UNIVERSITY OF COPENHAGEN
FACULTY OF SOCIAL SCIENCES
DEPARTMENT OF ECONOMICS

PhD thesis

Kasper Brandt

Essays in Development Economics:

Human Capital and Armed Conflicts

Advisor: Edward Samuel Jones

Handed in: April 30, 2020

Contents

Acknowledgements

Despite a joyful time in the Danish education system, it was not in the cards that I should write a PhD thesis. My career goals gradually changed from football player to shop assistant to bank clerk to business manager, and finally, development economist. I have met many inspiring teachers along the way guiding me towards this moment of submitting a PhD thesis. I owe a great deal of thanks to these teachers in my primary school, high school, and the university.

Above all, I wish to thank my supervisor, Sam Jones, for always being available during the first and most critical year of writing this thesis. There simply was not a time that he would not talk about the research projects I had in mind. Also, before taking a leave of absence in 2018, Sam did everything he could to feed me with new research ideas and suggestions for ongoing work.

Three of the four chapters in this thesis are joint work with other researchers. One chapter is co-authored with Sam. It taught me a lot, methodologically in particular, working on this project. A second chapter is co-authored with Beatrice Kalinda Mkenda. Beatrice has been invaluable in understanding the local context, gaining access to data, and setting up meetings and interviews. A third chapter is co-authored with two fellow PhD students, Gustav Agneman and Christoffer Pfeiffer Cappelen, and a data science expert, David Sjöberg. Working with this team has been a real pleasure and very insightful.

Further thanks go to the Department of Economics at the University of Copenhagen for giving me the opportunity to work on this thesis. I am grateful to the many colleagues at the department for participating in seminars and giving feedback when I have presented my work. Special thanks go to the members of the Development Economics Research Group (DERG) and John Rand, in particular, for arousing my interest in development economics research.

I thank all my fellow PhD students for making the time as a PhD student much more enjoyable than would have been the case without them. To many of them, I dare say they have become great friends both within and outside the work place. In particular, thanks go to my office mates and the other talented economists sitting in building 26.

Lastly, I am very fortunate to have a loving girlfriend, a supportive family, and good friends around me. Without them, the process of writing this thesis would have been substantially more onerous.

Kasper Brandt
Copenhagen, April 2020

Dansk introduktion

Denne ph.d.-afhandling består af fire selvstændige kapitler inden for udviklingsøkonomi. De tre første kapitler omhandler opbygning af humankapital i Østafrika, mens det fjerde kapitel omhandler måling af 'statskapacitet' på et lokalt niveau. Det første kapitel undersøger effekten på eksamensresultater af at gå i privatskole, i forhold til at gå i offentlig skole, i Tanzania. Eksamensresultater på det 9. klassetrin sammenlignes for elever som gik i den samme grundskole på den samme årgang og fik de samme eksamenskarakterer, men efterfølgende fortsatte nogle i offentlig skole og andre i privatskole. Det andet kapitel undersøger effekten af privatskoler på distriktsniveau. Det analyseres med data fra Kenya hvorvidt en stigning i andelen af børn som er privatskoleelever har en indflydelse på hvordan børnene gennemsnitligt klarer sig i kognitive tests. Det tredje kapitel undersøger hvordan en reform der eliminerede skolepenge i Tanzania har påvirket skoleindskrivning og læring. Det fjerde og sidste kapitel udvikler en metode til at måle 'statskapacitet' på et lokalt niveau i lande syd for Sahara. Dette mål bruges til at teste hvorvidt områder med høj 'statskapacitet' har lavere risiko for at opleve en væbnet konflikt.

Kapitel 1 – When Private Beats Public: A Flexible Value-Added Model with Tanzanian School Switchers

