanness indices with treatment interactions. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

The next question is why this result may exist. I explore potential explanations, including that high capacity areas already fully paying and that the results are driven by differences in underlying political opinions or the types of businesses found in lower capacity areas, and generally rule them out. Instead, I will argue that the results may be driven by a subset of areas with recent investments in state capacity, and that these findings imply an important role for fiscal exchange.

As mentioned, one potential explanation for these results is that in high-capacity areas, payments are already at such a high level that it is unlikely there are more taxpayers to reach. Strong evidence against this explanation, however, is that at the time of treatment, only 30% of the sample had paid any of these individual taxes in 2018-19. While there is reason to believe that the entirety of the sample may not have needed to pay one of Uganda's individual taxes in the current year (such as if someone's business had closed in recent years), 30% is sufficiently low that it is plausible to suggest there is room for growth in all regions. As well, the difference in compliance rates for the control group with a one standard deviation increase in predicted capacity is only 1.5 percentage points; this small magnitude represents further evidence that a "ceiling" on tax compliance is unlikely to have been reached, even in the highest tax capacity areas.

Tax Compliance and Political Views

Table 2.10: Predicted Tax Capacity and State Perceptions

	Predicted Capacity
To find out what taxes and fees you are supposed to pay is:	
Not very difficult	$+^{***}$
Not paying the taxes they owe on their income is:	
Wrong and punishable	$+^{***}$
A good citizen should always:	
Pay taxes they owe to government	$+^{***}$
The tax authority has the right to make people pay taxes:	
Agree or Strongly Agree	$+^{***}$
Ease of Service Access Index	$+^{***}$

Each row in this table is the result of a different regression correlating predicted tax capacity estimated using the Google Maps method of geocoding and local averaged Afrobarometer opinions based on the reported questions. Ease of Service Access Index is an index combining z-scores of a variety of questions on service access; all other regressions include only a single question, defined as in the table. All results reported are statistically significant, and all regressions control for district fixed effects and index measures of local health inputs, education inputs, urbanness, court inputs and police inputs.

In order to understand better what the tax capacity measure means in this context, I also correlate this measure directly with a handful of Afrobarometer survey questions in Table 2.10.[38] Note that these results are robust to whether Afrobarometer questions are included or excluded from the construction of predicted tax capacity, and involve district fixed effects and controls for all above-mentioned indices.

On the whole, the Afrobarometer data suggests that knowledge of taxation and perceptions of the government are significantly different between low and high capacity areas, in ways that may shed light on the mechanisms of the results presented in this paper. These results also rule out another plausible explanation for the strong results on tax capacity: namely, that for whatever reason, higher tax capacity is associated with worse attitudes with regard to tax payment. One example of this would be if higher capacity areas had

[38]Direct interactions of Afrobarometer measures with the treatment (not reported) produce insignificant results; this is likely due to both the large amount of missing data in the Afrobarometer measures and the non-random nature of the missing data, though this cannot be determined with certainty. For a replication of this table using the ArcGIS method of geocoding, see Appendix Table C.8; the results are robust to this alternate method of data construction.

lower beliefs in governmental legitimacy; in that case, it might be those beliefs which lead to the results presented herein, and not some aspect of capacity itself. However, the results which I find when correlating the data with the Afrobarometer survey seem, if anything, to go in the opposite direction, likely ruling out this explanation.

I find that in areas with high predicted tax capacity, nearby villages were significantly more likely to report that it was not "very difficult" to find out what taxes and fees they are supposed to pay to the government. This result suggests that part of the reason the information treatment may have worked is that the government may, simply, be providing less information at baseline to more remote areas.

Similarly, I find a number of statistically significant correlations in beliefs about state engagement and state legitimacy. First, individuals in areas with higher predicted tax capacity were more likely to report that not paying the taxes they owe on their income was wrong and punishable, as opposed to answering either wrong but understandable, or not wrong. It is difficult to know what to make of the difference between punishable and understandable in this context; it could either mean that they have a greater belief in such areas that they will be punished for evading, or it could be a moral belief about what is owed to the state. None the less, these results may help explain why punishment-focused messages worked better in low tax capacity areas.

Along the lines of the second explanation, I find villages in high capacity areas more likely to report that a good citizen should always pay the taxes they owe to the government; this is suggestive of the sort of reciprocal beliefs found elsewhere, such as in Martin (2016), and consistent with the fiscal exchange model of tax compliance. I also find villages more likely to agree or strongly agree with the statement that the tax authority has the right to make people pay taxes. Although this question is commonly interpreted as a measure of state legitimacy, it could also potentially be interpreted as referring to the capacity of the tax authority to make individuals pay, rather than the legitimacy of the tax authority to do so; disentangling the two explanations is difficult.

Reassuringly, given the construction of the capacity indices, I find villages in high predicted capacity areas report a higher ease of access to services, including education, health, bureaucratic services, water, police and courts; recall that these findings control for district, urbanness, and a host of other service input factors. These findings are reassuring, as they suggest the differences found using administrative data in public good inputs are also meaningfully observed by the local population.

People vs. Place

Next, I return to the individual heterogeneity results from earlier, and assess, to the extent possible, whether the treatment heterogeneity that I identify seems to be driven by business characteristics as opposed to local capacity.

For example, one might hypothesize the treatment is more effective among newer businesses, and that lower capacity areas - having seen less government investment in various public goods - have a higher proportion of relatively new businesses; an analogous argument

Table 2.11: Predicted Capacity and Individual Heterogeneity

	Paid Tax			
	(1)	(2)	(3)	(4)
Treatment	0.398**	0.408**	0.407**	0.407**
	(0.030)	(0.025)	(0.025)	(0.025)
Treatment x Predicted Tax Capacity	-1.264***	-1.207***	-1.094***	-1.087***
	(0.000)	(0.000)	(0.000)	(0.000)
Treatment x Years Since Registration	-0.118**	-0.110**	-0.108*	-0.108*
	(0.038)	(0.048)	(0.053)	(0.052)
Treatment x Last Tax Payment (1k USD)	-0.279*	-0.250	-0.256	-0.257
	(0.085)	(0.108)	(0.102)	(0.101)
Observations	82165	82165	82165	82165
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Control	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. All regressions include strata fixed effects; standard errors are bootstrapped using 1000 replications of the two-step procedure. Depending on the specification, regressions may include district fixed effects, individual controls, an index for the urbanness of the respondent's village and the human capital, enforcement and urbanness indices with treatment interactions. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

could be presented for tax payment amount. In other words, heterogeneous results which seem to be driven by capacity may be the result of underlying differences in business type. To see whether this is the case, I run a regression interacting the treatment with years since registraiton, size of most recent payment, and the measure of predicted tax capacity, under various specifications.

The results of this undertaking are presented in in Table 2.11.[39] I find that the results on predicted tax capacity persist strongly, with the coefficients only slightly smaller than in specifications without the additional interactions. The results for recent payment amount also persist, suggesting that there are distinct interaction effects of the treatment with both how recently the business was established and underlying levels of tax capacity. Though loca-

[39]Results disaggregated by treatment can be found in Appendix Table B.9; results estimated using the ArcGIS method of geocoding can be found in Appendix Table C.9. These results are robust to both alternate specifications.

tion is in itself endogenous, one can nonetheless conclude that measures of local public goods remain significant in determining the magnitude of the treatment effect when interactions with more individual characteristics are included.

It is more challenging to interpret the results on payment size, largely because tax payments explicitly vary based on location for the presumptive tax (particularly at the lower end, where the majority of the business sample is located). The results decrease in significance, and are just above the threshold for statistical significance in most specifications, although the coefficient remains negative and roughly the same magnitude as the regression without the inclusion of predicted tax capacity.

On the whole, the tax capacity heterogeneity and individual heterogeneity results are, separately and together, most consistent with the explanation that the text message treatment works well to reach individuals not previously reached by the state. This may be because the state has less (or a less effective) presence in the business's vicinity, or because the business is newer (or smaller). These results are nonetheless promising, as they suggest that low-cost, high feasibility methods of encouraging tax compliance may reach taxpayers who are more challenging for low-capacity states to reach with other, more traditional methods of tax encouragement and enforcement.

Recent Investments in Capacity

Last, in Table 2.12, I ask whether growth in service capacity might explain the strong results on capacity.[40] Using the school, health, police and courts datasets, I construct a measure of new investment in public goods New Services$_j$, which looks at the proportion of schools, police posts and local courts proximate to a business which were built relatively recently.[41]

The magnitude of the direct interaction between the treatment and predicted tax capacity decreases when we include the interaction with new services. That said, the directionality of the effect remains similar. We see a treatment effect of roughly .4 percentage points at the mean of predicted capacity and new services (as both are normalized prior to inclusion). Holding new services fixed, the treatment effect one standard deviation below the mean for predicted capacity increases dramatically, to 1.3 percentage points. If we look at the treatment effect one standard deviation above the mean for new services, the same group would infact have an estimated treatment effect of 1.6 percentage points (leaving out the insignificant and near zero direct effects of new services interacted with tax capacity and treatment).

[40] Results disaggregated by treatment are available in Appendix Table B.10; results estimated using the ArcGIS method of geocoding can be found in Appendix Table C.10. Note that, uniquely among the replication tables, the triple interaction between new services, tax capacity and treatment is not significant using the ArcGIS estimation, although the direction of the coefficients remains consistent. It is difficult to discern whether this is due to the change in sample which results from using the ArcGIS data or from differences in the results of the geocoding process.

[41] The New Services index is not included in the prediction of tax capacity, but is based on the same underlying datasets. Note that there appear to be no detectable direct interaction effects of the New Services index with the treatment (unreported).

Table 2.12: Predicted Capacity and New Services Heterogeneity

	Paid Tax			
	(1)	(2)	(3)	(4)
Treatment	0.361**	0.372**	0.361**	0.361**
	(0.050)	(0.041)	(0.047)	(0.047)
Treatment x Predicted Tax Capacity	-0.976***	-0.962***	-0.939***	-0.940***
	(0.000)	(0.000)	(0.000)	(0.000)
Treatment x New Services Intensity Index	-0.028	-0.028	-0.005	-0.005
	(0.890)	(0.885)	(0.981)	(0.980)
New Services x Tax Capacity x Treatment	-0.461***	-0.427***	-0.326*	-0.326*
	(0.004)	(0.006)	(0.051)	(0.050)
Observations				
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Control	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made from the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the method of geocoding. All regressions include strata fixed effects; standard errors are bootstrapped using 1000 replications of the two-step procedure. Depending on the specification, regressions may include district fixed effects, individual controls, an index for the urbanness of the respondent's village and the human capital, enforcement and urbanness indices with treatment interactions. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

In other words, the treatment effect is larger at lower levels of predicted capacity, even when holding fixed new services; however, it is even larger in areas where there were relatively greater increases in new services, suggesting the treatment worked best in low-capacity areas where the government had recently invested more intensively in capacity. Note that there is actually a positive correlation in the data between tax capacity and new service investment, suggesting that the government is not exclusively investing in low capacity areas; nonetheless, the results on treatment intensity seem to only occur when the investment has occured in lower capacity areas.

These results further suggest that while traditionally defined fiscal exchange messages are not shown to work in this context via the average treatment effects of this experiment, there may still be evidence that fiscal exchange matters. One way to interpret these results is that the combination of new service investment and low-cost messaging may expand the fiscal exchange reach of the state, leading not only the enforce, but also the inform and encourage messages to work most effectively in low capacity areas with higher levels of recent investment.

2.7 Conclusion

In this paper, I combine the test of a low-cost, high feasibility method for improving tax compliance with an extensive, granular dataset in order to explore the interaction between increasing tax compliance and existing levels of state capacity. Tax encouragement randomized control trials of the sort conducted here – sending letters, emails or texts to prospective taxpayers reminding them of their obligations – are common across the developed world, and have helped us understand the levers of tax compliance.

This paper joins a nascent literature showing that such interventions can work even in relatively low-capacity contexts. Specifically, I undertake a tax encouragement randomized control trial in Uganda, among a population with self-reported taxation and very little in the way of direct enforcement. The experiment works: the pooled treatment effect is roughly 6% of the control mean, and the rate of return is considerable, on average 6x and as much as 13x for the enforcement-focused message. I find that, like in many other contexts, information- and enforcement-focused appeals are more effective than encouragement-focused style appeals for the average respondent. I also find that the treatment is relatively more effective among newer and lower-paying businesses.

Next, I utilize a novel subnational database of public good provision to assess the variation in treatment effectiveness across low and high state capacity areas. By measuring service inputs at a highly local level (within five kilometers of a business or in its village), I can construct proximity-based measures of education, health, police and court services, as well as controlling more broadly for local economic activity and local government spending. I construct two different measures of state capacity, one a purely input-based measure, and one an output based measure which uses LASSO regression to filter these inputs through their ability to predict repayment in the control group.

The results suggest that tax compliance is generally higher in higher capacity areas, but that the treatment closes a significant fraction (forty to eighty percent, depending on the estimating equation) of this gap. Evidence suggests, given relatively low rates of tax compliance across the board, that these results are not driven by higher capacity areas having already reached a "ceiling" on tax compliance. These results are also robust to controlling for individual heterogeneity, and suggest that both people and places matter in thinking about treatment effectiveness (in other words, newness and smallness do not simply proxy for low state capacity, and vice versa). Looking at political opinions in high vs. low capacity areas also suggests that the difference in treatment effectiveness is not driven by a negative correlation between tax-related attitudes and capacity; if anything, individuals in high capacity areas seem to generally have a more favorable view of the state's right to tax.

I do find, however, that the intensity of recent investment in services strongly predicts treatment effectiveness in low capacity areas. Specifically, the interaction effect between the treatment and predicted tax compliance is strongest in places where investments have recently been made in public goods, suggesting a potential channel for the extension of fiscal exchange.

An important caveat to these results is that it is not necessarily clear from these short-run findings what the long-run implications of the results may be. If the treatment works primarily through increasing perceived beliefs about direct enforcement, then it seems likely that either the results will fade over time, or that there may be a potential backlash in the absence of true increases in enforcement likelihood. On the other hand, if the treatment works primarily through extending the reach of fiscal exchange, the results may in fact persist over time. I intend to continue studying the effects of this treatment when feasible, and hope to ultimately answer these questions.[42]

Furthermore, while these results suggest an important extension of fiscal exchange as a result of new public goods provision in low-capacity areas, it is not clear what the underlying mechanism is for this change. Several possibilities include the theory that the marginal fiscal exchange returns are higher when capacity is low (or perceived to be higher by taxpayers) or that new investments spark a sort of behavioral reciprocity in potential taxpayers. More work is necessary to better understand and potentially differentiate between the different channels of this theory.

Overall, these results indicate that simple, low-cost, easy to implement interventions can have a significant effect, even in low-capacity environments. Second, they suggest that low-cost treatments may serve to reach areas where otherwise the state struggles to extend its presence. Taken together, these findings suggest that simple, text-message based interventions – at least in the short term – are a promising way for the state to not just increase revenue, but potentially to build capacity.

[42]A planned extension of this study was put on hold due to COVID-19, but is expected to resume in future years.

Chapter 3

The Costs of Splitting: Administrative Unit Proliferation and Economic Growth in Uganda

3.1 Introduction

The question of the optimal size of governments, states and subnational units alike, is both old and important.

Intuitively, not all administrative unit proliferation is created equal. One can think of context-specific factors in determining costs and benefits: populations and their distribution, economies of scale or lack thereof, the degree of local tastes and knowledge, and more. Such parameters might lead to an "optimal" size. Against this, a split might be deemed beneficial if existing units are large relative to the optimal, and non-beneficial otherwise.

In Uganda, however, I find evidence which suggests that even when smaller units may be beneficial in terms of economic development, the creation of new districts can still disadvantage the population within them if there is a lack of adequate local governance. These findings may shed light on prior contradictory results of the impacts of such reforms on service provision, and suggest that governmental capacity should be taken seriously as an intervening factor in understanding similar policies.

I undertake this analysis by building a rich and detailed database which allows me to look at the subnational allocation of resources governmental and extra-governmental, inputs into education, a variety of measures of well-being from different remote sensing and household survey data sources, data on economic activity via taxation, and political response in terms of both levels of approval and voting behavior. Crucially, the majority of this data is at a level below that of the newly created administrative units (districts), allowing me to track individual units in a differences-in-differences framework through the splits. As some of this data includes all units (villages in some case, schools in others), I can also examine the distributional effects of unit creation.

The wealth of data which I am able to leverage for this paper allows me to look at different aspects of administrative unit proliferation, contextualizing the broader, more theoretical construct in specific ways.

In the Ugandan context, parent districts inherit the name and governmental structure of the district which existed prior to the split; new districts are headquartered elsewhere, and must build their own governmental structure. I find compelling evidence that the creation of new districts in Uganda leads to improvements in service delivery for all splitting districts, including both parent and new. Evidence on broader indicators of welfare, however, suggests that parent districts benefit overall from the shift, but new districts do not; a variety of indicators suggest that new districts may actually be relatively disadvantaged, at least in the short- and medium-run. These findings, plus results on shifts in resources available to new districts following splits, suggest that the creation of new districts comes with a trade-off. Smaller districts seem to be more beneficial for growth at Uganda's current level of district size, but not when they come at the cost of lower quality leadership.

Specifically, while new and parent districts both see relative improvements in the provision of education, only parent districts benefit in terms of economic and welfare gains for households. In many cases, new districts actually see a relative worsening on such measures, including nighttime lights, well-being, assets, and economic activity as measured by tax behavior. Unlike other studies, I do not find electoral or political benefits to district creation, once I account for regional trends.[1]

In the rest of the paper, I begin by discussing the background of broader research on decentralization and administrative unit proliferation, as well as Uganda's specific history of administrative unit creation. I then discuss in some detail the variety of administrative, secondary and remote sensing datasets utilized in this analysis, and my econometric strategy. I discuss a variety of results on the effects of administrative unit proliferation on budgetary allocations, educational inputs, development and well-being broadly defined, economic activity, distribution of resources, and political behavior. Taken together, these results suggest that even in a context where smaller units may be beneficial, newer units may be unable to reap these benefits if they do not inherit sufficient governmental capacity.

3.2 Background

Understanding the optimal size of political units is an important and much studied research question. There is still much to be understood as to whether smaller political units result in improved service quality. If they do, as has been shown in some contexts, the question of why they do so is crucial to understanding in assessing the optimal number and size of districts.

Decentralization, or the process of extending power and authority to subnational units, is a fundamentally distinct process from that of administrative unit creation, though not unre-

[1]Note that the lack of such decentralized gains do not mean that district creation is not poltiically motivated, nor that it is not politically beneficial.

lated. Decentralization, for example, may lead to increased demand for new administrative units by increasing their value (Grossman and Lewis, 2014).

Nonetheless, the broad literature on decentralization can offer insights into the effects of increasing the number of administrative units. A large literature has studied the effects of decentralization on public good provision and service delivery. In many contexts, including Bolivia (Faguet, 2012), Argentina (Habibi et al., 2003; Galiani, Gertler, and Schargodsky, 2008) and Indonesia (Kis-Katos and Sjahir, 2014), there is evidence that decentralization reforms increased investments in education and health, and even improved outcomes. In Uganda, research by Akin, Hutchinson, and Strumpf (2005) found decentralization led to decreased public health expenditures, suggesting that results may vary with context.

One challenge in understanding these results is that the composition of decentralization policies often differs widely across contexts. In particular, decentralization generally implies that fiscal authority devolves; this is not necessarily a feature of administrative unit proliferation. A narrower literature focuses specifically on administrative unit creation, looking at causes and consequences in Senegal (Gottlieb et al., 2019), Kenya (Hassan, 2016), Vietnam (Malesky, 2009), and Indonesia (Kimura, 2013; Pierskalla, 2016), often approaching the question from a perspective of understanding how administrative unit proliferation affects the political structure of a country.

Cross-country studies are more rare. Grossman, Pierskalla, and Dean (2017) examine the concept of decreasing marginal returns to administrative unit proliferation, using a cross-country panel model and examining the relationship between government fragmentation and the quality of health and education services. The authors find a positive and significant effect of the number of regional governments, but one which decreases as the number of regional governments increases. They attribute this leveling out of effects to inefficiencies arising from a lack of economies of scale, though they note that a lack of local administrative staff may also contribute.

A broader question which much of this literature addresses is why administrative district proliferation might result in improved service delivery. One theory is homogeneity, whether ethnic or otherwise; if newly formed districts are more homogenous, it might be easier for governments to respond to the precise needs of citizens (Alesina, Baqir, and Easterly, 1999); collective action is also more easily achieved in homogenous units via the ability to sanction (Miguel and Gugerty, 2005). Yardstick competition (Maskin, Qian, and Xu, 2000), in which local governmental leaders are able to show competence through their performance, may be heightened by the creation of more administrative units, leading to improved service delivey (Grossman, Pierskalla, and Dean, 2017). As well, if there are costs of administrative distance (Asher, Nagpal, and Novosad, 2018), shrinking districts may improve service quality by reducing inattentiveness.

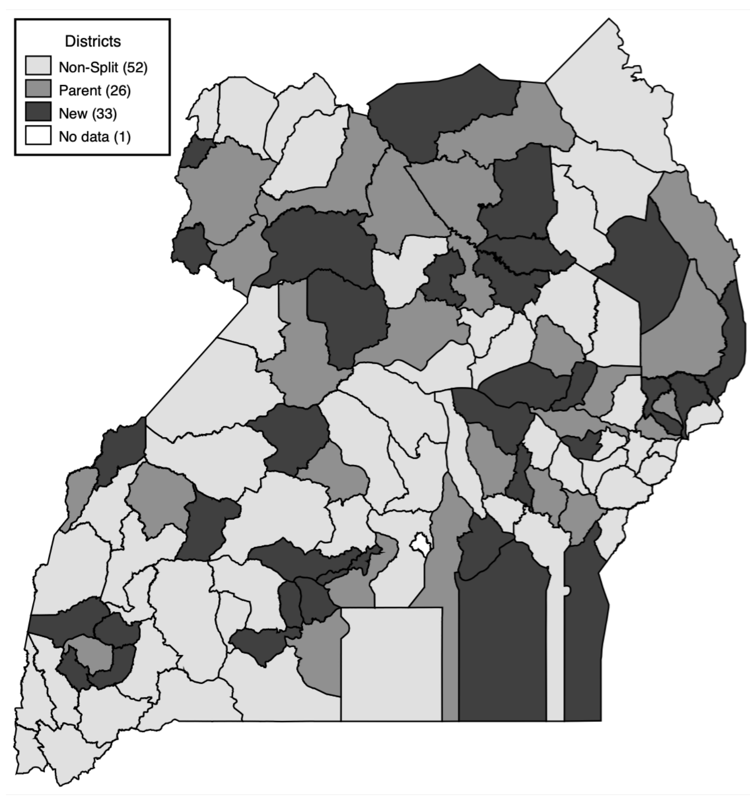
There are also potential downsides to administrative unit creation when it come to service delivery. Elite capture may become more likely as political units become smaller (Bardhan and Mookherjee, 2006); administrative costs may increase, making service delivery more expensive (Zax, 1989).