Det første kapitel undersøger hvorvidt et skifte fra en offentlig skole til en privatskole har betydning for elevers eksamensresultater i Tanzania. En lang række studier påviser at privatskoler i flere udviklingslande i gennemsnit opererer under de samme eller lavere omkostninger sammenlignet med offentlige skoler, hvorfor en potentiel læringseffekt er interessant at studere. Jeg bruger et unikt datasæt bestående af individuelle eksamensresultater fra det 7. og 9. klassetrin. Elevernes eksamensresultater på 9. klassetrin sammenlignes med andre elever, som gik i den samme skole på 7. klassetrin og fik de samme eksamensresultater på dette klassetrin. Forskellen mellem eleverne er dog at nogle fortsatte i en offentlig skole efter det 7. klassetrin, mens andre fortsatte i en privatskole. Desuden kontrolleres der for skolekammerateffekter og uobserverede faglige evner. For at belyse problemstillingen ved uobserverbar selektion, sammenlignes desuden elever som er på hver deres side af at dumpe 7. klassetrin. Såfremt en elev dumper, så kan vedkommende ikke søge videre til en offentlig skole på 8. klassetrin. Derved er flere elever nødsaget til at vælge privatskole på trods af de ønsker at komme i en offentlig skole. Elever som lige akkurat består 7. klassetrin kan frit vælge, og disse elever vælger i høj grad offentlige skoler.

Resultaterne af den empiriske analyse viser at privatskoler i høj grad forbedrer elevernes eksamenskarakterer. Eksamensbesvarelser vurderes af en central enhed som ikke har noget forhold til eleverne. I løbet af to års skolegang på 8. og 9. klassetrin forbedrer privatskoler i gennemsnit elevernes karakterer med 0.54 standarddeviationer. En analyse af hvorvidt andre faktorer driver denne effekt viser ingen tegn på at det skyldes ændringer i brug af privatlærer, arbejde ved siden af skolen eller sundhed. Ved at analysere forskellige fag enkeltvist, i stedet for et gennemsnit, viser at privatskoler i gennemsnit forbedrer eksamenskaraktererne i kiswahili, engelsk og matematik med henholdsvis 0.38, 0.48 og 0.69 standarddeviationer.

Kapitel 2 – Private Schools and Learning: A Dynamic Model of Learning in Kenya *med Sam Jones*

Det andet kapitel tager analysen af privatskolers læringseffekter fra det individuelle plan til distriktsniveau. Mens analysen på det individuelle plan fortæller hvorvidt den enkelte elev drager fordel af at gå i en privatskole, så tages der ikke højde for potentielle afsmittende effekter til andre børn. Vi gør brug af et datasæt bestående af 235.000 tests af børns kognitive evner over en femårig periode i Kenya. Baseret på disse tests beregner vi et underliggende mål for et barns faglige evner. Vi tager derefter et gennemsnit for hver årgang mellem 6 og 10 år i hvert distrikt i hvert af de fem år. Derefter undersøger vi hvordan en ændring i andelen af børn som går i privatskole påvirker det gennemsnitlige testresultat for gruppen. Dette gøres ved hjælp af system generalized method of moments (GMM) estimatoren, som hjælper os med at omgå endogenitetsproblemer i en simpel regressionsmodel.

Resultaterne viser at en stigning på 10 procentpoint i andelen af børn som går i en privatskole medfører en forventet stigning i gennemsnitlige faglige evner på 0.11 standarddeviationer. Dette er vel at mærke et gennemsnit for alle børn, inklusiv de 90 procent som ikke vil være påvirket af et skoleskifte. Mens denne effekt er særdeles robust i forhold til at ændre modellen på forskellige punkter, så er konfidensintervallet stort. Således dækker 95 procent konfidensintervallet over en effekt fra 0.04 til 0.17 standarddeviationer ved at øge andelen af børn i privatskole med 10 procentpoint. Desuden viser resultaterne at andelen af børn som aldrig har gået i skole har en stærk negativ effekt på gennemsnitlige faglige evner, samt at en stor del af den læring som opbygges i ét år ikke materialiserer sig i det følgende år.

Kapitel 3 – The Impacts of Eliminating Secondary School Fees: Evidence from Tanzania *med Beatrice Kalinda Mkenda*

Det tredje kapitel evaluerer en skolereform i Tanzania som eliminerede skolepenge på det 8. til 11. klassetrin. Det undersøges hvorvidt reformen påvirkede indskrivning og læring forskelligt for distrikter med initialt få elever på disse klassetrin relativt til distrikter med mange elever på disse klassetrin. Denne strategi bygger på at distrikter med initialt få elever har et større potentiale for at øge indskrivningen. Analysen benytter sig af et "difference-in-differences"-design, hvor vi sammenligner udviklingen fra før reformen til efter reformen for distrikter der i høj grad er påvirket af reformen med distrikter der i lav grad er påvirket af reformen. Denne strategi forudsætter at der er et klart skifte i "behandling" over tid. Eftersom reformen blev annonceret i februar 2015 så er årgangene som tog eksaminerne på det 9. klassetrin i 2013 og 2014 upåvirket af reformen.