Uganda is an ideal setting for the study of administrative district proliferation, chiefly because over the last decades, it has created a substantial number of new administrative units.

When the National Resistance Movement first took power in 1986, Uganda had 33 districts, which are the chief subnational unit of governmental administration; this was an increase from the 16 districts Uganda had when it became an independent country in 1962. The number of districts increased rapidly, to 55 by 2001 and 79 by 2007. In 2009, the government of Uganda created an additional 7 districts, followed by another 26 in 2010. Starting in 2016, the district creation process began again; as of 2021, Uganda has a total of 134 districts. From 1986 to 2021, the number of districts roughly quadrupled; over the same time period, Uganda's population tripled, meaning the pace of district creation has outpaced population growth.

In this paper, I focus on the 33 districts created in 2009 and 2010, which took the total number of districts from 79 to 112 (including the capital city of Kampala). Figure 3.1 below divides Uganda into "non-split" districts which saw no changes in 2009 and 2010, "parent" districts which lost part of their territories, and "new" districts which were created by the split. The breadth and depth of data I am able to access from the mid-2000s through the mid-2010s allows for rigorous study of the impact of the creation of these districts, enabling me to control for long pre-trends, and analyze both the short-run and medium-run impacts of district creation.

Figure 3.1: District Status as of 2011



There are past studies that have examined the determinants of administrative unit proliferation in Uganda. Grossman and Lewis (2014) look at the determinants and electoral effects of district creation from 1996 to 2011, and find that intradistrict ethnic heterogeneity increases the likelihood of a future split. They also find that new districts electorally reward the president of Uganda in the election following the split, suggesting that political considerations likely play a key role in determining district creation.

The chief stated rationale for creating new sub-national units, including in Uganda, has been to impove the quality of services; Uganda's 1995 constitution allowed for the creation of new districts based on "the need to bring services closer to the people." In practice, the demand for new districts has often been attributed to calls by local leaders (B, 2007). Although there has generally been local demand for most new districts, Green (2010) notes that this explanation does not adequately capture the timing aspect of new districts. Instead, Green argues that new districts may be created as a form of patronage, as the creation of a new district not only results in a number of new seats in parliament, but also a variety of new district-level administrative positions.

Other research also focuses on the service quality effects of new district creation. Grossman, Pierskalla, and Dean (2017) explore the effects of district creation on service provision in Uganda, focusing on the 2005-2006 wave of district creation. By using retrospective fertility information from the Demographic Household surveys (DHS), they find that child mortality decreases in newly created districts in the years after the split, with supplemental evidence suggesting that the findings may be driven by increased access to antenatal care.

Other work, including Green (2015), examines the challenges that Uganda has faced in decentralization, including a lack of local funding, and staffing challenges. Other authors, including Lewis (2014), also note the governmental capacity in new districts tends to be low, at least initially, and such concerns have been broadly echoed in the Ugandan press as well. This is perhaps unsurprising: only at the upper echelon of government, the set of new technical and administrative staff a district requires include a Chief Administrative Officer (CAO), Resident District Comissioner (RDC), deputy CAO, deputy RDC, District Auditor, Clerk, Community Based Services Manager, Education Officer, Engineer, Extension Coordinator, Finance Officer, Director of Health Services, Information Officer, Inspector of Schools, Land Officer, National Agricultural Advisory Services Officer, Personnel Officer and Planner (Green, 2010).

There are a number of challenges in assessing the effects of administrative unit proliferation. One is that data from the level below that of the unit is often challenging, meaning tracking the same area in distinct data before and after a split can prove challenging. Another is that splits are non-random, and may in fact be deeply purposeful.

Another subsantial challenge in assessing the effects of administrative unit proliferation on service quality is that new districts are not created in isolation. New administrative units may be created for political reasons, meaning there is reason to expect that governmental resources will be more heavily allocated to the same areas which see administrative unit creation. New units, by definition, are accompanied by new leadership, whatever form that may take in a particular context.

In this paper, I address all three of these challenges. I use a differences-in-differences design to account for non-random selection of districts for splitting. I also build a rich secondary dataset which enables me to look both at inputs and outputs of the development process. By utilizing data on governmental and extra-governmental funding, I can carefully examine changes in funding, accounting for them in the development process. Finally, I am able to utilize data at a level lower than that of the unit which is newly created; this not only allows for richer analysis, but also lets me look at changes in distributions within newly created units, directly enabling me to test for whether newly created units might be more equitable in their distribution of resources.

3.3 Model

In this simple model, I construct a world in which the creation of new districts may be globally welfare enhancing, and examine how doing so might result in shifts in the distribution of govenrment resources between different technologies for development. Specifically, I allow there to be local technologies whose effectivenesses are more or less dependent on governmental capacity, and show that under the creation of a new district and certain specific (but plausible) conditions, a new district will see a (relative) shift in spending from the more complex to the simpler technology, potentially prioritizing simple gains over those which require greater support in their implementation.

Consider a government, or some other actor, deciding how to allocate resources with a goal of economic development. For the purposes of simplification, I assume that the actor is maximizing short-term development gains, whether due to electoral pressures or organizational ones. The actor is allocating a set budget Y across a population N. The population is divided in to two regions, A and B, such that $N_A + N_B = N$.

I assume that there are two available technologies for transforming the budget Y in to development, f and g, both with diminishing marginal returns. Both technologies work the same in regions A and B, with the distinction that f is a function only of investment y and population n and g is a function of investment y, population n, and regional capacity z. One can think here of regional capacity as the region's efficacy in arranging projects, in preventing elite capture and patronage, or any other locally-specific ability to increase project success.

Let us assume that returns to the two technologies are such that it makes sense for the government to make some investment of y in both f and g in both regions in order to maximize development; therefore, the initial allocation might look something like $(y_a^f, y_a^g, y_b^f, y_b^g)^*$ where $Y = y_a^f + y_a^g + y_b^f + y_b^g$.

Think of district creation as the splitting of the existing district B into B' and a new district C.[2] The total popluation N is unchanged, but now distributed such that $N_A + N_B + N_C = N$; the total budget Y is also unchanged. One can suppose as well that new capacity

[2]Note that given how f and g vary with N, the functions themselves, and how the allocation of z functions, such a split may be development-increasing for the system as a whole.

z does not appear instantaneously at district creation, such that $z_{B'} + z_C = z_B + \varepsilon$, with the caveat that $\varepsilon < z_B$. This means that by construction, $z_{B'} + z_C$ must be weakly less than $2z_B$.

The actor's decision, then, is to allocate resouces to maximize development under this new model. Intuitively, one can see that if either B' or C has a lower z than B did, then there would be an expected reallocation of resources from the g technology to the f technology. Thinking of the new optimal allocation $(y_a^f, y_a^g, y_{b'}^f, y_{b'}^g, y_c^f, y_c^g)^*$, one would expect that $y_{b'}^f + y_c^f > y_b^f$, and $y_{b'}^g + y_c^g < y_b^g$.

In other words, in response to the decrease in local capacity, one would expect the government to invest relatively more in development which has no capacity requirements, and relatively less in development which is increasingly effective in capacity.

3.4 Datasets

I utilize a variety of data sources in this analysis, including data from the Uganda Ministry of Education and Sports, Ministry of Finance, Planning and Economic Development, and the Electoral Commision of Uganda. The full set of data is summarized in Table 3.1, including years and frequency of observations, and discussed in further detail below.

Table 3.1: Data Sources

Data Source	Citation	
Uganda Ministry of Finance	2003 to 2015, annual	MoFPED (2020)
Uganda Ministry of Education	2006 to 2016, annual	MoES, 2016
GeoQuery Database		Goodman et al. (2019)
DSMP-OLS nighttime lights	2000 to 2013, annual	Elvidge et al. (2014)
NDVI Index	2008 to 2018, annual	Pedelty, Devadiga, and Masouka (2007)
Population	2000 to 2015, imputed	CIESIN (2016)
World Bank Aid	2000 to 2014, annual	AidData (2017)
Chinese Aid	2000 to 2014, annual	Bluhm et al. (2018)
Other Donor Aid	2000 to 2014, annual	AidData (2016)
Afrobarometer	2005 to 2018, intermittent	Afrobarometer Data (2018), BenYishay et al. (2017)
Uganda National Panel Survey	2001 to 2018, intermittent	Uganda Bureau of Statistics (2018)
Electoral Commission of Uganda	2006, 2011, 2016	Electoral Commission of Uganda (2021)
Uganda Revenue Authority	2003 to 2018, annual	URA, 2018

Budgetary Data

I utilize annual budgetary data from the Ministry of Finance, Planning and Development, available from 2003 to 2015 (MoFPED, 2020). This data contains information on all sources of funding for district governments, broken down by funding type as well as by spending categories. I combine this with population data (CIESIN, 2016) to get measures of annual per capita spending in total, by type and by category for each district. One caveat is that data from this source is missing for 2008 and 2009, but available for all other years in the sample.

An obvious challenge with using budgetary data is that prior to the existence of a district, there is no district-level budgetary data. Take for example the district of Buyende, which was created from the district of Kamuli in 2009. Although there are district-level budgets for Kamuli starting in 2003, the first district budget for Buyende appears in 2010. To deal with this challenge, I impute per capita amounts for the earlier years by combining the populations of the two districts, and assuming the budget of the parent district is spread equally across the whole population.[3] Once a district is created, I then separate the population and budgetary allocations.

It is possible that this population-based imputation either over-allocates or under-allocates funding to new districts; this method would be an over-allocation if pre-split districts tend to fund the areas which will be kept in the district more heavily, and an under-allocation if pre-split districts tend to allocate more funds to area which will become new districts. However, there does not exist any centralized repository of budgets for levels lower than the district; as such, I believe this method provides a reasonable proxy for estimating per-capita spending.

The average district receives roughly 90% of its funding in this period from the central government, with the other 10% coming from local revenues or external sources.[4] Conditional funding accounts for 70% to 80% of governmental funding, and is allocated fairly specifically to individual line items, such as teacher salaries, library maintenance, secondary school construction, teacher pensions, road rehabiliation, infrastructure support, and more. The remainder of central government funding comes in discretionary funding, which accounts for 10% to 20% of annual budgets, and which is much more loosely specified: for example, wage vs.non-wage spending, or district vs. urban spending.

Using line-item budgets, it is also possible to disaggregate spending by sector, including isolating per-capita spending on education, health, agriculture, water, and transportation by district. Education spending is one of the largest shares, comprising 30% of district spending on average from 2003 to 2015; that number grows over time, with the average district spending 36% of its budget on education in 2015. Administrative spending is another large category, although spending has declined over time relative to the total budget and per capita.

[3]Note that population is actually measured at the village level for each year; therefore, it is possible to know the population of the area where a district will be prior to the formal creation of that district.

[4]Historically, the contribution of local revenue was much greater, due to the graduated tax, an annual district-level tax on every adult which was abolished in 2004; the steep decline can be seen in Appendix Figure A.7.

Education Data

A second dataset which I take advantage of contains data at the school level for all schools in Uganda from 2006 to 2016 (MoES, 2016). This dataset includes, for each school, the number of teachers and students, as well as data on infrastructure, including classroom equipment and school water source. I use administrative unit data on the district, county and subcounty location of the school to place it in the correct post-2011 district.

The number of schools in Uganda increases over time, with roughly 11,817 schools in the data as of 2006, and 12,305 schools in the data as of 2016. The panel is unbalanced, with schools entering and exiting the data, due either to school openings, school closing, or data entry errors. In total, roughly 14,183 schools are represented in the data. Of those, 9,327 are represented in the full eleven years of data, and another 1,853 are represented in ten years of data. The remaining 3,000 or so schools have nine or fewer years of data; there are approximately 1,185 schools with four or fewer years of data.

In general, metrics like student-teacher ratios are fairly poor. In 2015, the average student-teacher ratio was 110, and the median student-teacher ratio was 101. About thirty five percent of schools did not have access to piped water or a borehole, meaning they relied on potentially unhygienic water sources. The average school had seven rooms, which is equivalent to one room per one hundred and sixty five students, and only 32% of students had what is referred to as adequate space, meaning they had both a chair and a desk.

GeoQuery Data

The GeoQuery dataset contains a combination of remote sensing data and data from research projects at AidData, which are mapped to each village in Uganda (Goodman et al., 2019). I utilize data on economic activity, including nighttime lights and vegetation indices, as well as data on aid volumes from three different sources; I discuss each in turn below.

I utilize two sources of remote sensing data, measuring economic activity via the proxies of nightime lights and vegetatian. I use the DSMP-OLS nighttime lights data, available annually from 2000 to 2013 (Elvidge et al., 2014). This dataset measures the intensity of nighttime lights, averaged at a village-level. I also use the normalized difference vegetation index (NDVI) data, available annually from 2008 to 2018 for every village in Uganda (Pedelty, Devadiga, and Masouka, 2007). The index measures the intensity of vegetation in the village in question.

I use population data, available in 2000, 2005 and 2015, in conjunction with the budget data above, to derive per capita spending estimates (CIESIN, 2016). The data is available at the village level, which allows me to easily aggregate it into the district units which exist post-2010. In order to smooth between years, I assume steady population growth between 2000 and 2005, 2005 and 2010, and 2010 and 2015, and impute accordingly.

I utilize three sources of data on project aid in Uganda, including World Bank aid (AidData, 2017), Chinese aid (Bluhm et al., 2018) and aid from other donors (AidData, 2016), all of which are available annually from 2000 to 2014.[5] This type of aid difers from budgetary or general support in that it is allocated to specific projects, and therefore specific locations.[6] Projects can take various forms, such as supporting infrastructure, education, employment, or other developmentally-related goal. I utilize population data to normalize each type of aid by population, creating per capita aid measures.

Afrobarometer Data

The Afrobarometer Survey is a pan-African series of national public attitude surveys on democracy, governance and society. Although not a panel survey, it was conducted on a nationally representative sample in Uganda in 2005, 2008, 2012, 2015 and 2018 (Afrobarometer Data, 2005; Afrobarometer Data, 2008; Afrobarometer Data, 2012; Afrobarometer Data, 2015; Afrobarometer Data, 2018). I use the geocoding conducted by BenYishay et al. (2017) to map households in each year to districts, and accordingly to treatment status; since the geocoding was done post-2011, I am able to map all households and survey areas to their post-district creation locations. In theory, the data includes roughly 10,000 households, relative to a goal of roughly 2,400 per survey from 2005 through 2015, and 1,200 in 2018; in practice, the number is slightly lower in most years, and only roughly 8,975 respondents answered the survey. These individuals are distributed across approximately 1,150 enumeration areas, most of which do not repeat across years.

Using the Afrobarometer data, I construct measures of central government support and local government support. The survey also includes measures of well-being and local development, which I construct into indices. In Appendix Section C.3, I detail the construction of each variable and index.

Uganda National Panel Survey Data

The Uganda National Panel Survey is a multi-topic panel household survey conducted in Uganda. The survey dates back to 2005, and in that round included some retrospective questions focused on 2001 (Uganda Bureau of Statistics, 2009). More recent rounds were conducted in 2009, 2010, 2011, 2013, 2015, and 2018, and include nation-wide sampling of households, and tracking of households and communities across years (Uganda Bureau of Statistics, 2010; Uganda Bureau of Statistics, 2011; Uganda Bureau of Statistics, 2013; Uganda Bureau of Statistics, 2015; Uganda Bureau of Statistics, 2018). The household survey involves questions on household members, education, health, labor force status, housing conditions, assets, sources of household income, consumption and more.

[5] The largest donor in the “other” group is the African Development Fund; other prominent donors include the European Union, United States, Norway, and the United Kingdom.

[6] The construction of the underlying dataset involves certain assumptions – for example, that aid which cannot be located at the sub-district level is spread evenly across the district – which might bias the data towards or against higher volumes of aid for new districts in the pre-period.

The panel is unbalanced, representing a total of 7,169 households which appeared in at least one round later than 2011. Of these, 1,336 households were surveyed in at least five of the seven rounds; another 1,162 were surveyed in three or more rounds. The remaining 4,670 were surveyed in either one or two rounds, due to household splits, households leaving or entering the sample, or a refusal to be surveyed. District information is recorded in the survey, and I use post-2011 locations to determine treatment status.

Using the household survey, I construct measures of access to electricity and safe water. I also conduct indices measuring employment levels, assets owned, house quality, and overall welfare. In Appendix Section C.3, I detail the construction of each variable and index.

Electoral Data

Uganda releases data on electoral outcomes at various levels for different elections, including at the subcounty elections for local councils, constituency level for parliamentary elections, at the polling station level for some presidential elections, and at the district level for all elections (Electoral Commission of Uganda, 2021). As subcounties and constituencies are at a lower level than districts, I am able to follow subcounties and consituencies from 2006 to 2016, including some which change districts during the 2009 and 2010 district creation processes, or to aggregate to the post-2010 district level. I use data on presidential elections in 2006, 2011, and 2016 to contsruct measures of voter turnout and support for the incumbent president, Yoweri Museveni. I use data on the parliamentary elections in 2006, 2011 and 2016 to construct measures of whether a candidate from the National Resistance Movement (NRM), Uganda's primary political party, won a given constituency seat. I use data on local council elections at the subcounty level in 2006, 2011 and 2016 to construct measures of whether a candidate from the NRM won a given council seat.

In general, both the NRM and Museveni himself were electorally successful during this period; in the 2016 election, Museveni won the presidential election with 60.7% of the vote, and the NRM won roughly 68% of the seats in Uganda's multi-party parliament. Approximately two-thirds of Ugandans voted in the election, a rate of turnout consistent with elections over the decade prior.

Taxation Data

I utilize data on registration for tax identification numbers and payment of individual taxes from the Uganda Revenue Authority (URA, 2018). Individuals must register themselves with URA in order to create a tax identification number (TIN), which is used to pay taxes or for the payment of wage taxes by one's employer; a TIN is also required for various other transactions, such as land titling or transfering a vehicle registration. The data includes both year of registration and location, dating back to the early 2000s. I also have access to tax payment rates for Uganda's individual taxes, including both the personal income tax and the presumptive tax, which is levied on small businesses; this data is available from 2013, as that was the year in which the tax collection data was first digitized.

Using this data, I construct measures of the proportion of new registrants relative to population in each post-2010 district in a given year, the proportion of total registrants relative to population in each post-2010 district in a given year, and the proportion of individuals paying one of Uganda's individual taxes relative to total registration in each post-2010 district in a given year. Registration levels began to increase dramatically in the early 2010s, spiking in 2014, though growth has been consistent in subsequent years. On average, from 2013 to 2018, roughly 2% of the population paid Uganda's individual taxes, although rates of payment vary by district.

3.5 Econometric Specification

My primary specification is

$$Y_{ijtr} = \gamma_i + \nu_t + \zeta_{rt} + \beta_1 \text{ParentTreated}_{it} + \beta_2 \text{NewTreated}_{it} + \varepsilon_{ijtr}$$

for a given subunit j (village, school, household, etc.) of district i in year t and subregion r.[7] For some datasets and specifications, I instead use the following:

$$Y_{itr} = \gamma_i + \nu_t + \zeta_{rt} + \beta_1 \text{ParentTreated}_{it} + \beta_2 \text{NewTreated}_{it} + \varepsilon_{itr}$$

where the only difference is that there is no subunit j smaller than the district i. In all regressions, I cluster standard errors on the smallest geographic unit, such as school, village or district.

In both specifications, ν_t is a vector of year fixed effects, γ_i is a vector of district fixed effects, and ζ_{rt} is a vector of subregion and year interactons to account for time trends. $\text{ParentTreated}_{it}$ and NewTreated_{it} are dummy variables which take the value one if an observation is from a parent or new district, respectively, in the time period after the split. Throughout, the coefficients of interest are β_1 and β_2, which, conditional on the assumption of parallel pre-trends conditional on district-level fixed effects, represent the difference-in-difference estimates of the impact of district creation on parent and new districts, relative to unchanged districts.

Crucially, in this specification, I differentiate between districts which are newly created by a split (new) and those which inherit the headquartes of the prior unit (parent). In the context of Uganda, this distinction is meaningful; parent districts tend to inherit the administrative infrastructure which existed pre-split, whereas new districts must build capacity. Correspondingly, one would predict that new and parent districts may see different changes in service delivery post-split, and distinguishing between them becomes analytically important.

Throughout, I exclude Kampala from the analysis; as the largest city in the country and the largest financial unit, Kampala is an extreme outlier on a variety of measures. I instead compare districts which have undergone shifts only to other district units.

[7] In the time frame in question, Uganda is divided into four regions and ten subregions; each subregion contains as few as seven or as many as twenty three districts.

3.6 Results

In Table 3.2, I show the effects of district splits on the level of resources allocated to a given post-2010 district, whether from the government or from foreign aid. In the first column, I look at the level of total government spending per capita, which averages to roughly $18 USD per person per year in the group of non-splitting districts. There are no statistically significant effects on district spending, and no statistically significant differences between parent and new districts.

It is worth noting that the allocation within districts of governmental spending prior to splits is considered to be equitable, i.e. the total pre-split budget is averaged evenly over the entire population. The lack of a difference between the parent and new districts in the post-period relative to the pre-period is another way of saying that, following a district split, parent and new districts have roughly equivalent per-capita spending. If, in the pre-period, resources were allocated more heavily towards the parent district, these results would underestimate the spending effects of district splits for new districts.

Spending is funded generally by the central government, which gives mostly conditional funds (i.e. funds allocated to specific purposes) but also some discretionary funds. For the average non-split district, discretionary funds total roughly $2 USD per year, or 11% of the total budget, and are largely allocated only to the broad purpose of development. There is a statistically significant increase in the discretionary funds allocated to both parent and new districts relative to other districts; the volume is large relative to the general discretionary budget level, amounting to 25% of the non-split mean in new districts and 12.5% of the non-split mean in parent districts.[8] However, these differences are a very small fraction of the overall budget – roughly 2% for new districts and 1% for parent districts. In Panel B, I find evidence of an increase in per-capita spending on the administrative sector in both parent and new districts, equivalent to increases of 2.5% and 4%, respectively, relative to the non-split mean. I also find evidence of different budget priorities in new districts, relative to both parent and other districts, where per capita resource allocation shifts slightly away from health services and towards water and sewage services.

There are larger results for foreign aid. In these columns, I look at the per capita project-specific spending in a given village within a district, separating aid into World Bank aid, Chinese aid, and other aid, for the period from 2000 to 2014. I find large and significant decreases in per capita spending in villages in new districts from all three aid sources, and in parent districts for World Bank aid. Initially, these results may be surprising. I would suggest that there are two potential explanations for these findings. It is possible that these funding choices could be deliberate selection, in which new districts are deliberately avoided by donors; this theory seems fairly unlikely, however, particularly given the attention paid to new districts. Another, potentially more likely option, is that new districts may have lower quality administration, particularly in the first few years after the split. Concerns about

[8]The difference between the increase in per capita discretionary funding for parent and new districts is statistically insignificant, altough only just, with a p-value of 0.13.

district administration and challenges of implementation, elite capture, or patronage, may lead donors to be less likely to spend in these new districts. In other words, these districts receive less funding not because they are new per se, but because they may (be perceived to) have lower quality local governments.

Given these results – that per capita government spending is roughly equal between parent and new districts, but aid revenues decrease – one might expect that new district creation in Uganda in 2010 actually decreased service quality, for example by redirecting funds from service delivery towards administrative costs. I utilize primary school-level data to explore this directly.

In Table 3.3, I look at primary school-level effects of being located in a split district, whether a parent or new district. I also examine per-capita education spending from the government budget, and see no significant post-split changes, suggesting equivalent per-capita spending between parent and new districts in the post-period.

Despite these results, evidence suggests that several key indicators of school quality improve. Specifically, I see that the ratio of teachers per hundred students increases slightly in both new and parent districts; the change is equivalent to going from 90 students per teacher to 88 teachers per student, and is highly statistically significant. There are no statistically significant changes in the number of schools per students, suggesting no differential infrastructure investment; however, the number of rooms per student increases in parent but not new districts, suggesting that parent districts receive some infrastructure investment. However, there is a small increase in the proportion of students with adequate space in new districts, which does not occur in old districts.

These results suggest that while new districts are neither more nor less likely to receive bulky investments – i.e. the building of new schools – they are likelier to receive personnel and equipment investments in the six years following the creation of new distrists, though the magnitude of the changes is relatively small. Parent districts also see increases in their teacher-student ratios, but benefit as well from some improvements in infrastructure.

Appendix Figure A.3 also shows that, generally, new and parent districts both have lower ratios of schools, teachers and classrooms to students, as well as a lower proportion of students with adequate space; these differences are true prior to district splits, and persist after them. Teacher-student ratios, in particular, are lower in new districts than parent districts prior to splits, and equalize afterwards.

These findings suggest that smaller districts may provide better services to those who live there, but also continue to suggest that the mechanisms are different for parent and new districts. I next look at longer-term and broader meaures of well-being to assess whether overall development appears to improve as a result of the splitting of districts.

In Table 3.4, I look at a variety of indicators for overall well-being, including night-time lights, measures of hunger, infratsructure, assets, and service access. In general, these measures represent medium to long-term development indicators of some importance, representing a variety of ways of measuring quality of life and economic development.

The evidence in these tables suggests that the splitting of parent districts into parent and new districts seems to benefit parent districts in measures of longer-term development, but may disadvantage new districts, consistent with the theory of decreased short-run administrative capacity making difficult more complex investments in development. Specifically, new districts see a relative decrease in performance in nighttime lights and access to electricity, as well as a general measure of well-being (focused on adequate clothing, blanket and hunger) and an asset index. Parent districts, now smaller, see improvements in their NDVI indices, suggesting increased farming activity, though they alo see relative decreases in asset indices relative to non-splitting districts.

There are some exceptions – parent districts have slightly lower rates of access to safe water, and new districts see a slight, just significant increase in various types of employment – but on the whole, new districts seem to be disadvantaged in terms of more substantive developmen by their splits, whereas parent districts likely benefit, but at least not harmed.

In Table 3.5, I assess the effects of district creation on both local tax revenue and central government tax revenue collected from local sources. I find no particular change in per capita local tax revenue; recall that the imputation process means that per capita revenue is assumed to be equivalent in parent and new districts prior to a split, suggesting that the ratios remain roughly even in the post-period. These results, like the lack of an electoral gain, do not particularly suggest a mechanism by which local governments are rewarded for the creation of new districts.

I also look at several measures of the central tax process. First, I look at the percent of individuals in a given district, relative to population, who registered with a tax payer identification number in a given year, or cumulatively. On average, in a given non-split district, .15% of the population registers in a given year, and .66% of the population in total has registered. By the end of the sample, something like 2% of the population has registered in an average district. These rates are low, but not unexpectedly so; the average Ugandan needs a taxpayer identification number only if they themselves have a large source of income or are working for a formal employer required to pay taxes, both of which are fairly uncommon; no children have taxpayer identification numbers, further reducing the percent calculations.

Whether in terms of new registrants or total registrants, the data suggests that rates of registration in parent districts are significantly higher than in non-split districts and new districts, and that the rates of registration in new districts are significantly lower than in non-split districts and parent districts. There are several potential explanations for these findings, but the most likely is further evidence that the splitting of districts benefited parent districts in longer-term economic outcomes, but not new districts. This interpretation is reasonable chiefly because the increase likely represents an underlying increase in the proportion of individuals who need to pay taxes.

I also look at the proportion of those who pay one of Uganda's individual taxes relative to the proportion of registrants. Note that this dataset is first available in 2013, and so is not a true difference-in-differences, but rather a simple comparison of individual tax payment rates between parent, new and non-split districts in the post period. I find that individuals in parent and non-split districts pay individual taxes at roughly equivalent rates relative to registration, but that individuals in new districts pay individual taxes at a much lower rate. This finding may be evidence of differences in opinion regarding tax compliance, but is also potentially further evidence of lower economic activity in new districts; individual taxes apply only above a certain annual threshold for earnings, which many businesses in the more remote parts of Uganda do not reach. As such, this difference is plausibly further evidence of a decrease in economic development among new districts.

I am able to utilize the nation-wide coverage of the GeoQuery database to look at the distributional effects of district creation in Table 3.6. In this table, the outcomes are not the levels of the development or funding indicators (including nighttime lights, vegetation indices, and project aid from various donors), but rather the post-2010 district level annual standard deviation of each measure. This analysis is made possible by two particular qualities of the underlying data sources: first, that they are sub-district, at the level of villages, and second, that they have universal coverage, meaning there are no concerns about representativeness in the aggregation.

I find statistically significant decreases in the standard deviation of aid in particular, relative to non-split districts but not relative to parent districts. Recall from Table 3.2 that on the whole, aid projects are less common in villages in new districts, relative to both parent and non-split districts. These results suggest that what aid projects are undertaken in these districts have wider coverage than those in parent or non-split districts. It is difficult to assess from this table whether these differences indicate a greater taste for redistribution or equity in new districts specifically, or whether more broadly implemented projects may be more desirable to donors in a context of relatively low local governmental capacity. Nonetheless, these results are further evidence of differences between new districts and both parent and non-split districts.

It is worth noting that the universal coverage of schools makes a similar analysis possible for educational inputs; the results can be found in Appendix Table C.1. There is a slight increase in the standard deviation of safe water in parent districts, but no other statistically significant results. Although it is always difficult to interpet a null result, the lack of changes in equity in new districts is consistent with the results on project-level financing coming not from increased governmental focus on equity, but rather an explanation like different project types making more sense for lower-capacity local governments.

I look at the effects of district creation on political support and voter behavior in Table 3.7. Unlike prior studies, I do not necessarily find evidence of electoral rewards for the creation of new districts in Uganda, whether at the presidential, parliamentary or local level. There is a potential suggestion of gains in electoral support for the president in new districts, but the results are just shy of statistical significance and fairly small in magnitude, a two percentage point increase relative to a non-splitting district mean of 67% incumbent support. There is some evidence of increased participation in elections in new districts, consistent with a potential feeling of empowerment by local actors following a succesful lobbying of the govenrment on a desired policy outcome.

In Appendix Table C.2, I also assess the possibility of short vs. medium-term political gains resulting from district creation, separating out effects in the period from 2011 to 2015 and 2015 onwards. I find suggestive evidence that there may be temporary gains in support for the central government in the new district in the period immediately following district creation, but no changes in support in the short or medium run in the parent district. I also find some suggestive evidence of medium-run increases in support for President Museveni among new districts, and positive – though not significant – coefficients for new distritcs in the short-run and parent districts in the medium-run.

Overall, however, these results are far from the evidence of increases in electoral support found by Grossman and Lewis (2014). These discrepancies may be due to the fact that this study focuses on a later period, specifically the district splits which occured in 2009 and 2010. It may be that only earlier splits benefited from greater electoral rewards than later ones, potentially consistent with their broader findings of that study on diminishing marginal returns to district proliferation overall.

3.7 Conclusion

In this paper, I explore the effects of administrative unit proliferation in Uganda, specifically looking at the effects of the creation of thirty three new districts in 2009 and 2010, which increased the number of districts in Uganda from seventy nine to one hundred and twelve.

More broadly, it is important to contextualize work on any individual instance of new administrative unit creation in several ways. All unit splits are not created equal; the current size of existing units and the overall level of local development may both matter considerably when thinking about a specific event. As well, what is meant by the creation of a new administrative unit may vary; efforts to analyze the creation of new units are made challenging if other policies or increases in attention go alongside the creation.

By using a combination of uniquely rich and disaggregated administrative, secondary and remote sensing data, I am able to assess the differential effects of district creation for new and parent districts relative to non-splitting districts in Uganda, using a difference-in-differences framework. Specifically, I have rich data on funding from governmental and non-governmental sources, categories of budgetary spending, educational inputs at the school-level, well-being measured from a variety of remote sensing and household survey data, taxation data, and political opinion and voting data. All data is disaggregated at least at the level of post-2010 districts, although with imputation-related caveats in the case of budgetary data.

Crucially, the richness and disaggregated nature of this data allows me to assess other changes – such as in budget or aid – which may occur alongside the creation of new districts. I find that while overall budgetary allocations do not change, the focus of budgets shifts slightly; discretionary and administative spending increases for all parent and new districts, and new districts see a slight reallocation of funding from health to water. Education inputs increase for all new and parent districts, particularly those which do not require as much governmental involvement, such as the number of teachers per student and the share of students with adequate space. I also see evidence that per capita aid spending decreases significantly in new districts relative to both parent and non-splitting districts, and that which occurs is much more spread out across the district.

On broader measures of welfare, the new and parent districts diverge futher. Parent districts, if anything, see slight relative welfare gains; new districts, however, seem to be relatively worse off on a variety of well-being measures, including economic activity as measured by nighttime lights and taxation data, and a general index of well-being focused on access to resources (food, water, etc.). Unlike other studies which geenrally looked at earlier waves of district splits, I do not see evidence of electoral rewards following the creation of new districts, potentially suggesting further evidence of lower returns to the creation of new districts.

The gains to parent districts on a variety of measures suggest that in 2009 and 2010, Uganda's districts may have been too large, such that there was the potential for welfare gains by making the districts smaller. The lack of comparable welfare gains for new districts, however, suggests that problems with district creation may have prevented new districts from achieving comparable results. The most likely culprit, and one which would also explain the shift in project aid-based resources, is lower governmental capacity following the creation of a new district. This difference is also one of the biggest between parent and new districts, as parent districts inherit existing governmental capacity, whereas new districts do not.

On the whole, the evidence suggests that the experience of newly created administrative units may be mixed. Even when smaller districts are potentially welfare-enhancing, the district creation process can lead to lower welfare and economic development if there is not sufficient governmental capacity to sustain the newly created units. These findings add an important consideration to the existing decentralization and administrative unit proliferation literatures. They suggest that context and implementation may be an even more important component of understanding such reforms than previously believed, and even may help to reconcile some of the potentially contradictory existing results on the benefits of decentralization.

Table 3.2: Resource Effects of District Creation

Panel A. Budgetary Sources					
	Per Capita		Foreign Aid		
	Total	Discretionary	World Bank	China	Other
Parent District	1.125	0.291**	-4.307	4.664	-24.617**
	(0.790)	(0.138)	(4.117)	(22.297)	(10.391)
New District	1.005	0.527**	-19.206***	-31.665**	-53.774***
	(0.854)	(0.208)	(3.753)	(13.346)	(9.968)
Observations	1221	1221	78585	78585	78585
Dataset	Budget	Budget	GeoQuery	GeoQuery	GeoQuery
Level	District	District	Village	Village	Village
Years	2003-15	2003-15	2000-14	2000-14	2000-14
Year & District FE	Yes	Yes	Yes	Yes	Yes
Subregion x Year FE	Yes	Yes	Yes	Yes	Yes
Non-Split Mean	18.38	2.02	55.31	15.90	103.41
Parent v New	0.89	0.15	0.00	0.09	0.00

Panel B. Spending Categories					
	Per Capita				
	Admin	Transport	Health	Agriculture	Water
Parent District	0.113**	0.018	0.054	-0.025	-0.009
	(0.044)	(0.072)	(0.171)	(0.108)	(0.099)
New District	0.207***	0.084	-0.340**	0.128	0.201*
	(0.054)	(0.075)	(0.159)	(0.123)	(0.114)
Observations	1221	1221	1221	1221	1221
Dataset	Budget	Budget	Budget	Budget	Budget
Level	District	District	District	District	District
Years	2003-15	2003-15	2003-15	2003-15	2003-15
Year & District FE	Yes	Yes	Yes	Yes	Yes
Subregion x Year FE	Yes	Yes	Yes	Yes	Yes
Non-Split Mean	4.47	4.07	1.51	0.59	0.63
Parent v New	0.02	0.42	0.02	0.28	0.03

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Outcomes include total government spending per capita in a given district, discretionary government spending per capita in a given district, a village's level of project aid per capita from the World Bank, China and other donors in Panel A, and government spending per capita in a given district in the administrative, transportation, health, agriculture, and water and sewage sectors in Panel B. Per capita spending is imputed based on an even distribution across a district's total population in years prior to a split. Standard errors are clustered at the district level. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Table 3.3: Education Input Effects of District Creation

	Per Capita	Ratio to 100 Students of			Prop has	
	Educ Spending	Schools	Teachers	Rooms	Adq Space	Safe Water
Parent District	0.307	0.000	0.030***	0.028**	-0.003	-0.002
	(0.394)	(0.003)	(0.009)	(0.012)	(0.003)	(0.007)
New District	-0.060	-0.003	0.024***	0.013	0.006**	0.006
	(0.388)	(0.004)	(0.009)	(0.010)	(0.003)	(0.008)
Observations	1221	1221	122173	117225	122327	131343
Dataset	Budget	MoES	MoES	MoES	MoES	MoES
Level	District	District	School	School	School	School
Years	2003-15	2006-16	2006-16	2006-16	2006-16	2006-16
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Subregion x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Non-Split Mean	3.45	0.11	1.11	0.95	0.35	0.56
Parent v New	0.37	0.49	0.61	0.29	0.01	0.37

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include per capita spending on education in a given district, the ratio of schools, teachers and classrooms per one hundred students, and the proportion of students at a given school with adequate space (defined as a desk and chair) and access to water from a safe source. Per capita spending is imputed based on an even distribution across a district's total population in years prior to a split. Standard errors are clustered at the district level for budget data and school-student ratio data, and the school level for all other regressions. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Table 3.4: Well-Being Effects of District Creation

Panel A. Development

	Village		Index of	Prop Has	
	Night Lights	NDVI Index	Local Dev	Electricity	Safe Water
Parent District	-0.004	17.535***	0.208	-0.013	-0.054***
	(0.019)	(6.171)	(0.169)	(0.010)	(0.017)
New District	-0.053***	10.203	-0.164	-0.025***	0.007
	(0.020)	(6.557)	(0.219)	(0.009)	(0.019)
Observations	73402	57665	996	20987	20994
Dataset	GeoQuery	GeoQuery	Afrobrmtr	UNPS	UNPS
Level	Village	Village	Village	Household	Household
Years	2000-13	2008-18	2008-18	2001-18	2001-18
Year FE	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes
Subregion x Year FE	Yes	Yes	Yes	Yes	Yes
Non-Split Mean	0.79	5705.41	0.08	0.11	0.69
Parent v New	0.01	0.32	0.11	0.24	0.00

Panel B. Household

	Index of				
	Well-Being	Employment	Assets	House Quality	Welfare
Parent District	0.039	0.009	-0.073*	0.012	-0.010
	(0.066)	(0.018)	(0.038)	(0.031)	(0.035)
New District	-0.164**	0.036*	-0.089**	0.041	-0.048
	(0.082)	(0.020)	(0.036)	(0.031)	(0.037)
Observations	8975	15609	18727	21002	17931
Dataset	Afrobrmtr	UNPS	UNPS	UNPS	UNPS
Level	Individual	Household	Household	Household	Household
Years	2005-18	2005-15	2005-18	2001-18	2001-15
Year FE	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes
Subregion x Year FE	Yes	Yes	Yes	Yes	Yes
Non-Split Mean	0.89	0.70	-0.01	-0.05	-0.03
Parent v New	0.02	0.23	0.72	0.44	0.38

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes in Panel A include nighttime lights measured via DSMP-OLS, the NDVI index, an index of local development, and the proportion of households with access to electricity and safe water. Outcomes in Panel B include indices of general well-being, employment, asset ownership, house quality, and general welfare. Standard errors are clustered at the village level for all outcomes. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Table 3.5: Taxation Effects of District Creation

	Per Capita	Percent of Pop		Percent of Reg
	Local Taxes	New Registration	Total Registration	Tax Paid
Parent District	0.100	0.076***	0.357***	0.002
	(0.079)	(0.024)	(0.073)	(0.147)
New District	0.136	-0.059***	-0.219***	-0.792***
	(0.090)	(0.022)	(0.066)	(0.133)
Observations	1332	1443	1443	666
Dataset	Budget	URA	URA	URA
Level	District	District	District	District
Years	2003-15	2006-18	2006-18	2013-18
Year FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	No
Subregion x Year FE	Yes	Yes	Yes	Yes
Non-Split Mean	1.04	0.15	0.66	1.88
Parent v New	0.44	0.00	0.00	0.00

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include per capita collection of local taxes in a given district, the percent of new tax identification registrations in a given district relative to its population in a given year, the percent of total individuals registered for tax purposes relative to a district's population in a given year, and the proportion of registrants who paid one of Uganda's individual taxes in a given year. Per capita estimations are imputed based on an even distribution across a district's total population in years prior to a split. Standard errors are clustered at the district level for all regressions. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects except for the regression on tax payment, which does not include district fixed effects.

Table 3.6: Distributionary Effects of District Creation

	Village		Foreign Aid		
	Nighttime Lights	NDVI Index	World Bank	China	Other
Parent District	0.005	-6.044	-14.232	57.881	-39.876
	(0.049)	(12.398)	(20.804)	(102.564)	(37.450)
New District	-0.055	-13.044	-40.704**	-108.043	-66.865**
	(0.044)	(11.392)	(18.779)	(92.578)	(33.804)
Observations	1554	1221	1665	1665	1665
Dataset	GeoQuery	GeoQuery	GeoQuery	GeoQuery	GeoQuery
Level	District SD	District SD	District SD	District SD	District SD
Years	2000-13	2008-18	2000-14	2000-14	2000-14
Year FE	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes
Subregion x Year FE	Yes	Yes	Yes	Yes	Yes
Non-Split Mean	1.04	395.78	78.64	40.69	118.19
Parent v New	0.26	0.62	0.24	0.14	0.51

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include district-level standard deviation calculations for village-level data on nighttime lights measured via DSMP-OLS, the NDVI index, and village-level allocations of project aid from the World Bank, China, and other donors. Standard errors are clustered at the district level for all regressions. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Table 3.7: Political Effects of District Creation

	Support for		Voting		NRM Won	
	Central Gov	Local Gov	Turnout	Pres Share	MP Seat	SC Chair
Parent District	0.019	0.022	-0.004	-0.001	0.007	-0.039
	(0.056)	(0.058)	(0.008)	(0.012)	(0.105)	(0.043)
New District	0.054	0.026	0.017*	0.023	0.104	-0.009
	(0.084)	(0.079)	(0.010)	(0.014)	(0.107)	(0.047)
Observations	8975	8971	333	333	678	3544
Dataset	Afrobrmtr	Afrobrmtr	Elections	Elections	Elections	Elections
Level	Individual	Individual	District	District	Const	Subcounty
Years	2005-18	2005-18	2006-16	2006-16	2006-16	2006-16
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Subregion x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Non-Split Mean	0.89	0.89	0.66	0.67	0.76	0.72
Parent v New	0.68	0.96	0.08	0.11	0.45	0.58

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include indices capturing levels of support for central and local governments, turnout relative to potential voters in presidential elections, the share of presidential election support for President Museveni, and whether a given parliamentary seat or subcounty chair seat was won by the National Resistance Movement. Standard errors are clustered at the village level for Afrobarometer data, the district level for presidential elections data, and the constituency and subcounty level, respectively, for other election data. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Bibliography

Acemoglu, Daron, Camilo García-Jimeno, and James A Robinson (2015). "State Capacity and Economic Development: A Network Approach". In: *American Economic Review* 105.8, pp. 2364–2409.

Acemoglu, Daron, Tristan Reed, and James A. Robinson (2014). "Chiefs: Economic development and elite control of civil society in Sierra Leone". In: *Journal of Political Economy* 122.2, pp. 319–368. ISSN: 1537534X. DOI: 10.1086/674988.

Afrobarometer Data (2005). *Uganda, Round 3*. Accessed at http://www.afrobarometer.org/data/uganda-round-3-data-2005.

— (2008). *Uganda, Round 4*. Accessed at http://www.afrobarometer.org/data/uganda-round-4-data-2008.

— (2012). *Uganda, Round 5*. Accessed at http://www.afrobarometer.org/data/uganda-round-5-data-updated-july-2015.

— (2014). *Uganda, Round 6*.

— (2015). *Uganda, Round 6*. Accessed at https://afrobarometer.org/data/uganda-round-6-data-2015.

— (2018). *Uganda, Round 7*. Accessed at https://afrobarometer.org/data/uganda-round-7-data.

AidData (2016). *UgandaAIMS_GeocodedResearchRelease_Level1_v1.4.1*. Williamsburg, VA and Washington, DC: AidData.

— (2017). *WorldBank_GeocodedResearchRelease_Level1_v1.4.2 geocoded dataset*. Williamsburg, VA and Washington, DC: AidData. Accessed in 2018. http://aiddata.org/research-datasets.

AidData (2018). *Uganda_Budget_Health_GeocodedResearchRelease_Level1_v1.0.0*. Williamsburg, VA: AidData.

Akin, John, Paul Hutchinson, and Koleman Strumpf (2005). "Decentralisation and Government Provision of Public Goods: The Public Health Sector in Uganda". In: *Journal of Development Studies* 41.8, pp. 1417–1443.

Alesina, Alberto, Reza Baqir, and William Easterly (1999). "Public Goods and Ethnic Divisions". In: *Quarterly Journal of Economics* 114.4, pp. 1243–1284.