Den empiriske analyse viser at reformen havde en stor påvirkning på indskrivning, hvilket især er drevet af distrikter med initialt få elever på det 9. klassetrin. Resultaterne understøtter desuden antagelsen om at årgangene før annonceringen af reformen ikke bliver påvirket, idet vi ikke kan afvise at udviklingen i indskrivning fra 2013 til 2014 er den samme for distrikter med initialt få elever relativt til distrikter med initial mange elever. Modellen forudsiger at distriktet på den 80. percentil, i forhold til graden af eksponering over for reformen, har haft en stigning i indskrivning på 12 procentpoint sammenlignet med distriktet på den 20. percentil. Desuden viser vi at den positive effekt på indskrivning skete på bekostning af læring, og at dette ikke kan forklares af lavere faglighed hos de elever som reformen påvirkede til at lade sig indskrive. Distriktet på den 80. percentil har haft en forventet nedgang i eksamensresultater på mellem 0.10 til 0.15 standarddeviationer sammenlignet med distriktet på den 20. percentil.

Kapitel 4 – Predicting Local State Capacity in Africa: A Machine Learning Approach *med Gustav Agneman, Christoffer Pfeiffer Cappelen og David Sjöberg*

Det fjerde kapitel udvikler en metode til at måle 'statskapacitet' på et subnationalt niveau. Vi konstruerer et indeks for 'statskapacitet' baseret på svar fra den samme spørgeskemaundersøgelse i forskellige lande. Derefter identificerer vi forklarende variable som er til rådighed for alle områder i Afrika syd for Sahara. Disse variable inkluderer rejsetid til hovedstaden, belysning om natten, nutidig og historiske befolkningstal, terræn, m.fl. Vi anvender disse forklarende variable til at forusige indekset ved hjælp af "machine learning"-modeller og ekstrapolerer indekset til hele Afrika syd for Sahara. Vi bekræfter validiteten af det ekstrapolerede mål ved hjælp af empiriske mål for historisk lokal magt, lokale etniciteters politiske magt, samt vaccinationstilslutning.

Vores mål anvendes til at undersøge om 'statskapacitet' mindsker risikoen for væbnet konflikt. Specifikt undersøger vi om ændringer i olieprisen har en forskellig påvirkning på risikoen for væbnet konflikt i olieområder kontra andre områder. Derefter inkluderer vi et tredobbelt interaktionsled mellem olieområde, oliepris, og 'statskapacitet'. Dette gør vi for at undersøge om områder med høj 'statskapacitet' har lavere risiko for væbnet konflikt som følge af en olieprisstigning sammenlignet med områder med lav 'statskapacitet'.

Det ekstrapolerede mål for 'statskapacitet' er blandt andet positivt korreleret med nattebelysning og befolkningstal, samt negativt korreleret med rejsetid til hovedstaden og øvrige byer, rejsetid per kilometer til hovedstaden, og skovdække. Dertil er målet positivt korreleret med tidligere anvendte mål for historisk lokal magt, lokale etniciteters politiske magt, samt vaccinationstilslutning. Analysen af væbnet konflikt indikerer at områder med høj 'statskapacitet' har lavere risiko for væbnet konflikt som følge af en olieprisstigning.

English introduction

This PhD dissertation consists of four self-contained chapters in the field of development economics. The topic of the first three chapters is human capital in East Africa, whereas the topic of the fourth chapter is measurement of sub-national state capacity. The first chapter studies the effect on exam scores from attending private school relative to attending public school in Tanzania. I compare secondary school exam scores for students who went to the same primary school, in the same cohort, and achieved the same primary school exam scores, but one went on to public secondary school and the other to private secondary school. The second chapter studies the effect of private schools at the district level. Based on data from Kenya, we analyse whether an increase in the share of children enrolled in private schools affect average performance in cognitive tests. The third chapter evaluates the impacts on enrolment and learning from a reform eliminating secondary school fees in Tanzania. The fourth and final chapter develops a method for measuring state capacity at a local level in Sub-Saharan Africa. This measure is used to test whether areas with high predicted state capacity have lower risk of armed conflict relative to areas with low predicted state capacity.