Alesina, Alberto and David Dollar (2000). "Who Gives Foreign Aid to Whom and Why?" In: *Journal of Economic Growth* 5, pp. 33–63.

Ali, Merima, Odd Helge Fjeldstad, and Ingrid Hoem Sjursen (2014). "To pay or not to pay? Citizens' attitudes toward taxation in Kenya, Tanzania, Uganda, and South Africa". In: *World Development* 64.March 2013, pp. 828–842. ISSN: 0305750X. DOI: 10.1016/j.worlddev.2014.07.006. URL: http://dx.doi.org/10.1016/j.worlddev.2014.07.006.

Allingham, Michael G and Agnar Sandmo (1972). "Income Tax Evasion: A Theoretical Analysis". In: *Jounal of Public Economics* 1, pp. 323–338.

Anderson, Michael (2008). "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects". In: *Journal of the American Statistical Association* 103.484, pp. 1481–1495.

Ariel, Barak (2012). "Deterrence and moral persuasion effects on corporate tax compliance: Findings from a randomized controlled trial". In: *Criminology* 50.1, pp. 27–69. ISSN: 00111384. DOI: 10.1111/j.1745-9125.2011.00256.x.

Arndt, Channing, Sam Jones, and Finn Tarp (2016). "What is the Aggregate Economic Rate of Return to Foreign Aid". In: *The World Bank Economic Review* 30.3, pp. 446–474.

Asher, Sam, Karan Nagpal, and Paul Novosad (2018). *The Cost of Distance: Geography and Governance in Rural India*. Unpublised manuscript.

B, Rubongoya J (2007). *Regime hegemony in Museveni's Uganda: Pax Musevenica*. New York: Palgrave Macmillian.

Bardhan, Pranab and Dilip Mookherjee (2006). "Decentralisation and Accountability in Infrastructure Delivery in Developing Countries". In: *Economic Journal* 116.508, pp. 101–127.

Barr, Abigail and Marcel Fafchamps (2006). "A client-community assessment of the NGO sector in Uganda". In: *The Journal of Development Studies*.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2011). "LASSO Methods for Gaussian Instrumental Variables Models". Working Paper.

BenYishay, Ariel et al. (2017). "Geocoding Afrobarometer Rounds 1-6: Methodology and Data Quality". AidData. Available online at http://geo.aiddata.org.

Benzarti, Youssef (2015). "How taxing is tax filing? Leaving money on the table because of hassle costs." Mimeo, University of California, Los Angeles.

Bergolo, Marcelo et al. (2019). *Tax Audits as Scarecrows. Evidence from a Large-Scale Field Experiment*. IZA DP No. 12335.

Berthélemy, Jean-Claude and Ariane Tichit (2004). "Bilateral donors' aid allocation decisions - a three-dimensional panel analysis". In: *International Review of Economics and Finance* 13, pp. 253–274.

Besley, Timothy and Torsten Persson (2013). *Taxation and Development*. Vol. 5. Elsevier B.V., pp. 51–110. ISBN: 9780444537591. DOI: 10.1016/B978-0-444-53759-1.00002-9. URL: http://dx.doi.org/10.1016/B978-0-444-53759-1.00002-9.

— (2014). "Why do developing countries tax so little?" In: *Journal of Economic Perspectives* 28.4, pp. 99–120. ISSN: 08953309. DOI: 10.1257/jep.28.4.99.

Bielefeld, Wolfgang and James C Murdoch (2004). "The Locations of Nonprofit Organizations and Their For-Profit Counterparts: An Exploratory Analysis". In: *Nonprofit and Voluntary Sector Quarterly* 33.2, pp. 221–246.

Bluhm, Richard et al. (2018). *Connective Financing: Chinese Infrastructure Projects and the Diffusion of Economic Activity in Developing Countries*. Tech. rep. AidData Working Paper #64. Williamsburg, VA: AidData at William & Mary.

Blumenthal, Marsha, Charles Christian, and Joel Slemrod (2001). "Do normative appeals affect tax compliance? Evidence from a controlled experiment in Minnesota". In: *National Tax Journal* 54.1, pp. 125–136. ISSN: 00280283. DOI: `10.17310/ntj.2001.1.06`.

Bodea, Cristina and Adrienne LeBas (2013). "The Origins of Social Contracts: Attitudes toward Taxation in Urban Nigeria". CSAE working paper WPS/2013-02.Oxford: Centre for the study of African Economies.

Bouguen, Adrien et al. (2018). "Using RCTs to Estimate Long-Run Impacts in Development Economics". NBER Woking Paper 25356.

Burnside, Craig and David Dollar (2000). "Aid, Policies and Growth". In: *The American Economic Review* 90.4, pp. 847–868.

Bwire, Thomas, Tim Lloyd, and Oliver Morrissey (2017). "Fiscal Reforms and the Fiscal Effects of Aid in Uganda". In: *The Journal of Development Studies* 53.7, pp. 1019–1036.

Canavire, Gustavo et al. (2005). *Assessing the Allocation of Aid: Developmental Concerns and the Self-Interest of Donors*. Tech. rep. Kiel Working Paper No. 1253.

Castro, Lucio and Carlos Scartascini (2013). "Tax Compliance and Enforcement in the Pampas: Evidence from a Field Experiment". IDB Working Paper No. IDB-WP-472.

Chetty, Raj, Adam Looney, and Kory Kroft (2009). "Salience and taxation: Theory and evidence". In: *American Economic Review* 99.4, pp. 1145–1177. ISSN: 00028282. DOI: `10.1257/aer.99.4.1145`.

Christian, Paul and Christopher B Barrett (2017). *Revisiting the Effect of Food Aid on Conflict: A Methodological Caution*. Tech. rep. World Bank Policy Research Working Paper 8171.

CIESIN (2016). *Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision UN WPP Country Totals*. Palides, NY: NASA Socioeconomic Data and Applications Center (SEDAC). http://dx.doi.org/10.7927/H4HX19NJ.

Donaldson, Dave and Adam Storeygard (2016). "The view from above: Applications of satellite data in economics". In: *Journal of Economic Perspectives* 30.4, pp. 171–198. ISSN: 08953309. DOI: `10.1257/jep.30.4.171`.

Dreher, Axel and Steffen Lohmann (2015). "Aid and growth at the regional level". In: *Oxford Review of Economic Policy* 31.3-4, pp. 420–446.

Dreher, Axel et al. (2017). *Aid, China, and Growth: Evidence from a New Global Development Finance Dataset*. Tech. rep. AidData Working Paper 46.

Easterly, William (2002). "The cartel of good intentions: The problem of bureaucracy in foreign aid". In: *The Journal of Policy Reform* 5.4, pp. 223–250.

Easterly, William, Ross Levine, and David Roodman (2004). "Aid, Policies, and Growth: Comment". In: *American Economic REview* 94.3, pp. 774–780.

Eissa, Nada and Andrew Zeitlin (2014). "Using mobile technologies to increase VAT compliance in Rwanda". Mimeo, McCourt School of Public Policy, Georgetown University.

Electoral Commission of Uganda (2021). *Elections and Results*. Available online at https://www.ec.or.ug/?q=by-elections.

Elvidge, Christopher et al. (2017). "VIIRS night-time lights". In: *International Jounal of Remote Sensing*, pp. 1–20. ISSN: 10.1080/01431161.2017.1342050.

Elvidge, Christopher D et al. (2014). "National trends in satellite-observed lighting". In: *Global urban monitoring and assessment through earth observation* 23, pp. 97–118.

Fagernäs, Sonja and John Roberts (2004a). *Fiscal Impact of Aid: A Survey of Issues and Synthesis of Country Studies of Malawi, Uganda and Zambia*. Tech. rep. ESAU Working Paper 11.

— (2004b). *The Fiscal Effects of Aid in Uganda*. Tech. rep. ESAU Working Paper 0.

— (2004c). *The Fiscal Effects of Aid in Zambia*. Tech. rep. ESAU Working Paper 10.

Fagernäs, Sonja and Cedrik Schurich (2004). *The Fiscal Effects of Aid in Malawi*. Tech. rep. ESAU Working Paper 7.

Faguet, Jean-Paul (2012). *Decentralization and Popular Democracy: Governance from Below in Bolivia*. Ann Arbor: University of Michigan Press.

Franco-Rodriguez, Susana (2000). "Recent Developments in Fiscal Response with an Application to Costa Rica". In: *Journal of International Development* 12, pp. 429–441.

Franco-Rodriguez, Susana and Oliver Morrissey (1998). "Aid and the Public Sector in Pakistan: Evidence with Endogenous Aid". In: *World Development* 26.7, pp. 1241–1250.

Galiani, Sebastian, Paul Gertler, and Ernesto Schargodsky (2008). "School Decentralization: Helping the Good Get Better, but Leaving the Poor Behind". In: *Journal of Public Economics* 92.10-11, pp. 2106–2120.

Gates, Scott and Anke Hoeffler (2004). *Global Aid Allocation: Are Nordic Donors Different?* Tech. rep. CSAE WPS/2004-34.

Gehring, Kai, Lennart Kaplan, and Melvin H L Wong (2017). "Aid and conflict at the local level - Mechanisms and causality".

Gomanee, Karuna et al. (2005). "Aid, Government Expenditure and Aggregate Welfare". In: *World Development* 33.3, pp. 355–370.

Goodman, Seth et al. (2019). "GeoQuery: Integrating HPC systems and public web-based geospatial data tools". In: *Computers & Geosciences* 122, pp. 103–112.

Gottlieb, Jessica, Adrienne LeBas, and Nonso Obikili (2018). *Formalization, Tax Appeals, and Social Intermediaries in Lagos, Nigeria*. EGAP registry, available at: https://osf.io/gx9s7.

Gottlieb, Jessica et al. (2019). "The Effect of Administrative Unit Creation on Electoral Behavior: Evidence from Senegal". In: *Journal of Politicss* 81.2, pp. 631–647.

Green, Elliot D (2010). "Patronage, District Creation, and Reform in Uganda". In: *Studies in Comparative International Development* 45.1, pp. 83–103.

— (2015). "Decentralization and Development in Contemporary Uganda". In: *Regional & Federal Studies* 25.5, pp. 491–508.

Grossman, Guy and Janet I Lewis (2014). "Administrative Unit Proliferation". In: *American Political Science Association* 108.1, pp. 196–217.

Grossman, Guy, Jan H. Pierskalla, and Emma Boswell Dean (2017). "Government Fragmentation and Public Goods Provision". In: *The Journal of Politics* 79.3, pp. 823–840.

Habibi, Nadir et al. (2003). "Decentralization and Human Development in Argentina". In: *Journal of Human Development* 4.1, pp. 73–101.

Hansen, Henrik and Finn Tarp (2000). "Policy Arena: Aid Effectiveness Disrupted". In: *Journal of International Development* 12, pp. 375–398.

— (2001). "Aid and growth regressions". In: *Journal of Development Economics* 64, pp. 547–570.

Hassan, Mai (2016). "A State of Change: District Creation in Kenya after the Beginning of Multi-Party Elections". In: *Political Research Quarterly* 69.3, pp. 510–21.

Heller, Peter S (1975). "A Model of Public Fiscal Behavior in Developing Countries: Aid, Investment and Taxation". In: *The American Economic Review* 65.3, pp. 429–445.

International Development Association (2018). *Uganda: Support to Municipal Infrastructure Development Program Additional Financing.* Tech. rep. 122547-UG. World Bank.

International Monetary Fund (2020). *Government Finance Statistics.* https://data.imf.org/?sk=a0867067-d23c-4ebc-ad23-d3b015045405.

Iqbal, Zafar (1997). "Foreign Aid and the Public Sector: A Model of Fiscal Behavior in Pakistan". In: *The Pakistan Development Review* 36.2, pp. 115–129.

Isopi, Alessia and George Mavrotas (2009). "Development Aid: A Fresh Look". In: ed. by George Mavrotas and Mark McGillivray. United Nations University. Chap. Aid Alloca, pp. 114–157.

JLoS (2013). *GIS Map Locations of JLOS Institutions.* Uganda Justice, Law and Order Sector. Available at: https://jlos.go.ug/index.php/about-jlos/projects/justice-information-systems/gis.

Kettle, Stewart et al. (2016). "Behavioral Interventions in Tax Compliance : Evidence from Guatemala". Policy Research Working Paper;No. 7690. World Bank, Washington, DC. World Bank. https://openknowledge.worldbank.org/handle/10986/24530 License: CC BY 3.0 IGO.

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

— (2019). "Making Moves Matter: Experimental Evidence on incentivizing Bureaucrats through Performance-based Transfers". In: *American Economic Review* 109.1, pp. 237–70.

Kimura, Ehito (2013). *Political Change and Territoriality in Indonesia: Provincial Proliferation.* New York: Routledge.

Kirchler, Erich, Erik Hoelzl, and Ingid Wahl (2008). "Enforced versus voluntary tax compliance: The slippery slope framework". In: *Journal of Economic Psychology* 29.2, pp. 210–225.

Kis-Katos, Krisztina and Bambang Suharnoko Sjahir (2014). "The Impact of Fiscal and Political Decentralization on Local Public Investments in Indonesia". Unpublished manuscript.

Kleven, Henrik Jacobsen et al. (2011). "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark". In: *Econometrica* 79.3, pp. 651–692. ISSN: 0012-9682. DOI: 10.3982/ecta9113.

Koch, Dirk-Jan (2009). "Development Aid: A Fresh Look". In: ed. by George Mavrotas and Mark McGillivray. United Nations University. Chap. Blind Spot, pp. 26–57.

Lewis, Janet I (2014). "When Decentralization Leads to Recentralization: Subnational State Transformation in Uganda". In: *Regional & Federal Studies* 24.5, pp. 571–588.

Lloyd, Tim et al. (2009). "Development Aid: A Fresh Look". In: ed. by George Mavrotas and Mark McGillivray. United Nations University. Chap. The Fiscal, pp. 158–179.

Luttmer, Erzo F.P. and Monica Singhal (2014). "Tax morale". In: *Journal of Economic Perspectives* 28.4, pp. 149–168. ISSN: 08953309. DOI: 10.1257/jep.28.4.149.

Malesky, Edmund (2009). "Gerrymandering - Vietnamese Style: Escaping the Partial Reform Equilbrium in a Nondemocratic Regime". In: *Journal of Politics* 71.1, pp. 132–159.

Martin, Lucy (2016). "Taxation, Loss Aversion, and Accountability: Theory and Experimental Evidence for Taxation's Effect on Citizen Behavior".

Martins, Pedo M G (2010). *Fiscal Dynamics in Ethiopia: The Cointegrated VAR Model with Quarterly Data*. Tech. rep. University of Sussex Economics Department Working Paper Series No. 9-2010.

Mascagani, Giulia and Emilija Timmis (2014). *Fisacl Effects of Aid in Ethiopia: Evidence from CVAR applications*. Tech. rep. CREDIT Research Paper No. 14/06.

Mascagni, Giulia (2018). "From the Lab To the Field: a Review of Tax Experiments". In: *Journal of Economic Surveys* 32.2, pp. 273–301. ISSN: 14676419. DOI: 10.1111/joes.12201.

Mascagni, Giulia, Christopher Nell, and Nara Monkam (2017). "One Size Does Not Fit All: A Field Experiment on the Drivers of Tax Compliance and Delivery Methods in Rwanda". ICTD Working Paper 58. Brighton, UK.

Maskin, Eric, Yingyi Qian, and Chenggang Xu (2000). "Incentives, Information, and Organizational Form". In: *Review of Economic Studies* 67.2, pp. 359–378.

Mavrotas, George and Bazoumana Ouattara (2006). "Aid Disaggregation and the Public Sector in Aid-Recipient Economies: Some Evidence from Cote D'Ivoire". In: *Review of Development Economics* 10.3, pp. 434–451.

McGillivray, Mark and Oliver Morrissey (2001). *A Review of Evidence on the Fiscal Effects of Aid*. Tech. rep. CREDIT Research Paper No. 01/13.

McKerchar, Margaret and Chris Evans (2009). "Sustaining growth in developingeconomies through improved taxpayer compliance: Challenges forpolicy makers and revenue authorities." In: *eJournal of Tax Research* 7, pp. 171–201.

Meiselman, Ben S (2018). "Ghostbusting in Detroit: Evidence on nonfilers from a controlled field experiment". In: *Jounal of Public Economics* 158, pp. 180–193.

Miguel, Edward and Mary Kay Gugerty (2005). "Ethnic Diversity, Social Sanctions, and Public Goods in Kenya". In: *Journal of Public Economics* 89.11-12, pp. 2325–2368.

MoES (2016). *Uganda Education Management and Information System*. Ministry of Education and Sports.

MoFPED (2018). *Know Your Budget 2018: A Citizens Guide to the Budget of FY 2018/19*. Ministry of Finance, Planning and Economic Development.

MoFPED (2020). *Budget Library*. Ministry of Finance, Planning and Economic Development. Available at: http://budget.go.ug/library/.

MoH (2016). *Uganda Health Management and Information System*. Ministry of Health.

Morrissey, Oliver (2015). "Aid and Government Fiscal Behavior: Assessing Recent Evidence". In: *World Development* 69, pp. 98–105.

Murphy, Kevin and Robert Topel (1985). "Estimation and Inference in Two-Step Econometric Models". In: *Journal of Business & Economic Statistics* 3.4, pp. 370–379.

Neumayer, Eric (2003a). "The Determinants of Aid Allocation by Regional Multilateral Development Banks and United Nations Agencies". In: *International Studies Quarterly* 47, pp. 101–122.

— (2003b). "What Factors Determine the Allocation of Aid by Arab Countries and Multilateral Agencies?" In: *Journal of Development Studies* 39.4, pp. 134–147.

Nunn, Nathan and Nancy Qian (2014). "US Food Aid and Civil Conflict". In: *American Economic Review* 104.6, pp. 1630–1666.

Ortega, Daniel, Lucas Ronconi, and Pablo Sanguinetti (2016). "Reciprocity and Willingness to Pay Taxes: Evidence from a Survey Experiment in Latin America". In: *Economia* 16.2, pp. 55–87.

Osei, Robert, Oliver Morrissey, and Tim Lloyd (2005). "The Fiscal Effects of Aid in Ghana". In: *Journal of International Development* 17, pp. 1037–1053.

Pedelty, JA, S Devadiga, and E Masouka (2007). *Generating a Long-term Land Data Record from the AVHRR and MODIS Instruments*. Proceedings of IGARRS 2007, pp. 10211025. Institute of Electricaland Electronics Engineers, NY, USA. NY, USA.

Petrin, Amil and Kenneth Train (2003). *Omitted Product Attributes in Discrete Choice Models*. NBER Working Paper 9452. http://www.nber.org/papers/w9452.

Pierskalla, Jan H (2016). "Splitting the Difference? The Politics of District Creation in Indonesia". In: *Comparative Politics* 48.2, pp. 249–268.

Pomeranz, Dina (2015). "No taxation without information: Deterrence and self-enforcement in the value added tax". In: *American Economic Review* 105.8, pp. 2539–2569. ISSN: 00028282. DOI: 10.1257/aer.20130393.

Pomeranz, Dina, Cristobal Marshall, and Pamela Castellon (2014). "Randomized Tax Enforcement Messages: A Policy Tool for Improving Audit Strategies". In: *Tax Administration Review* 36, pp. 1–21. DOI: 10.1039/C2EE02542F.

Pomeranz, Dina and José Vila-Belda (2019). "Taking State-Capacity Research to the Field: Insights from Collaborations with Tax Authorities". In: *Annual Review of Economics* 11, pp. 755–81.

Rajan, Raghuram and Arvind Subramanian (2008). "Aid and Growth: What Does the Cross-Country Evidence Really Show?" In: *Review of Economics and Statistics* 90, pp. 643–665.

Remmer, Karen L (2004). "Does Foreign Aid Promote the Expansion of Government". In: *American Journal of Political Science* 48.1, pp. 77–92.

Slemrod, Joel (2018). "Tax Compliance and Enforcement". NBER Working Paper 24799. URL: `http://www.nber.org/papers/w24799`.

Slemrod, Joel, Marsha Blumenthal, and Charles Christian (2001). "Taxpayer response to an increased probability of audit: Evidence from a controlled experiment in Minnesota". In: *Journal of Public Economics* 79.3, pp. 455–483. ISSN: 00472727. DOI: `10.1016/S0047-2727(99)00107-3`.

Snavely, Keith (1990). "Governmental policies to reduce tax evasion: coerced behavior versus services and values development". In: *Policy Sciences* 23, pp. 57–72.

Svensson, Jakob (2003). "Why conditional aid does not work and what can be done about it?" In: *Journal of Development Economics* 70, pp. 381–402.

The World Bank, World Development Indicators (2021a). *Net ODA received (% of central government expense) [Data file]*. Retrieved from https://data.worldbank.org/indicator/DT.ODA.ODAT.XP.ZS?locations=UG.

— (2021b). *Net ODA Received per capita [Data file]*. Retrieved from https://data.worldbank.org/indicator/DT.ODA.ODAT.PC.ZS?locations=UG.

UBOS (2017). *Uganda National Household Survey*. Kampala, Uganda.

Uganda Bureau of Statistics (2009). *Uganda-National Panel Survey (UNPS) 2005-2009, Ref. UGA_2005-2009_UNPS_v01_M*. Dataset downloaded from https://microdata.worldbank.org/index.php/catalog/1001.

— (2010). *Uganda-National Panel Survey (UNPS) 2010-2011, Ref. UGA_2010_UNPS_v01_M*. Dataset downloaded from https://microdata.worldbank.org/index.php/catalog/2166.

— (2011). *Uganda-National Panel Survey (UNPS) 2011-2012, Ref. UGA_2011_UNPS_v01_M*. Dataset downloaded from https://microdata.worldbank.org/index.php/catalog/2059.

— (2013). *Uganda-National Panel Survey (UNPS) 2013-2014, Ref. UGA_2013_UNPS_v01_M*. Dataset downloaded from https://microdata.worldbank.org/index.php/catalog/2663.

— (2015). *Uganda-National Panel Survey (UNPS) 2015-2016, Ref. UGA_2015_UNPS_v01_M*. Dataset downloaded from https://microdata.worldbank.org/index.php/catalog/3460.

— (2018). *Uganda-National Panel Survey (UNPS) 2018-2019, Ref. UGA_2018_UNPS_v01_M*. Dataset downloaded from https://microdata.worldbank.org/index.php/catalog/3795.

URA (2018). *Taxation Data*. Uganda Revenue Authority.

Wagstaff, Adam (2011). "Fungibility and the impact of development assistance: Evidence from Vietnam's health sector". In: *Journal of Development Economics* 94, pp. 62–73.

Walle, Dominique van de and Ren Mu (2007). "Fungibility and the flypaper effect of project aid: Micro-evidence for Vietnam". In: *Journal of Development Economics* 84, pp. 667–685.

Weigel, Jonathan (2020). "The Participation Dividend of Taxation: How Citizens Engage More with the State When it Tries to Tax Them". In: *The Quarterly Journal of Economics* 135.2, pp. 1849–1903.

Weiss, D.J. et al. (2018). "A global map of travel time to cities to assess inequalities in accessibility in 2015". In: *Nature*.

World Bank (2021a). *Skills Development Project*. Retrieved from https://projects.worldbank.org/en/projects-operations/project-detail/P145309.

World Bank (2021b). *Uganda Multisectoral Food Security and Nutrition Project*. Retrieved from https://projects.worldbank.org/en/projects-operations/project-detail/P149286.

Zax, Jeffrey S (1989). “Is There a Leviathan in Your Neighborhood?” In: *American Economic Review* 79.3, pp. 560–567.

Appendix A

Appendix for "Crowd-In or Crowd-Out?: The Subnational Fiscal Response to Aid"

A.1 OLS and Reduced Form Tables

Table A.1: OLS: Revenue Impacts of Aid (1M USD)

	Current				**Lagged**			
Outcome	Coef	SE	p/q	Obs	Coef	SE	p/q	Obs
Total	-0.033	0.029	0.26/0.65	555	-0.001	0.030	0.98/0.98	555
Salary	-0.017	0.030	0.58/0.71	555	0.001	0.024	0.98/0.98	555
Non-Salary	0.002	0.006	0.71/0.71	555	0.011	0.011	0.33/0.55	555
Development	-0.018	0.006	0.00/0.03	555	-0.019	0.010	0.07/0.33	555
Unconditional	0.001	0.002	0.62/0.71	555	0.002	0.002	0.32/0.55	555

Note: 2010-2014 aid sample, district and year fixed effects. Total is the sum of Salary, Non-Salary and Development; Unconditional contains both Salary and Non-Salary items. Q-values are constructed using the Benjamini-Hochberg procedure to control for false discovery rates.

Table A.2: OLS: Revenue Impacts of Aid (1M USD)

	Current				**Lagged**			
Outcome	Coef	SE	p/q	Obs	Coef	SE	p/q	Obs
Education	-0.042	0.032	0.19/0.49	555	-0.014	0.022	0.53/0.88	555
Health	-0.001	0.008	0.89/0.89	555	-0.012	0.005	0.02/0.08	555
Agriculture	-0.002	0.004	0.61/0.80	555	-0.002	0.003	0.49/0.88	555
Transport	-0.004	0.002	0.09/0.43	555	0.002	0.007	0.77/0.97	555
Water	-0.001	0.002	0.64/0.80	555	0.000	0.002	0.99/0.99	555

Note: 2010-2014 aid sample, district and year fixed effects. Q-values are constructed using the Benjamini-Hochberg procedure to control for false discovery rates.

Table A.3: OLS: Other Donor Impacts of Aid (1M USD)

	Current				**Lagged**			
Outcome	Coef	SE	p/q	Obs	Coef	SE	p/q	Obs
All Donors	0.369	0.367	0.32/0.44	555	0.452	0.089	0.00/0.00	498
Japan	-0.269	0.170	0.12/0.24	555	-0.033	0.038	0.39/0.46	498
Norway	0.004	0.006	0.52/0.61	555	-0.003	0.006	0.57/0.57	498
ADF	0.772	0.147	0.00/0.00	555	0.497	0.044	0.00/0.00	498
US	-0.001	0.006	0.87/0.87	555	-0.011	0.006	0.07/0.09	498
UK	0.001	0.000	0.00/0.00	555	0.001	0.000	0.00/0.00	498
EU	-0.063	0.042	0.14/0.24	555	-0.023	0.011	0.04/0.08	498

Note: 2010-2014 aid sample, district and year fixed effects. Q-values are constructed using the Benjamini-Hochberg procedure to control for false discovery rates.

Table A.4: OLS: Chinese Donation Impacts of Aid (1M USD)

	Current				Lagged			
Outcome	Coef	SE	p/q	Obs	Coef	SE	p/q	Obs
Other Aid	-0.083	0.074	0.26/0.55	555	0.025	0.023	0.28/0.65	498
All Aid	-0.153	0.168	0.36/0.55	555	-0.045	0.099	0.65/0.65	498
ODA-like Aid	-0.070	0.137	0.61/0.61	555	-0.070	0.096	0.47/0.65	498

Note: 2010-2014 aid sample, district and year fixed effects. Q-values are constructed using the Benjamini-Hochberg procedure to control for false discovery rates.

Table A.5: Reduced Form: Revenue Impacts of Aid (1M USD)

	Current				Lagged			
Outcome	Coef	SE	p/q	Obs	Coef	SE	p/q	Obs
Total	2.031	0.763	0.01/0.04	555	2.900	1.027	0.01/0.01	444
Salary	1.722	0.693	0.01/0.04	555	3.106	0.865	0.00/0.00	444
Non-Salary	0.122	0.176	0.49/0.70	555	-0.101	0.306	0.74/0.76	444
Development	0.151	0.438	0.73/0.73	555	-0.059	0.190	0.76/0.76	444
Unconditional	0.068	0.115	0.56/0.70	555	0.027	0.070	0.70/0.76	444

Note: 2010-2014 aid sample, district and year fixed effects. Total is the sum of Salary, Non-Salary and Development; Unconditional contains both Salary and Non-Salary items. Q-values are constructed using the Benjamini-Hochberg procedure to control for false discovery rates.

Table A.6: Reduced Form: Revenue Impacts of Aid (1M USD)

	Current				Lagged			
Outcome	Coef	SE	p/q	Obs	Coef	SE	p/q	Obs
Education	2.121	0.695	0.00/0.01	555	2.578	0.947	0.01/0.04	555
Health	0.235	0.203	0.25/0.42	555	0.568	0.252	0.03/0.06	555
Agriculture	-0.108	0.084	0.20/0.42	555	-0.118	0.072	0.11/0.18	555
Transport	-0.092	0.128	0.47/0.59	555	-0.083	0.118	0.48/0.60	555
Water	0.010	0.032	0.75/0.75	555	-0.012	0.034	0.73/0.73	555

Note: 2010-2014 aid sample, district and year fixed effects. Q-values are constructed using the Benjamini-Hochberg procedure to control for false discovery rates.

Table A.7: Reduced Form: Other Donor Impacts of Aid (1M USD)

	Current				Lagged			
Outcome	Coef	SE	p/q	Obs	Coef	SE	p/q	Obs
All Donors	-32.497	11.498	0.01/0.02	555	-9.754	12.269	0.43/0.51	444
Japan	-9.610	9.125	0.29/0.34	555	-2.176	8.983	0.81/0.81	444
Norway	-0.352	0.248	0.16/0.22	555	-0.553	0.713	0.44/0.51	444
ADF	-30.818	5.847	0.00/0.00	555	-5.101	1.313	0.00/0.00	444
US	0.345	0.386	0.37/0.38	555	0.582	0.675	0.39/0.51	444
UK	-0.047	0.030	0.13/0.22	555	0.058	0.038	0.13/0.30	444
EU	2.690	1.221	0.03/0.07	555	2.367	1.355	0.08/0.29	444

Note: 2010-2014 aid sample, district and year fixed effects. Q-values are constructed using the Benjamini-Hochberg procedure to control for false discovery rates.

Table A.8: Reduced Form: Chinese Donation Impacts of Aid (1M USD)

	Current				**Lagged**			
Outcome	Coef	SE	p/q	Obs	Coef	SE	p/q	Obs
Other Aid	-11.706	11.465	0.31/0.42	555	-6.674	6.539	0.31/0.93	444
All Aid	-14.002	11.937	0.24/0.42	555	-8.662	21.653	0.69/0.93	444
ODA Aid	-2.296	2.840	0.42/0.42	555	-1.988	21.616	0.93/0.93	444

Note: 2010-2014 aid sample, district and year fixed effects. Q-values are constructed using the Benjamini-Hochberg procedure to control for false discovery rates.

Appendix B

Appendix for "Low-Cost Tax Capacity: A Randomized Evaluation on Tax Compliance with the Uganda Revenue Authority"

B.1 Supplementary Tables and Figures

Figure A.1: Geocoding Visualization

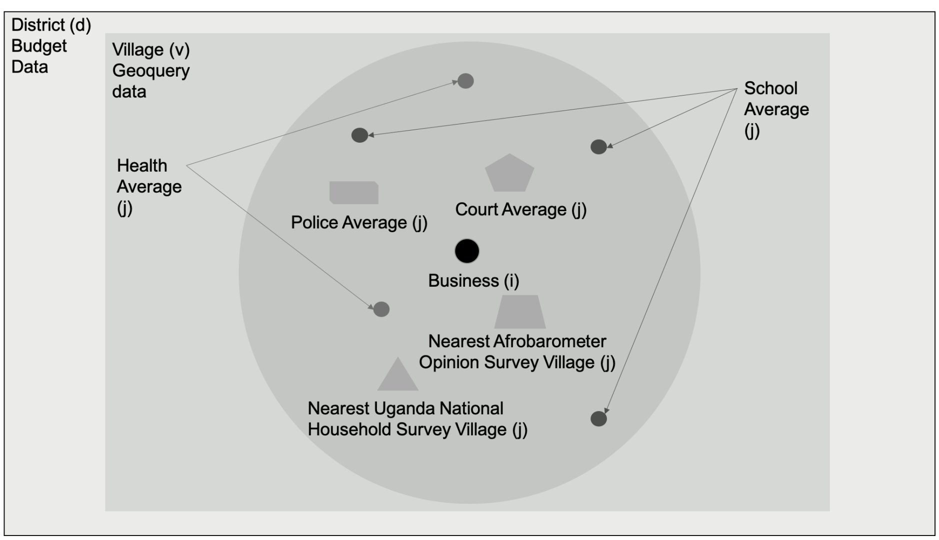


Table A.1: Treatment Effects (Conditional on Payment)

Panel A. Separate Treatments						
	Tax Amount (USD)		Presumptive		PIT	
Inform	-1.117	-0.007	2.649	2.109	-1.932	-0.792
	(0.588)	(0.997)	(0.566)	(0.636)	(0.397)	(0.716)
Encourage	1.132	1.212	-3.707	-2.345	2.646	2.492
	(0.594)	(0.541)	(0.378)	(0.556)	(0.252)	(0.257)
Enforce	-3.067	-1.716	-2.107	-2.121	-1.008	-0.049
	(0.141)	(0.384)	(0.612)	(0.601)	(0.657)	(0.982)
Inform vs. Enforce	0.332	0.376	0.244	0.234	0.681	0.732
Inform vs. Encourage	0.275	0.531	0.124	0.218	0.045	0.138
Encourage vs. Enforce	0.043	0.137	0.659	0.942	0.108	0.247
Panel B. Pooled Treatment						
	Tax Amount (USD)		Presumptive		PIT	
Any Treatment	-1.119	-0.233	-0.993	-0.788	-0.146	0.515
	(0.516)	(0.886)	(0.787)	(0.830)	(0.938)	(0.772)
Individual Controls	No	Yes	No	Yes	No	Yes
Control Mean	54.441	54.441	73.641	73.641	81.954	81.954
Observations	4833	4833	1978	1978	2917	2917

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Outcomes consist of the amount paid for any tax, the presumptive, and the personal income tax, respectively, between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding, restricted to those who paid the tax in the column heading during the time period. All regressions include strata and district fixed effects and are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table A.2: Geographic Covariates Balance Table

Variable	Control	Inform	Encourage	Enforce	p	p (T v C)
Human Capital Index	-2.1e-03	-2.1e-03	4.8e-03	3.1e-03	0.763	0.614
Education Index	-3.9e-03	-5.5e-03	-1.6e-04	6.6e-03	0.460	0.598
Health Index	-7.3e-03	6.3e-03	-8.8e-03	1.9e-03	0.277	0.367
Court Index	7.5e-03	-4.6e-03	.018	-3.0e-03	0.039	0.603
Enforcement Index	-2.1e-03	5.7e-04	2.8e-03	8.8e-03	0.682	0.438
Capacity Index	-2.1e-03	-6.4e-04	1.4e-03	1.3e-03	0.751	0.267
Urban Index	.013	-6.5e-03	.015	6.0e-03	0.078	0.302
New Services Index	-8.1e-03	-2.1e-03	-.014	-.011	0.405	0.880
N=84,955						

The first four columns represent the group-specific mean for the variables listed on the lefthand side, based on geocoded datasets on public goods and economic activity using the Google Maps method. p(All) represents a p-value from a test of joint equality for coefficients on inform, encourage and enforce; p(T v C) represents the p-value for the coefficient for a dummy variable indicating the respondent was assigned to any treatment group

Table A.3: Human Capital Heterogeneity Components

	Components		
	Education	Health	Court
Treatment	0.354**	0.356**	0.357**
	(0.046)	(0.045)	(0.045)
Component x Treatment	-0.350*	-0.075	-0.348
	(0.069)	(0.670)	(0.125)
Observations	84955	84955	84955

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. In each regression, the column title indicates the name of the index which was interacted with treatment in that specification, which include the education quality index, health quality index, and court quality index. All regressions include strata and district fixed effects, individual controls, and an index for the urbanness of the respondent's village. All regressions are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table A.4: All Human Capital Heterogeneity Components

Panel A. Education				
	Piped Water	Teach Ratio	Prop Space	Advance P4P5
Treatment	0.356**	0.355**	0.355**	0.353*
	(0.045)	(0.046)	(0.046)	(0.051)
Component x Treat	-0.147	-0.434	-0.573	-0.576
	(0.648)	(0.116)	(0.139)	(0.237)
Observations	84955	84955	84955	82587
Panel B. Health				
	Mother Visit	Baby Visit	Matrn Audit	Natal Audit
Treatment	0.373**	0.369**	0.249	0.331
	(0.043)	(0.045)	(0.395)	(0.527)
Component x Treat	0.348	0.057	0.090	0.057
	(0.159)	(0.854)	(0.865)	(0.542)
Observations	78452	78339	31627	20349
	Has ORS	Has Depo	Has Oxy	
Treatment	0.356**	0.356**	0.357**	
	(0.045)	(0.045)	(0.044)	
Component x Treat	-0.281	-1.500	0.708	
	(0.719)	(0.197)	(0.281)	
Observations	84955	84955	84955	
Panel C. Courts				
	Staff Nr	Elec	Internet	
Treatment	0.357**	0.293	0.327*	
	(0.044)	(0.109)	(0.069)	
Component x Treat	-0.071	-0.694***	-0.029	
	(0.790)	(0.005)	(0.892)	
Observations	84955	79407	82149	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding, and regressions are specified as in the human capital table.

Table A.5: Police Heterogeneity Components

	Components			
	Nr Staff	Elec	Internet	Nr Vehicle
Treatment	0.356**	0.361**	0.351**	0.357**
	(0.045)	(0.043)	(0.048)	(0.044)
Component x Treatment	0.102	0.107	-0.405	0.120
	(0.648)	(0.775)	(0.222)	(0.545)
Observations	84955	84713	84880	84955

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. In each regression, the column title indicates the name of the variable which was interacted with treatment in that specification, all of which are components of the police quality index. All regressions include strata and district fixed effects, individual controls, and an index for the urbanness of the respondent's village. All regressions are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table A.6: Input Capacity Heterogeneity Components

Panel A. Police				
	Nr Staff	Elec	Internet	Nr Vehicle
Treatment	0.352**	0.355**	0.347*	0.354**
	(0.047)	(0.046)	(0.051)	(0.046)
Component x Treatment	0.114	-0.179	-0.409	0.170
	(0.610)	(0.581)	(0.218)	(0.388)
Observations	84955	84713	84880	84955
Panel B. Courts				
	Nr Staff	Elec	Internet	
Treatment	0.355**	0.300*	0.317*	
	(0.046)	(0.099)	(0.077)	
Component x Treatment	-0.259	-0.646***	0.008	
	(0.192)	(0.002)	(0.970)	
Observations	84955	79407	82149	
Panel C. Health				
	Health Spend	Has Oxy	Has ORS	Has Depo
Treatment	0.353**	0.354**	0.352**	0.351**
	(0.046)	(0.046)	(0.047)	(0.048)
Component x Treatment	0.034	0.971	-0.625	-0.770
	(0.830)	(0.126)	(0.392)	(0.484)
Observations	84955	84955	84955	84955
Panel D. Education				
	Teach Ratio	Piped Water	Prop Space	
Treatment	0.351**	0.352**	0.352**	
	(0.048)	(0.047)	(0.048)	
Component x Treatment	-0.457*	-0.307	-0.732**	
	(0.097)	(0.299)	(0.044)	
Observations	84955	84955	84955	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding, and regressions are specified as in the input capacity table.

Table A.7: Predicted Capacity Heterogeneity (OLS)

	Paid Tax			
	(1)	(2)	(3)	(4)
Treatment	0.576***	0.574***	0.561***	0.596***
	(0.003)	(0.003)	(0.004)	(0.002)
Predicted Tax Capacity	6.396***	5.923***	5.018***	5.133***
	(0.000)	(0.000)	(0.000)	(0.000)
Treatment * Predicted Tax Capacity	-5.005***	-4.745***	-4.318***	-4.470***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	82165	82165	82165	82165
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Control	No	No	Yes	Yes
All Index Interactions	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. Predicted tax capacity is estimated in a first stage using OLS regression. All regressions include strata fixed effects and are estimated with robust standard errors. Depending on the specification, regressions may include district fixed effects, individual controls, an index for the urbanness of the respondent's village and the human capital, enforcement and urbanness indices with treatment interactions. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table A.8: Predicted Capacity Heterogeneity (Elastic Net)

	Paid Tax			
	(1)	(2)	(3)	(4)
Treatment	0.568***	0.568***	0.572***	0.600***
	(0.004)	(0.003)	(0.003)	(0.002)
Predicted Tax Capacity	3.060***	2.723***	1.809***	1.979***
	(0.000)	(0.000)	(0.000)	(0.000)
Treatment * Predicted Tax Capacity	-0.994***	-0.963***	-0.867***	-1.118***
	(0.000)	(0.000)	(0.001)	(0.000)
Observations	82165	82165	82165	82165
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Control	No	No	Yes	Yes
All Index Interactions	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. Predicted tax capacity is estimated in a first stage using elastic net regression. All regressions include strata fixed effects and are estimated with robust standard errors. Depending on the specification, regressions may include district fixed effects, individual controls, an index for the urbanness of the respondent's village and the human capital, enforcement and urbanness indices with treatment interactions. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table A.9: Predicted Capacity Heterogeneity (Ridge)

	Paid Tax			
	(1)	(2)	(3)	(4)
Treatment	0.567***	0.568***	0.572***	0.600***
	(0.004)	(0.003)	(0.003)	(0.002)
Predicted Tax Capacity	3.089***	2.748***	1.843***	2.013***
	(0.000)	(0.000)	(0.000)	(0.000)
Treatment * Predicted Tax Capacity	-1.028***	-0.995***	-0.894***	-1.143***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	82165	82165	82165	82165
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Control	No	No	Yes	Yes
All Index Interactions	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. Predicted tax capacity is estimated in a first stage using ridge regression. All regressions include strata fixed effects and are estimated with robust standard errors. Depending on the specification, regressions may include district fixed effects, individual controls, an index for the urbanness of the respondent's village and the human capital, enforcement and urbanness indices with treatment interactions. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

B.2 Separate Treatments Tables

Table B.4: Individual Heterogeneity

	Paid Tax		
	(1)	(2)	(3)
Inform	0.440**	0.429*	0.441**
	(0.046)	(0.051)	(0.046)
Enforce	0.802***	0.735***	0.799***
	(0.000)	(0.001)	(0.000)
Encourage	-0.055	-0.090	-0.058
	(0.800)	(0.675)	(0.786)
Inform x Last Tax Payment (1k USD)	-0.081		-0.065
	(0.681)		(0.741)
Enforce x Last Tax Payment (1k USD)	-0.353**		-0.330**
	(0.021)		(0.031)
Encourage x Last Tax Payment (1k USD)	-0.186		-0.184
	(0.311)		(0.305)
Inform x Years Since Registration		-0.134**	-0.134**
		(0.039)	(0.038)
Enforce x Years Since Registration		-0.124*	-0.108
		(0.066)	(0.111)
Encourage x Years Since Registration		-0.042	-0.037
		(0.519)	(0.575)
Observations	84955	84955	84955
Individual Controls	Yes	Yes	Yes
District FE	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. All regressions include strata and district fixed effects as well as individual controls and are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table B.5: Human Capital and Enforcement Heterogeneity

	Paid Tax		
	(1)	(2)	(3)
Inform	0.432**	0.434**	0.432**
	(0.049)	(0.049)	(0.049)
Enforce	0.731***	0.731***	0.730***
	(0.001)	(0.001)	(0.001)
Encourage	-0.093	-0.092	-0.093
	(0.664)	(0.670)	(0.665)
Inform x Human Capital Index	-0.702***		-0.698***
	(0.006)		(0.006)
Enforce x Human Capital Index	-0.857***		-0.848***
	(0.001)		(0.001)
Encourage x Human Capital Index	0.285		0.275
	(0.243)		(0.262)
Inform x Enforcement Capacity		0.098	0.037
		(0.699)	(0.885)
Enforce x Enforcement Capacity		0.153	0.077
		(0.562)	(0.772)
Encourage x Enforcement Capacity		-0.115	-0.087
		(0.637)	(0.721)
Observations	84955	84955	84955
Individual Controls	Yes	Yes	Yes
District FE	Yes	Yes	Yes
Urban Interaction	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. All regressions include strata and district fixed effects, individual controls, and an index for the urbanness of the respondent's village. All regressions are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table B.6: Input Capacity Heterogeneity

	Paid Tax			
	(1)	(2)	(3)	(4)
Inform	0.481**	0.428*	0.428*	0.432**
	(0.032)	(0.052)	(0.052)	(0.050)
Inform x Capacity Index	-0.557**	-0.565**	-0.565**	-0.566**
	(0.018)	(0.014)	(0.014)	(0.022)
Enforce	0.741***	0.728***	0.728***	0.730***
	(0.001)	(0.001)	(0.001)	(0.001)
Enforce x Capacity Index	-0.647**	-0.631***	-0.631***	-0.505*
	(0.010)	(0.010)	(0.010)	(0.054)
Encourage	-0.117	-0.098	-0.098	-0.096
	(0.593)	(0.648)	(0.648)	(0.656)
Encourage x Capacity Index	0.014	-0.001	-0.001	0.089
	(0.951)	(0.995)	(0.995)	(0.712)
Observations	84955	84955	84955	84955
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Index Interactions	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. All regressions include strata fixed effects and estimated with robust standard errors. Depending on the specification, regressions may include district fixed effects, individual controls, and an index for the urbanness of the respondent's village. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table B.7: Predicted Capacity Heterogeneity

	Paid Tax			
	(1)	(2)	(3)	(4)
Predicted Tax Capacity	2.854***	2.545***	1.354***	1.483***
	(0.000)	(0.000)	(0.000)	(0.000)
Inform	0.471**	0.452**	0.442*	0.464**
	(0.047)	(0.049)	(0.068)	(0.021)
Inform x Pred Tax Capacity	-1.188***	-1.141***	-1.030***	-1.165***
	(0.002)	(0.002)	(0.003)	(0.000)
Enforce	0.685***	0.703***	0.708***	0.754***
	(0.003)	(0.003)	(0.002)	(0.001)
Enforce x Pred Tax Capacity	-1.085**	-1.026**	-0.769**	-0.970***
	(0.014)	(0.016)	(0.023)	(0.007)
Encourage	-0.142	-0.101	-0.105	-0.073
	(0.507)	(0.642)	(0.646)	(0.739)
Encourage x Pred Tax Capacity	-1.479***	-1.439***	-1.441***	-1.508***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	82165	82165	82165	82165
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Control	No	No	Yes	Yes
All Index Interactions	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. All regressions include strata fixed effects; standard errors are bootstrapped using 1000 replications of the two-step procedure. Depending on the specification, regressions may include district fixed effects, individual controls, an index for the urbanness of the respondent's village and the human capital, enforcement and urbanness indices with treatment interactions. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table B.9: Predicted Capacity and Individual Heterogeneity

	Paid Tax			
	(1)	(2)	(3)	(4)
Inform	0.496**	0.474**	0.464**	0.465**
	(0.035)	(0.038)	(0.047)	(0.045)
Inform x Pred Tax Capacity	-1.220***	-1.166***	-1.066***	-1.058***
	(0.001)	(0.002)	(0.002)	(0.002)
Enforce	0.764***	0.778***	0.787***	0.787***
	(0.001)	(0.001)	(0.001)	(0.000)
Enforce x Pred Tax Capacity	-1.120***	-1.049**	-0.814**	-0.799**
	(0.009)	(0.014)	(0.018)	(0.027)
Encourage	-0.109	-0.064	-0.067	-0.067
	(0.629)	(0.775)	(0.772)	(0.753)
Encourage x Pred Tax Capacity	-1.486***	-1.444***	-1.446***	-1.447***
	(0.000)	(0.000)	(0.000)	(0.000)
Inform x Years Registration	-0.169**	-0.155**	-0.154**	-0.154**
	(0.012)	(0.016)	(0.023)	(0.020)
Inform x Last Payment (1k USD)	-0.069	-0.065	-0.067	-0.069
	(0.742)	(0.754)	(0.751)	(0.741)
Enforce x Years Registration	-0.135**	-0.127*	-0.125*	-0.125*
	(0.048)	(0.069)	(0.072)	(0.064)
Enforce x Last Payment (1k USD)	-0.357*	-0.318*	-0.327*	-0.327*
	(0.069)	(0.083)	(0.087)	(0.087)
Encourage x Years Registration	-0.065	-0.061	-0.059	-0.059
	(0.330)	(0.378)	(0.395)	(0.359)
Encourage x Last Payment (1k USD)	-0.196	-0.180	-0.186	-0.186
	(0.338)	(0.342)	(0.353)	(0.345)
Observations	82165	82165	82165	82165
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Control	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding, and specification is as in the predicted capacity and individual heterogeneity table.

Table B.10: Predicted Capacity and New Services Heterogeneity

	Paid Tax			
	(1)	(2)	(3)	(4)
Inform	0.495**	0.475**	0.460**	0.460**
	(0.044)	(0.015)	(0.034)	(0.039)
Inform x Predicted Tax Capacity	-0.896**	-0.882***	-0.878***	-0.877***
	(0.010)	(0.006)	(0.006)	(0.008)
New Services x Tax Capacity x Inform	-0.515**	-0.489***	-0.380*	-0.378*
	(0.016)	(0.006)	(0.064)	(0.075)
Enforce	0.718***	0.731***	0.726***	0.725***
	(0.004)	(0.001)	(0.002)	(0.001)
Enforce x Predicted Tax Capacity	-0.602	-0.607	-0.549	-0.548
	(0.138)	(0.199)	(0.111)	(0.121)
New Services x Tax Capacity x Enforce	-0.566**	-0.524***	-0.384*	-0.382*
	(0.011)	(0.007)	(0.072)	(0.085)
Encourage	-0.128	-0.090	-0.093	-0.093
	(0.582)	(0.645)	(0.652)	(0.683)
Encourage x Predicted Tax Capacity	-1.409***	-1.376***	-1.384***	-1.384***
	(0.000)	(0.000)	(0.000)	(0.000)
New Services x Tax Capacity x Encourage	-0.299	-0.263	-0.248	-0.247
	(0.117)	(0.251)	(0.167)	(0.196)
Observations	82165	82165	82165	82165
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Control	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the Google Maps method of geocoding. All regressions include strata fixed effects; standard errors are bootstrapped using 1000 replications of the two-step procedure.. Depending on the specification, regressions may include district fixed effects, individual controls, an index for the urbanness of the respondent's village and the human capital, enforcement and urbanness indices with treatment interactions. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

B.3 ArcGIS Tables

Table C.1: Balance Table

Variable	Control	Inform	Encourage	Enforce	p (All)	p (T v C)
Taxpayer Age (years)	42.6	42.7	42.6	42.6	0.234	0.680
Years Since Registration	5.62	5.6	5.65	5.66	0.550	0.598
Taxpayer is Male (d)	.703	.707	.711	.706	0.441	0.178
Located in Kampala (d)	.499	.492	.493	.499	0.292	0.234
Registered Business (d)	.699	.691	.699	.697	0.193	0.321
Paid Tax 2017-18 (d)	.498	.503	.496	.498	0.367	0.818
Unpaid by June 28 (d)	.699	.7	.706	.699	0.298	0.489
Last Payment (1k USD)	.197	.214	.235	.237	0.750	0.264
N=78,240						.

The first four columns represent the group-specific mean for the variables listed on the lefthand side, based on the URA administrative database. The sample size is based on the ArcGIS method of geocoding. p(All) contains the p-value from a test of joint equality for coefficients on inform, encourage and enforce; p(T v C) contains the p-value for the coefficient on dummy variable indicating the respondent was assigned to any treatment group.

Table C.2: Treatment Effects

Panel A. Separate Treatments						
	Paid Tax		Presumptive		PIT	
Inform	0.552**	0.513**	0.279*	0.279*	0.282	0.244
	(0.019)	(0.027)	(0.058)	(0.056)	(0.136)	(0.192)
Encourage	-0.111	-0.130	0.069	0.066	-0.166	-0.180
	(0.626)	(0.565)	(0.630)	(0.643)	(0.364)	(0.323)
Enforce	0.737***	0.721***	0.701***	0.727***	0.076	0.033
	(0.002)	(0.002)	(0.000)	(0.000)	(0.684)	(0.856)
Inform vs. Enforce	0.444	0.380	0.008	0.004	0.277	0.261
Inform vs. Encourage	0.005	0.005	0.156	0.145	0.016	0.022
Encourage vs. Enforce	0.000	0.000	0.000	0.000	0.187	0.242
Panel B. Pooled Treatment						
	Paid Tax		Presumptive		PIT	
Any Treatment	0.391**	0.367**	0.350***	0.357***	0.063	0.032
	(0.039)	(0.049)	(0.003)	(0.002)	(0.677)	(0.832)
Individual Controls	No	Yes	No	Yes	No	Yes
Control Mean	5.460	5.460	2.028	2.028	3.500	3.500
Observations	78240	78240	78240	78240	78240	78240

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Outcomes consist of indicators for whether any tax, the presumptive, and the personal income tax, respectively, were paid between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the ArcGIS method of geocoding. All regressions include strata and district fixed effects and are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table C.3: Treatment Effects

Panel A. Separate Treatments						
	Tax Amount (USD)		Presumptive		PIT	
Inform	0.194	0.168	0.293*	0.289*	0.119	0.095
	(0.285)	(0.351)	(0.062)	(0.064)	(0.497)	(0.584)
Encourage	-0.046	-0.050	0.027	0.030	-0.034	-0.038
	(0.799)	(0.779)	(0.848)	(0.832)	(0.844)	(0.825)
Enforce	0.162	0.132	0.531***	0.548***	0.028	-0.003
	(0.379)	(0.469)	(0.001)	(0.000)	(0.872)	(0.986)
Inform vs. Enforce	0.859	0.843	0.149	0.115	0.602	0.570
Inform vs. Encourage	0.187	0.227	0.079	0.084	0.381	0.443
Encourage vs. Enforce	0.259	0.319	0.001	0.001	0.720	0.838
Panel B. Pooled Treatment						
	Tax Amount (USD)		Presumptive		PIT	
Any Treatment	0.103	0.083	0.284**	0.289**	0.037	0.018
	(0.488)	(0.573)	(0.021)	(0.018)	(0.793)	(0.900)
Individual Controls	No	Yes	No	Yes	No	Yes
Control Mean	3.074	3.074	1.497	1.497	2.894	2.894
Observations	78240	78240	78240	78240	78240	78240

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Outcomes consist of the amount paid for any tax, the presumptive, and the personal income tax, respectively, between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the ArcGIS method of geocoding. All regressions include strata and district fixed effects and are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table C.4: Individual Heterogeneity

	Paid Tax		
	(1)	(2)	(3)
Treatment	0.446**	0.369**	0.436**
	(0.017)	(0.048)	(0.020)
Last Tax Payment (1k USD)	0.503***		0.470***
	(0.002)		(0.004)
Treatment x Last Tax Payment (1k USD)	-0.382**		-0.354**
	(0.031)		(0.045)
Years Since Registration		0.260***	0.235***
		(0.000)	(0.000)
Treatment x Years Since Registration		-0.181***	-0.161***
		(0.004)	(0.010)
Observations	78240	78240	78240
Individual Controls	Yes	Yes	Yes
District FE	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the ArcGIS method of geocoding. All regressions include strata and district fixed effects as well as individual controls and are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table C.5: Human Capital and Enforcement Heterogeneity

	Paid Tax		
	(1)	(2)	(3)
Treatment	0.367**	0.365*	0.366**
	(0.049)	(0.050)	(0.049)
Human Capital Index	0.325*		0.316
	(0.092)		(0.102)
Treatment x Human Capital Index	-0.548***		-0.543***
	(0.005)		(0.005)
Enforcement Capacity		0.233	0.258
		(0.334)	(0.286)
Treatment x Enforcement Capacity		0.098	0.067
		(0.655)	(0.759)
Observations	78240	78240	78240
Individual Controls	Yes	Yes	Yes
District FE	Yes	Yes	Yes
Urban Interaction	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the ArcGIS method of geocoding. All regressions include strata and district fixed effects, individual controls, and an index for the urbanness of the respondent's village. All regressions are estimated with robust standard errors. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table C.6: Input Capacity Heterogeneity

	Paid Tax			
	(1)	(2)	(3)	(4)
Treatment	0.393**	0.366**	0.366**	0.365*
	(0.038)	(0.049)	(0.049)	(0.050)
Capacity Index	0.134	0.488**	0.488**	0.485**
	(0.450)	(0.014)	(0.014)	(0.015)
Treatment x Capacity Index	-0.496**	-0.476**	-0.476**	-0.477**
	(0.017)	(0.019)	(0.019)	(0.018)
Observations	78240	78240	78240	78240
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Controls	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the ArcGIS method of geocoding. All regressions include strata fixed effects and are estimated with robust standard errors. Depending on the specification, regressions may include district fixed effects, individual controls, and an index for the urbanness of the respondent's village. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table C.7: Predicted Capacity Heterogeneity

	Paid Tax			
	(1)	(2)	(3)	(4)
Treatment	0.291	0.289	0.289	0.301
	(0.129)	(0.127)	(0.128)	(0.113)
Predicted Tax Capacity	3.066***	2.796***	1.869***	1.973***
	(0.000)	(0.000)	(0.000)	(0.000)
Treatment x Predicted Tax Capacity	-1.272***	-1.237***	-1.192***	-1.287***
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	76208	76208	76208	76208
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Control	No	No	Yes	Yes
All Index Interactions	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the ArcGIS method of geocoding. Predicted tax capacity is estimated in a first stage using LASSO regression. All regressions include strata fixed effects; standard errors are bootstrapped using 1000 replications of the two-step procedure. Depending on the specification, regressions may include district fixed effects, individual controls, an index for the urbanness of the respondent's village and the human capital, enforcement and urbanness indices with treatment interactions. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table C.8: Predicted Tax Capacity and State Perceptions

	Predicted Capacity
To find out what taxes and fees you are supposed to pay is:	
Very difficult	$-^{***}$
Not paying the taxes they owe on their income is:	
Wrong and punishable	$+^{***}$
A good citizen should always:	
Pay taxes they owe to government	$+^{***}$
The tax authority has the right to make people pay taxes:	
Agree or Strongly Agree	$+^{***}$
Ease of Service Access Index	$+^{***}$

Each row in this table is the result of a different regression correlating predicted tax capacity estimated using the ArcGIS method of geocoding and local averaged Afrobarometer opinions based on the reported questions. Ease of Service Access Index is an index combining z-scores of a variety of questions on service access; all other regressions include only a single question, defined as in the table. All results reported are statistically significant, and all regressions control for district fixed effects and index measures of local health inputs, education inputs, urbanness, court inputs and police inputs.

Table C.9: Predicted Capacity and Individual Heterogeneity

	Paid Tax			
	(1)	(2)	(3)	(4)
Treatment	0.360*	0.362*	0.363*	0.363*
	(0.062)	(0.057)	(0.057)	(0.057)
Treatment x Predicted Tax Capacity	-1.290***	-1.251***	-1.215***	-1.213***
	(0.000)	(0.000)	(0.000)	(0.000)
Treatment x Years Since Registration	-0.238***	-0.203***	-0.218***	-0.218***
	(0.006)	(0.006)	(0.005)	(0.005)
Treatment x Last Tax Payment (1k USD)	-0.377**	-0.351**	-0.355**	-0.355**
	(0.041)	(0.046)	(0.045)	(0.045)
Observations	76208	76208	76208	76208
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Control	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made from the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the ArcGIS method of geocoding. All regressions include strata fixed effects; standard errors are bootstrapped using 1000 replications of the two-step procedure. Depending on the specification, regressions may include district fixed effects, individual controls, an index for the urbanness of the respondent's village and the human capital, enforcement and urbanness indices with treatment interactions. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

Table C.10: Predicted Capacity and New Services Heterogeneity

	Paid Tax			
	(1)	(2)	(3)	(4)
Treatment	0.298	0.299	0.303	0.303
	(0.122)	(0.116)	(0.112)	(0.113)
Treatment x Predicted Tax Capacity	-1.265***	-1.220***	-1.140***	-1.137***
	(0.000)	(0.000)	(0.000)	(0.000)
Treatment x New Services Intensity Index	-0.018	0.003	0.021	0.025
	(0.928)	(0.988)	(0.913)	(0.899)
New Services x Tax Capacity x Treatment	-0.076	-0.100	-0.145	-0.149
	(0.671)	(0.572)	(0.430)	(0.420)
Observations	76208	76208	76208	76208
Individual Controls	No	Yes	Yes	Yes
District FE	No	No	Yes	Yes
Urban Control	No	No	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The outcome in each regression is an indicator for whether any tax payments were made between the date of treatment, June 28, and September 1, 2019. The sample size is based on the matched sample using the ArcGIS method of geocoding. All regressions include strata fixed effects; standard errors are bootstrapped using 1000 replications of the two-step procedure. Depending on the specification, regressions may include district fixed effects, individual controls, an index for the urbanness of the respondent's village and the human capital, enforcement and urbanness indices with treatment interactions. Individual controls include age, male, whether the individual registered with a business, whether the individual paid in 2017-18, whether the individual had paid by June 2019, years since TIN registration, and size of most recent tax payment.

B.4 Model

I construct a simple, one-period model of the state's tax revenue maximization problem. The state, working from a budget constraint, spends on some combination of enforcement, sensitization, and human capital investment. I introduce a non-linearity in enforcement, which reflects the real-world need for a country to build up meaningful infrastructure before enforcement is feasible, and allow for there to be positive tax compliance even in the absence of enforcement, consistent with evidence on the non-zero rate of compliance in Uganda. I explore the implications for the complementarity of the marginal change in compliance from sensitization (for which I argue my treatment is a good proxy) with enforcement and human capital spending, and their combination, which we might typically define as capacity, under zero and non-negative levels of enforcement.

Specifically, I model three components of state investment: enforcement investment e, which feeds into enforcement rate $\pi(e)$; sensitization investment s, which feeds into compliance rate $c(s)$; and human capital investment h, which feeds into taxable income $y(h)$.[1] The state allocates e, s and h to maximize revenue R_1, starting from some initial endowment R_0.

I introduce a non-linearity in e which reflects the need to build up meaningful infrastructure before enforcement in realistically possible. In other words, enforcement cannot be done at all unless e is higher than some $\bar{e}$.

Set-up

Think of the state of Uganda as solving the following:

$$\max_{s,h,e} R_1 = c(s)y(h)t + \pi(e)[1 - c(s)]y(h)tf - e - s - h$$

subject to

$$\begin{aligned}
&R_0 \geq e + s + h \\
&0 \leq \pi(e) \leq 1 \text{ and } \pi(0) = 0 \\
&0 \leq c(s) \leq 1 \\
&e = \max(0, e - \bar{e})
\end{aligned}$$

where t is a fixed tax rate and f is a fine proportionate to the evaded tax.

No Enforcement

I focus here on the case where $\pi(e^*) = 0$, i.e. the state is not enforcing.

[1] I model compliance as a function of sensitization rather than enforcement based both on the fact that compliance occurs in low enforcement environments, but also in accordance with recent evidence on the importance of salience in economic decision-making (Chetty, Looney, and Kroft, 2009), which has been suggested to be relevant to taxation by Meiselman (2018) and Bergolo et al. (2019).

I can then take first order conditions with respect to s:

$$\frac{dR}{ds} = c'(s^*)y(h)t - 1 = 0$$
$$\implies c'(s^*) = \frac{1}{y(h)t}$$

Based on the model's first order conditions:

- $c'(s^*) > 0 \implies$ the treatment effect is positive
- $\frac{dc'(s^*)}{dh} < 0 \implies$ the effect of the treatment should be decreasing in human capital investment
- $\frac{dc'(s^*)}{de} = 0 \implies$ the effect of the treatment should be unrelated to enforcement investment

State capacity, g, traditionally defined, is $g = e + h$; $\frac{dc'(s^*)}{dg}$ should be positive.

When there is no enforcement, the model suggests a substitution effect between $c'(s^*)$, a proxy for the treatment effect, and human capital investment h (or, alternatively, effective income $y(h)$). These results suggest that the state faces a trade-off in allocating s and h, and in some region of its objective function, they are actually substitutes. Intuitively, this result makes sense; the idea is that under scarce resources, the state would want to balance between investments in s and h, rather than putting all of its resources in one or the other, though the precise extent of the trade-off would depend on the functional forms for $c(s)$ and $y(h)$.

Some Enforcement

Under $\pi(e^*) > 0$, I derive fairly similar results.

The first order condition with respect to s becomes:

$$\frac{dR}{ds} = c'(s^*)y(h)t - \pi(e)c'(s^*)y(h)tf - 1 = 0$$
$$\implies c'(s^*) = \frac{1}{y(h)t[1 - \pi(e)f]}$$

Based on the model's first order conditions:

- $c'(s^*) \implies$ the treatment effect depends on whether $\pi(e)f$ is greater or less than one
- $\frac{dc'(s^*)}{dh} < 0 \implies$ the effect of the treatment should be decreasing in human capital investment
- $\frac{dc'(s^*)}{de} > 0 \implies$ the effect of the treatment should be increasing in enforcement investment

State capacity, g, traditionally defined, is $g = e + h$; $\frac{dc'(s^*)}{dg}$ is ambiguous.

Under the version of the model where some non-zero level of enforcement is revenue-maximizing, I find similar predicted heterogeneity results. Here, we see that the marginal change in c at s^* is decreasing in human capital investment h, and increasing in enforcement investment e under certain assumptions. In other words, optimally, sensitization is a substitute for human capital investment and complements enforcement capacity. In this version of the model, the direction of the interaction with state capacity would be ambiguous, suggesting the results that I find in Uganda may not extend to materially different contexts with regard to capacity.

B.5 Data Appendix

Index Construction

In this section, I outline the specific variables in the construction of each index used in the paper.

Human Capital Capacity$_j$ (MoES, 2016; MoH, 2016; JLoS, 2013): A measure of school inputs in 2016, health inputs in 2017, and local court inputs in 2013 in the vicinity of the business.

- An indicator for whether the school has piped water or a borehold, as opposed to well water, spring water, rain water, or lake or river water
- Teacher student ratio
- Proportion of students with adequate space (including the use of a desk and chair)
- Proportion of students in P4 who advanced to P5 from the last to the current year
- Percentage of audited maternal deaths
- Percentage of audited perinatal deaths
- Percentage of mothers receiving PNC checks within six days
- Percentage of babies receiving PNC checks within six days
- Percentage of days without stockout of oral rehydration solution (ORS)
- Percentage of days without stockout of depo provera
- Percentage of days without stockout of oxycontin
- Number of staff in the court
- Indicator for whether the court building has electricity
- Indicator for whether the court builidng has internet

Enforcement Capacity$_j$ (JLoS, 2013): A measure of police post inputs in 2013 in the vicinity of the business.

- Number of staff in the police post
- Number of vehicles the police post has
- Indicator for whether the police post has electricity
- Indicator for whether the police post has internet

Urban$_v$ (Goodman et al., 2019): A measure of the economic activity and urbanness in the village associated with a business.

- VIIRS nighttime lights, village mean (Elvidge et al., 2017)
- Distance to nearest city of 50,000 or more, village mean (Weiss et al., 2018)
- Population density, village mean (CIESIN, 2016)

Input Capacity$_{jv}$ (MoES, 2016; MoH, 2016; JLoS, 2013; AidData, 2018): A measure of local inputs towards state capacity from various sources and years.

- Number of staff in the police post
- Number of vehicles the police post has
- An indicator for whether the police post has electricity
- An indicator for whether the police post has internet
- An indicator for whethe the school has piped water or a borehold, as opposed to using well water, spring water, rain water, or lake or river water
- Teacher student ratio
- Proportion of students with adequate space (including the use of a desk and chair)
- Percentage of days without stockout of oral rehydration solution (ORS)
- Percentage of days without stockout of depo provera
- Percentage of days without stockout of oxycontin
- Spending in the nearby village on health projects in 2014-15
- Number of staff in the court
- Indicator for whether the court building has electricity
- Indicator for whether the court builidng has internet

New Services$_j$ (MoES, 2016; MoH, 2016; JLoS, 2013): A measure of recent service investments

- Proportion of schools near the business in 2016 which were built between 2011 and 2016
- Proportion of courts near the business in 2013 which were built between 2011 and 2013
- Proportion of schools near the business in 2013 which were built between 2011 and 2013

Appendix C

Appendix for "The Costs of Splitting: Administrative Unit Proliferation and Economic Growth in Uganda"

C.1 Supplementary Tables

Table C.1: Distributionary Effects of District Creation

	Ratio to 100 Students of		Prop has	
	Teachers	Rooms	Adequate Space	Safe Water
Parent District	0.031	0.025	0.001	0.014**
	(0.020)	(0.067)	(0.003)	(0.007)
New District	-0.007	-0.031	-0.002	0.003
	(0.018)	(0.060)	(0.003)	(0.007)
Observations	1221	1221	1221	1221
Dataset	MoES	MoES	MoES	MoES
Level	District SD	District SD	District SD	District SD
Years	2006-16	2006-16	2006-16	2006-16
Year FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Subregion x Year FE	Yes	Yes	Yes	Yes
Non-Split Mean	0.40	0.52	0.13	0.39
Parent v New	0.09	0.45	0.47	0.15

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include district-level standard deviation calculations for the ratio of teachers and classrooms per hundred students, and the school-level proportion of students with adequate space and access to safe water. Standard errors are clustered at the district level for all regressions. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Table C.2: Political Effects of District Creation

	Support for		Voting		NRM Won	
	Central Gov't	Local Gov't	Turnout	Pres Share	MP Seat	SC Chair
Parent, Short	0.019	0.058	-0.007	-0.011	-0.015	-0.030
	(0.074)	(0.069)	(0.009)	(0.015)	(0.117)	(0.049)
Parent, Medium	0.021	-0.015	-0.001	0.010	0.038	-0.044
	(0.062)	(0.072)	(0.010)	(0.015)	(0.118)	(0.049)
New, Short	0.190**	0.074	0.015	0.014	0.036	-0.029
	(0.093)	(0.089)	(0.010)	(0.017)	(0.129)	(0.052)
New, Medium	-0.077	-0.024	0.019	0.032**	0.176	0.011
	(0.093)	(0.091)	(0.012)	(0.015)	(0.115)	(0.052)
Observations	8975	8971	333	333	678	3544
Dataset	Afrobrmtr	Afrobrmtr	Elections	Elections	Elections	Elections
Level	Individual	Individual	District	District	Const	Subcounty
Years	2005-18	2005-18	2006-16	2006-16	2006-16	2006-16
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes
Subreg x Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Non-Split Mean	0.89	0.89	0.66	0.67	0.76	0.72
P v N, Short	0.08	0.86	0.09	0.17	0.74	0.99
P v N, Medium	0.29	0.93	0.16	0.19	0.32	0.35

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include indices capturing levels of support for central and local governments, turnout relative to potential voters in presidential elections, the share of presidential election support for President Museveni, and whether a given parliamentary seat or subcounty chair seat was won by the National Resistance Movement. Short run refers to the period from 2010 to 2015, and medium run refers to the period from 2016 on. Standard errors are clustered at the village level for Afrobarometer data, the district level for presidential elections data, and the constituency and subcounty level, respectively, for other election data. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Table C.3: Area Development Index Components

	Village Area Has									
	Elec	Water	Sewage	Cell	Postal	School	Police	Health	Market	Road
Parent District	0.052	0.110	0.043	0.155**	-0.011	0.138**	-0.049	0.049	0.040	-0.069
	(0.074)	(0.072)	(0.054)	(0.063)	(0.053)	(0.066)	(0.080)	(0.079)	(0.080)	(0.062)
New District	-0.017	-0.002	-0.017	0.048	0.015	0.034	0.022	-0.025	-0.233	-0.255**
	(0.101)	(0.104)	(0.065)	(0.132)	(0.052)	(0.130)	(0.143)	(0.176)	(0.153)	(0.111)
Observations	996	995	995	996	996	996	995	994	996	996
Dataset	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr
Level	Vill	Vill	Vill	Vill	Vill	Vill	Vill	Vill	Vill	Vill
Years	08-18	08-18	08-18	08-18	08-18	08-18	08-18	08-18	08-18	08-18
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SR x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N-S Mean	0.35	0.26	0.14	0.86	0.10	0.84	0.32	0.68	0.59	0.22
Parent v New	0.53	0.31	0.39	0.41	0.67	0.45	0.64	0.68	0.08	0.10

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include whether the local area in which surveying was conducted contains electricity infrastructure, piped water, sewage infrastructure, cell service, a post office, a school, a police station, a health center, a market, and whether the road to the area was well-maintained. Standard errors are clustered at the enumeration area level for all regressions. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Table C.4: Well-Being Index Components

	Household Not Without				
	Food	Clean Water	Medicine	Fuel	Income
Parent District	-0.016	0.016	0.017	0.026	0.013
	(0.031)	(0.036)	(0.029)	(0.030)	(0.018)
New District	-0.082*	-0.016	-0.039	-0.037	-0.066***
	(0.043)	(0.049)	(0.038)	(0.042)	(0.024)
Observations	8963	8971	8960	8966	8961
Dataset	Afrobrmtr	Afrobrmtr	Afrobrmtr	Afrobrmtr	Afrobrmtr
Level	Individual	Individual	Individual	Individual	Individual
Years	2005-18	2005-18	2005-18	2005-18	2005-18
Year FE	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes
Subregion x Year FE	Yes	Yes	Yes	Yes	Yes
Non-Split Mean	0.48	0.50	0.35	0.55	0.13
Parent v New	0.14	0.54	0.15	0.15	0.00

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include whether, in the last year, the household went without access to food, clean water, medicine, fuel or cash income. Standard errors are clustered at the enumeration area level for all regressions. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Table C.5: Asset Index Components

	Household Owns									
	House	Otr Bld	Frntr	Appl	Gnrtr	Solar	Bicycle	Moto	Jewelry	Mobile
Parent District	-0.024*	0.002	0.006	-0.042***	-0.003	-0.001	-0.013	0.005	-0.035**	-0.014
	(0.013)	(0.018)	(0.011)	(0.015)	(0.004)	(0.010)	(0.019)	(0.010)	(0.016)	(0.017)
New District	0.004	-0.044**	-0.014	-0.045***	0.001	0.001	-0.019	0.012	-0.015	0.000
	(0.014)	(0.019)	(0.010)	(0.016)	(0.005)	(0.008)	(0.020)	(0.009)	(0.018)	(0.017)
Observations	18727	18727	18727	18727	18727	18727	18727	18727	18727	18727
Dataset	UNPS	UNPS	UNPS	UNPS	UNPS	UNPS	UNPS	UNPS	UNPS	UNPS
Level	Hhd	Hhd	Hhd	Hhd	Hhd	Hhd	Hhd	Hhd	Hhd	Hhd
Years	05-18	05-18	05-18	05-18	05-18	05-18	05-18	05-18	05-18	05-18
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SR x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N-S Mean	0.85	0.20	0.91	0.13	0.01	0.10	0.37	0.08	0.15	0.60
Parent v New	0.09	0.04	0.10	0.86	0.44	0.84	0.79	0.50	0.34	0.48

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include whether the household reports currently owning their house, another building, furniture, appliances, a generator, solar panels, a bicycle, a motorcycle, jewelry, or a mobile phone. Standard errors are clustered at the enumeration area level for all regressions. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Table C.6: House Quality Index Components

	House Has			
	Sturdy Roof	Sturdy Walls	Sturdy Floor	Any Latrine
Parent District	0.007	0.014	0.000	-0.014
	(0.015)	(0.016)	(0.015)	(0.011)
New District	0.049***	0.032**	0.012	-0.020*
	(0.017)	(0.016)	(0.015)	(0.012)
Observations	20998	20999	20999	20984
Dataset	UNPS	UNPS	UNPS	UNPS
Level	Household	Household	Household	Household
Years	2001-18	2001-18	2001-18	2001-18
Year FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Subregion x Year FE	Yes	Yes	Yes	Yes
Non-Split Mean	0.71	0.58	0.30	0.92
Parent v New	0.04	0.32	0.53	0.67

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include whether the enumerator observes the house of the respondent as containing, based on their materials, a sturdy roof, sturdy walls, or a sturdy floor, and whether the household reports access to any sort of latrine. Standard errors are clustered at the enumeration area level for all regressions. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Table C.7: Local Government Support Index Components

	LG Manages Well		Level of		
	Roads	Markets	Trust	Listen	Approval
Parent District	0.006	0.019	-0.006	0.018	0.028
	(0.031)	(0.033)	(0.029)	(0.026)	(0.027)
New District	0.033	0.073	-0.012	-0.009	0.028
	(0.043)	(0.055)	(0.035)	(0.038)	(0.039)
Observations	8881	6643	8729	8749	8629
Dataset	Afrobrmtr	Afrobrmtr	Afrobrmtr	Afrobrmtr	Afrobrmtr
Level	Individual	Individual	Individual	Individual	Individual
Years	2005-18	2005-18	2005-18	2005-18	2005-18
Year FE	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes
Subregion x Year FE	Yes	Yes	Yes	Yes	Yes
Non-Split Mean	0.88	0.66	0.87	0.87	0.86
Parent v New	0.54	0.35	0.88	0.48	1.00

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include whether the respondent rates the local government as managing the roads and markets well, whether they trust their local government, whether their local government listens to them, and whether they approve of the performance of the local government. Standard errors are clustered at the enumeration area level for all regressions. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Table C.8: Central Government Support Index Components - Part 1

	CG Manages Well							
	Economy	Jobs	Prices	Inequal	Crime	Health	Educ	Water
Parent District	-0.005	-0.002	-0.004	0.006	0.060**	0.008	0.037	0.042
	(0.025)	(0.023)	(0.022)	(0.021)	(0.028)	(0.026)	(0.027)	(0.031)
New District	-0.024	0.041	-0.059*	-0.017	0.072*	-0.033	0.023	-0.053
	(0.035)	(0.034)	(0.032)	(0.028)	(0.043)	(0.038)	(0.039)	(0.045)
Observations	8762	8704	8841	8691	8840	8907	8889	8808
Dataset	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr
Level	Indiv	Indiv	Indiv	Indiv	Indiv	Indiv	Indiv	Indiv
Years	05-18	05-18	05-18	05-18	05-18	05-18	05-18	05-18
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SR x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N-S Mean	0.87	0.87	0.88	0.86	0.88	0.89	0.88	0.88
Parent v New	0.59	0.22	0.09	0.43	0.78	0.27	0.72	0.05

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include whether the respondent rates the central government as managing well the economy, job creation, inflation, inequality, crime, health services, education services, water and sewage, hunger, and corruption. Standard errors are clustered at the enumeration area level for all regressions. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

Table C.9: Central Government Support Index Components - Part 2

	CG Manages Well					Level Of	
	Hunger	Corrupt	Welfare	Infrastructure	Electricity	Trust Pres	Trust MP
Parent District	-0.003	0.007	0.041	-0.003	-0.034	-0.022	-0.027
	(0.024)	(0.024)	(0.027)	(0.035)	(0.038)	(0.025)	(0.026)
New District	0.049	-0.022	0.118***	0.065	0.010	0.028	0.011
	(0.037)	(0.041)	(0.041)	(0.080)	(0.059)	(0.039)	(0.038)
Observations	8719	8600	6755	6748	6483	8785	8661
Dataset	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr	Afrobr
Level	Indiv	Indiv	Indiv	Indiv	Indiv	Indiv	Indiv
Years	05-18	05-18	05-18	2008-18	2008-18	05-18	05-18
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
SR x Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N-S Mean	0.87	0.86	0.68	0.68	0.66	0.87	0.86
Parent v New	0.17	0.49	0.08	0.40	0.47	0.22	0.33

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include whether the respondent rates the central government well on managing improving living standards, maintaining infrastructure, and providing electricity, whether they trust the president and central government, whether any member of parliament would listen to them, whether they approve of the performance of the president and parliament, and whether they would vote for Uganda's primary political party, the National Resistance Movement, in an election held tomorrow. Standard errors are clustered at the enumeration area level for all regressions. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

	Level Of			
	Listen MP	Approval Pres	Approval MP	Vote NRM
Parent District	0.018	0.012	0.034	-0.033
	(0.023)	(0.024)	(0.031)	(0.025)
New District	0.042	0.045	0.027	0.004
	(0.038)	(0.034)	(0.048)	(0.033)
Observations	8611	8764	8623	8975
Dataset	Afrobrmtr	Afrobrmtr	Afrobrmtr	Afrobrmtr
Level	Individual	Individual	Individual	Individual
Years	2005-18	2005-18	2005-18	2005-18
Year FE	Yes	Yes	Yes	Yes
District FE	Yes	Yes	Yes	Yes
Subregion x Year FE	Yes	Yes	Yes	Yes
Non-Split Mean	0.86	0.88	0.86	0.89
Parent v New	0.51	0.33	0.88	0.29

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are in parenthesis. Outcomes include whether the respondent rates the central government well on managing improving living standards, maintaining infrastructure, and providing electricity, whether they trust the president and central government, whether any member of parliament would listen to them, whether they approve of the performance of the president and parliament, and whether they would vote for Uganda's primary political party, the National Resistance Movement, in an election held tomorrow. Standard errors are clustered at the enumeration area level for all regressions. All regressions include year fixed effects, district fixed effects, and an interaction between subregion and year fixed effects.

C.2 Figures

Figure A.1: Resource Effects Figures - Panel A

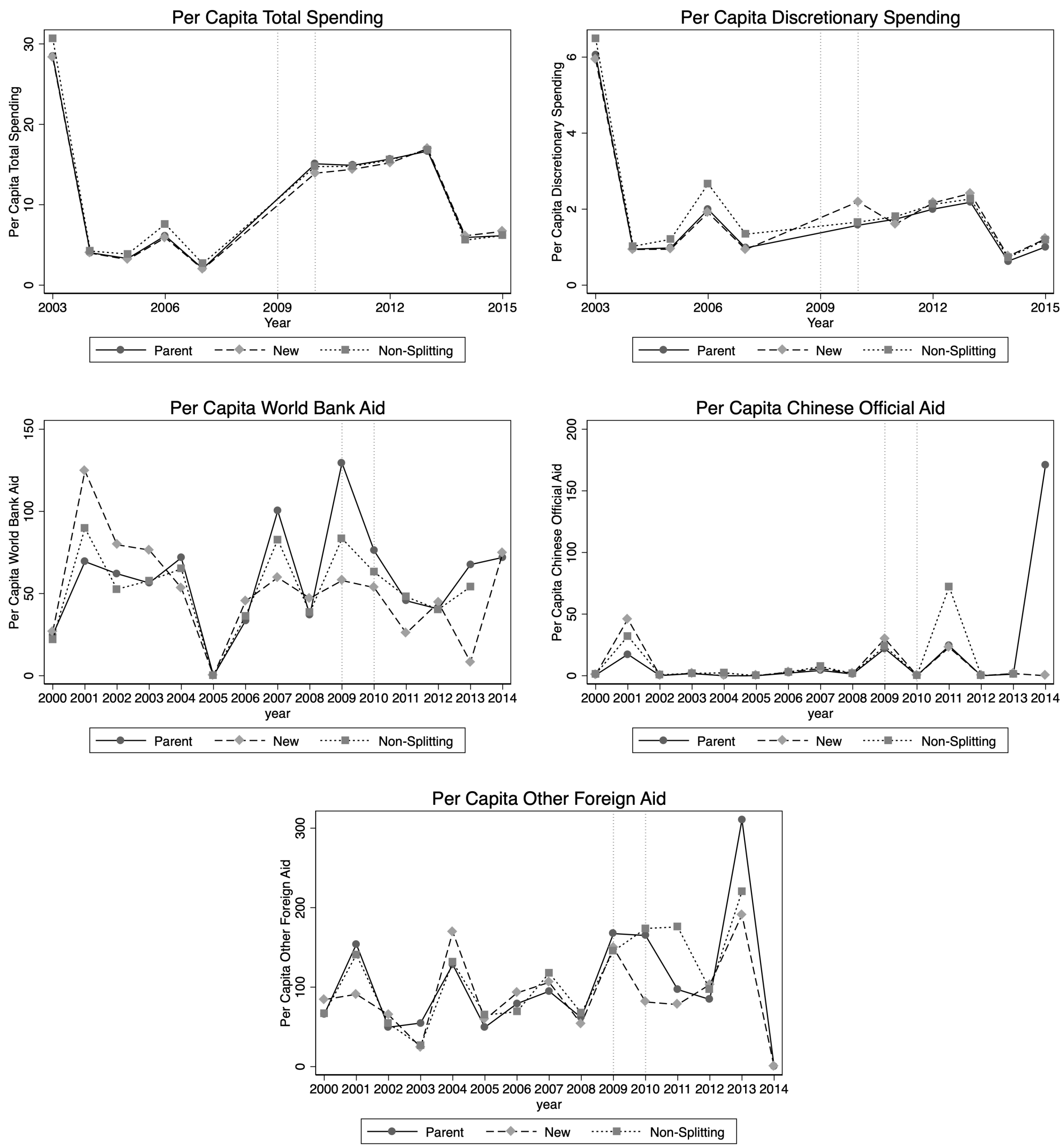


Figure A.2: Resource Effects Figures - Panel B

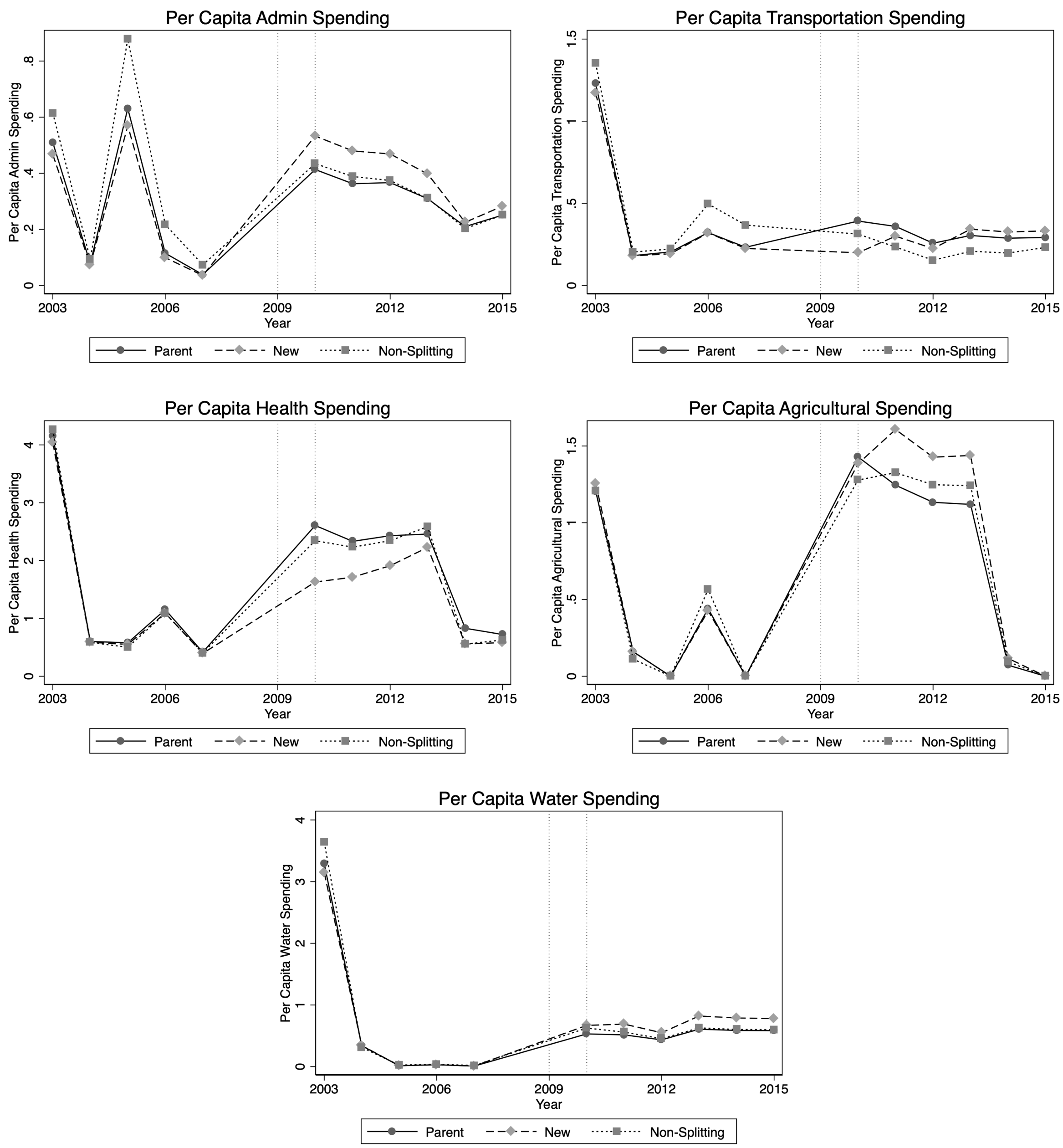


Figure A.3: Education Input Effects Figures

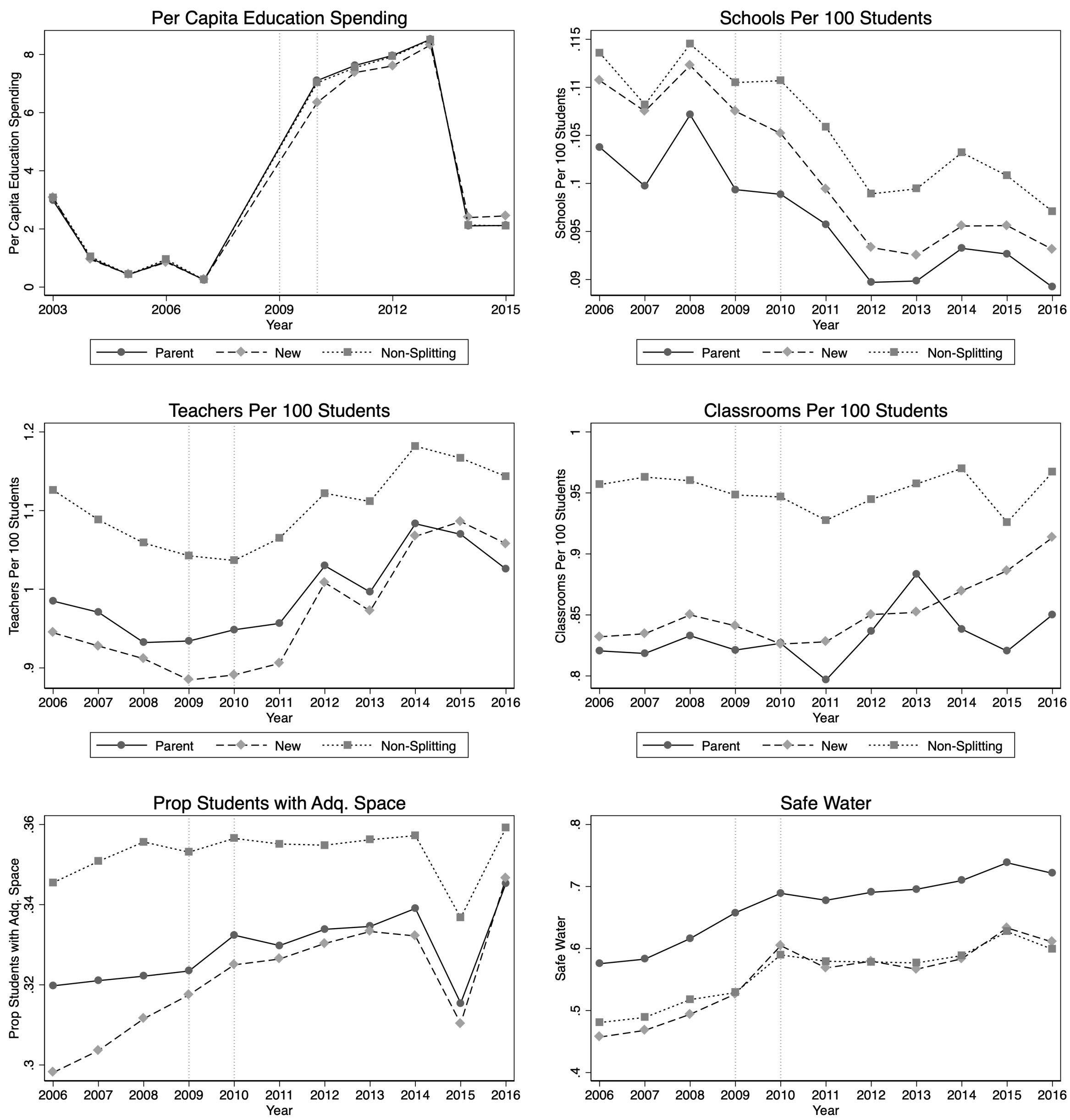


Figure A.4: Well-Being Effects Figures - Panel A

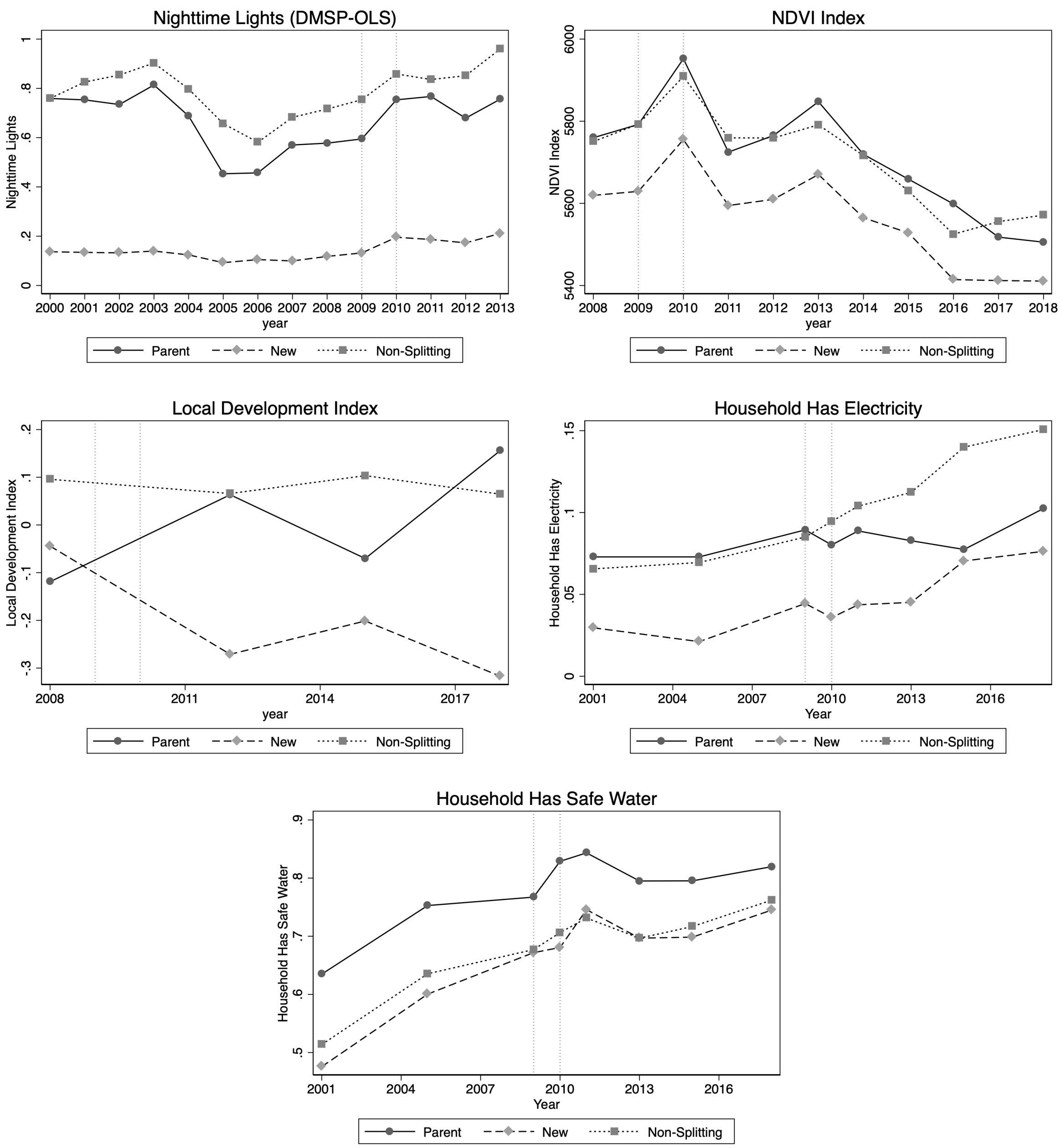


Figure A.5: Well-Being Effects Figures - Panel B

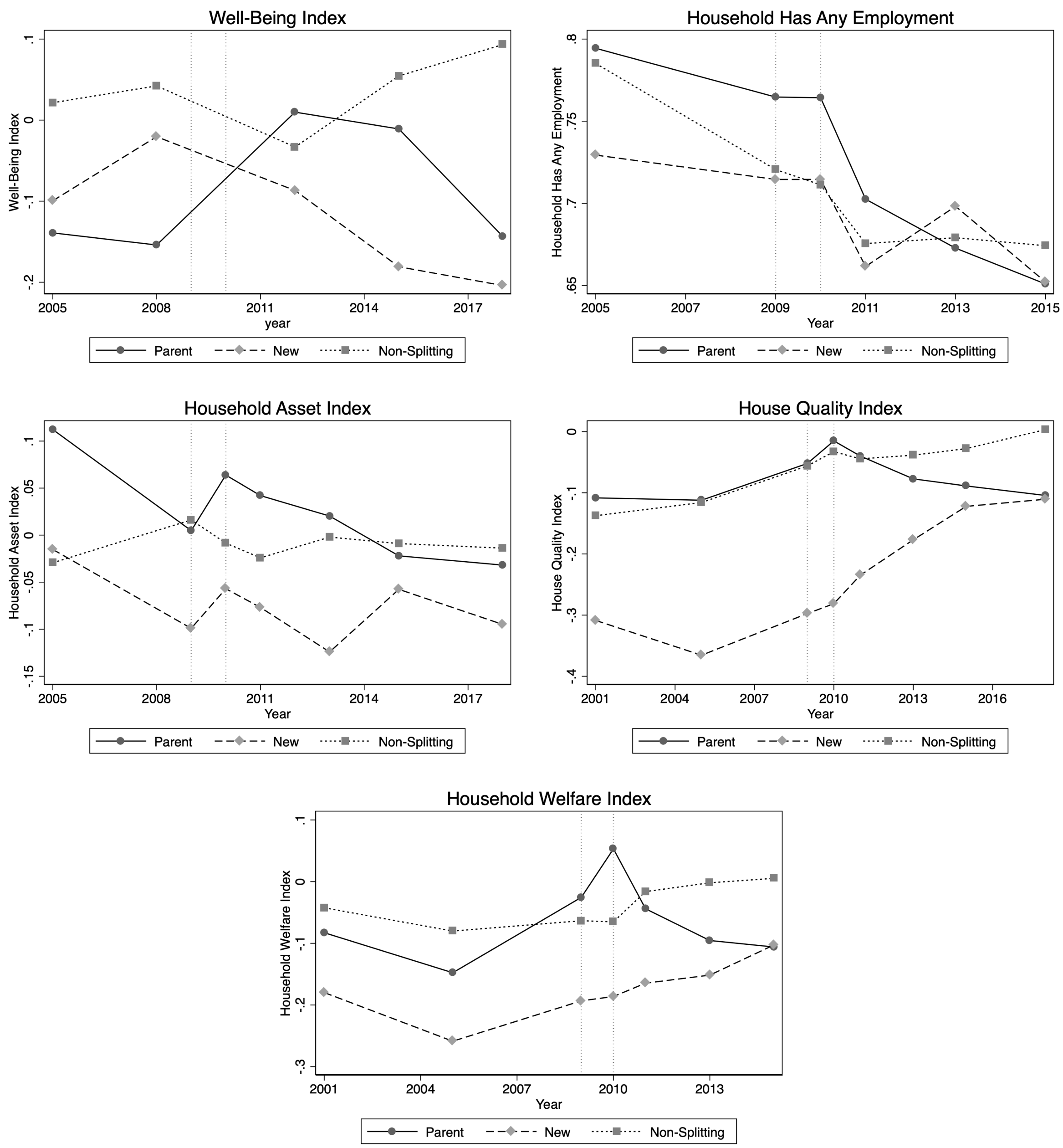


Figure A.6: Voting Effects Figures

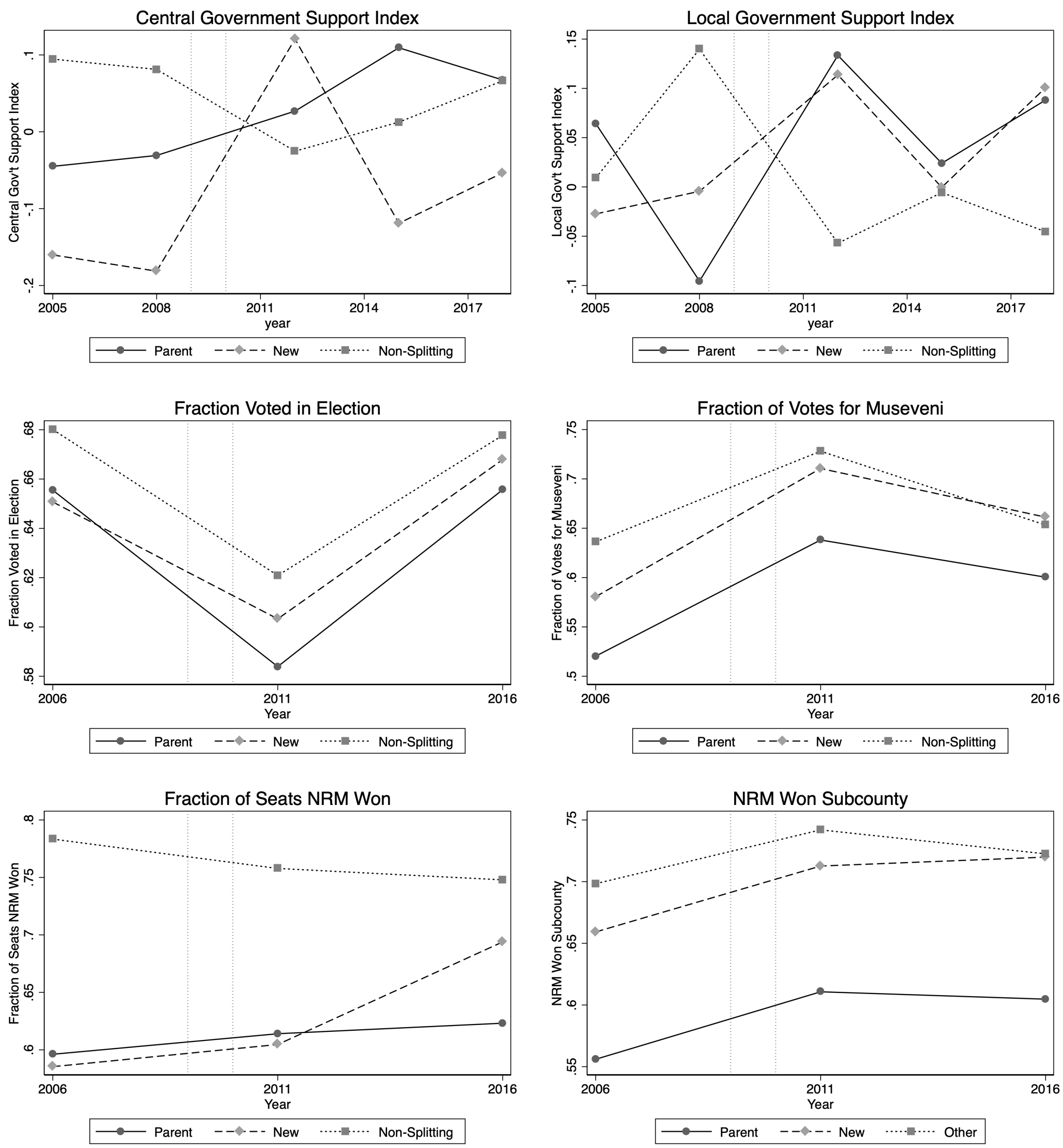
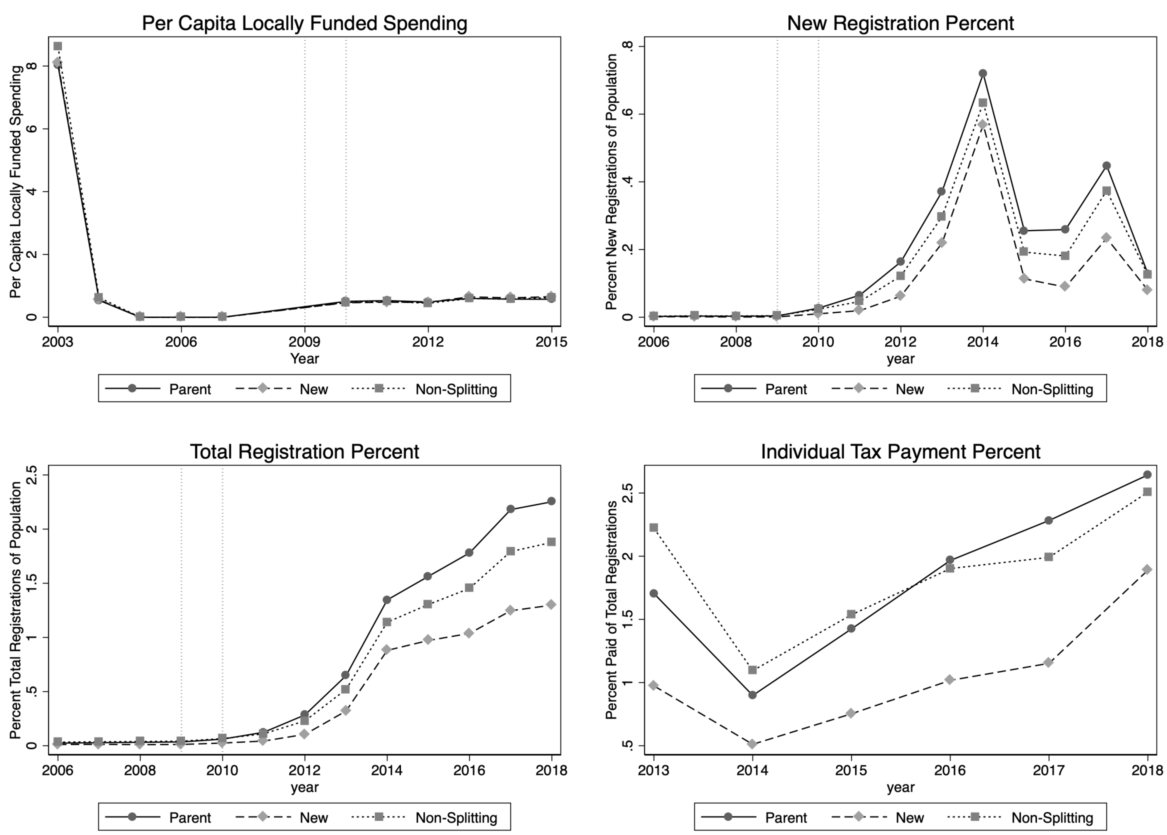


Figure A.7: Taxation Effects Figures



C.3 Data Appendix

Uganda National Panel Survey

I construct two variables and four indices from the Uganda National Panel Survey, measuring access to electricity and safe water, and indices of employment, assets, house quality and welfare. In this section, I detail the specific construction of these variables.

The variable Has Electricity$_{it}$ measures whether a houseshold indicates that they have access to electricity in a given survey round. For the 2001 retrospective and 2005 survey rounds, I construct this measure based on whether the household utilizes electricity for lighting, cooking or other activities. For 2009 onwards, I construct this measure using a question which specifically asks if the house has electricity.

The variable Safe Water$_{it}$ measures whether the household has access to a safe water source, defined as their main source of water being a private connection to pipeline, a public tap, borehole, or protected well or spring. Unsafe sources of water include an unprotected well or spring, river, stream, lake, or pond, or some method of rain water collection. A similar version of this question is asked in every year, including a retrospective dating back to 2001.

The variable Any Employment$_{it}$ is an annualized index measuring whether the household reports either wage employment or self-employment in a given year. This question was not asked retrospectively in 2001, and an equivalent was not asked in 2018; accordingly, this data is available from 2005 to 2018.

The variable Asset Index$_{it}$ is an annualized index measuring asset levels for a household starting in 2005 and ending in 2018. Specifically, it includes measures of whether the household owns a house; other building; appliances such as a kettle, flat iron, etc.; electronics such as a television, radio, cassette player, etc.; generator; solar panel, bicycle; motorcycle; other means of transportation; jewelry; and mobile phone. Each of these item is asked about in every survey round from 2005 to 2018. I normalize each relative to its own year, and sum the measure of normalized item ownership to construct an aggregate asset index. Note that prior to 2018, the survey asks if any member of the household owns the asset at present; in 2018, the question differentiates between owning individually and owning jointly. For consistency, I consider all forms of ownership as equivalent to answering yes in an earlier year.

The variable House Quality Index$_{it}$ is an annualized index measuring house quality starting in 2001 and ending in 2017. Specifically, it includes measures of whether the household has a sturdy roof, defined as a roof not made of thatch or mud; a sturdy floor, defined as a floor not made of earth; sturdy walls, defined as walls not made of thatch or mud; and any latrine, defined as any use of a toilet other than bushes. Each of these items is asked about in every survey round from 2005 to 2018, and retrospectively for 2001. I normalize each relative to its own year, and sum the measure of normalized house quality items to construct an aggregate house quality index.

The variable Welfare Index$_{it}$ is an annualized index measuring general household welfare starting in 2001 and ending in 2015. Specifically, it includes measures of whether every member of the houeshold has two sets of clothing; whether every member of the houeshold has at least one set of shoes; whether every child in the houeshold has a blanket; and the number of meals taken per day in the household. Each of these items is asked about in every survey round from 2005 to 2015, and retrospectively for 2001; however, these items do not appear to have been asked in the 2018 survey round. I normalize each relative to its own year, and sum the measure of normalized general welfare items to construct an aggregate welfare index.

Afrobarometer

I construct six indices from the Afrobarometer Survey, measuring support for the local government, central government, and president, as well as overall wellbeing, asset ownership, and local development at the enumeration area level. In this section, I detail the specific construction of these variables.

The variable Support for Central Government$_{it}$ is an index meassuring support for the central government. It includes whether a person indicates that they believe the central government handles each of the following matters fairly or very well, including: managing the economy, improving the living standards of the poor, creating jobs, keeping prices down, narrowing gaps between rich and poor, reducing crime, improving basic health services, addressing educational needs, providing water and sanitation services, ensuring everyone has enough to eat, fighting corruption in government, maintaining roads and bridges, and providing a reliable supply of electricity. It also includes variables indicating whether an individual indicated that they trusted the president of Uganda either somewhat or a lot, whether they trusted the parliament of Uganda either somewhat or a lot, and whether they rated members of parliament as listening to people either often or always. Finally, the variable includes measures of whether the individual approves or strongly approves of the performance of the president of Uganda and their own member or parliament, and whether they would vote for the national ruling party in a presidential election held tomorrow. Each of these items has been asked in every survey round from 2005 to 2018, with the exception of rating the maintenance of roads and bridges and providing a reliable supply of electricity, which are asked from 2008 on only. I normalize each relative to its own year, and sum the measures of central government support to construct an aggregate central government support index.

The variable Support for Local Government$_{it}$ is an index measuring support for the respondent's local government. It includes whether a person indicates that they believe the local government handles maintaining local roads and maintaining local market places either fairly well or very well, whether they indicated they trusted their local government either somewhat or a lot, whether they rated their local government as listening to people either often or always, and whether they approve or strongly approve of the general performance of their local government. Each of these items has been asked in every survey round from 2005 to 2018. I normalize each relative to its own year, and sum the measures of local government support to construct an aggregate local government support index.

The variable Index of Wellbeing$_{it}$ is an index of general well-being constructed from the Afrobarometer data. It contains measures of whether the household never, in the last year, lacked access to food, clean water, medicine, fuel and cash income. Each question was asked in every survey round from 2005 to 2018. I normalize each relative to its own year, and sum the measures of wellbeing to construct a general wellbeing index.

The variable Index of Local Development$_{it}$ is an index measuring the local development of the enumeration area where the survey is conducted. This measure is based on enumerator observations of the area in question, and specifically records whether the enumeration area is observed to have an electricity grid that most houses could access, piped water that most houses could access, a sewage system that most houses could access, and cell phone service. The enumerator also records whether each of the following facilities is either in the area or within walking distance: post office, school, police station, health clinic, and market stalls. Last, the enumerator records whether the road at the entry to the enumeration area was paved, tarred or concrete. Each of these questions is asked in every survey round from 2008 to 2018. I normalize each relative to its own year at the enumeration-area level, and sum the measures of local development to construct a development index.