Chapter 1 – When Private Beats Public: A Flexible Value-Added Model with Tanzanian School Switchers

The first chapter examines whether a switch from a public to a private school influences exam scores for Tanzanian students. A large number of studies show that private schools in several developing countries, on average, are as cheap or cheaper to operate compared to public schools. Consequently, any potential learning effect is arguably driven by higher efficiency. I use a unique panel dataset consisting of individual exam records from grade 7 in primary school and grade 2 in secondary school. Exam scores in secondary school are compared to other students who went to the same primary school in the same year and achieved the same primary school exam scores. Some of the students, however, went on to public secondary schools, while others went on to private secondary school. The analysis further controls for peer effects and unobserved academic abilities. To elucidiate the issue of unobserved selection, students on either side of failing the primary school exams are compared. If students fail, they cannot enrol in public secondary school. Consequently, some students are obliged to choose private despite wanting public.

The results from the empirical analysis show that private schools substantially improve exam scores. The exam scores are assessed by a central unit without any relationship to the students. During two years of secondary education, private schools, on average, improve exam scores by 0.54 standard deviations. An analysis of potentially confounding factors shows no sign of the learning effect being driven by either private tuition, work hours, or student health. Comparing students on either side of passing the primary school exams gives similar results as the baseline analysis. Subject-specific analyses show that private schools, on average, improve exam scores in Kiswahili, English, and math by 0.38, 0.48, and 0.69 standard deviations, respectively.

Chapter 2 – Private Schools and Learning: A Dynamic Model of Learning in Kenya *with Sam Jones*

The second chapter takes the analysis of the private school learning effect from the individual level to the district level. While analyses from individual-level studies demonstrate whether one student benefits from attending private school, potential spillover effects are not accounted for. We make use of a dataset consisting of 235,000 tests of children's cognitive abilities over a five-year period in Kenya. Based on these tests, we derive an individual-level latent measure of academic ability and take averages for each cohort aged 6 to 10 years in each district in each of the five years. Next, we examine how a change in the share of children attending private school affects average academic ability. To counter endogeneity issues, we employ the system generalized method of moments (GMM) estimator.

The results show that an increase of 10 percentage points in the share of children attending private school leads to an expected rise in average latent ability of 0.11 standard deviation. This is an average for all children, including the 90 percent who are not affected by switching school. While this effect is highly robust to changing key aspects of the analysis, the confidence interval is wide. The 95 percent confidence interval covers an effect from 0.04 to 0.17 standard deviations from increasing the share of students attending private school by 10 percentage points. The results, in addition, demonstrate that a higher share of students never having attended school causes average ability to drop, and a large part of what is learnt during one year does not materialize in the following year.

Chapter 3 – The Impacts of Eliminating Secondary School Fees: Evidence from Tanzania *with Beatrice Kalinda Mkenda*

The third chapter evaluates a school reform in Tanzania, which eliminated secondary school fees. We examine whether the reform had different impacts on enrolment and learning for districts with low pre-reform progression rates to secondary education relative to districts with high pre-reform progression rates. This strategy relies on districts with low pre-reform progression rates having higher potential for growth. The analysis makes use of a difference-in-differences design, where we compare the post-reform situation with the pre-reform situation for districts highly exposed to the reform relative to districts less exposed to the reform. The strategy assumes that there is a strict shift in treatment over time. Since the reform plans were announced in February 2015, cohorts taking the secondary school exams in 2013 and 2014 were unaffected.

The empirical analysis demonstrate that the reform had a massive impact on progression rates to secondary education, in particular for districts with low pre-reform rates. The results support the assumption regarding cohorts taking the secondary school exams before the announcement being unaffected, as we cannot reject the development in progression rates between 2013 and 2014 being the same for highly exposed and less exposed districts. The model predicts that the district at the 80th percentile, in terms of exposure to the reform, had a rise in the progression rate of 12 percentage points relative to the district at the 20th percentile. This positive enrolment effect came at the expense of learning, which cannot be explained by lower academic ability of students induced by the reform to progress. The district at the 80th percentile has had an expected drop in exam scores of 0.10 to 0.15 standard deviations relative to the district at the 20th percentile.

Chapter 4 – Predicting Local State Capacity in Africa: A Machine Learning Approach *with Gustav Agneman, Christoffer Pfeiffer Cappelen, and David Sjöberg*

The fourth chapter develops a method for measuring state capacity at a subnational level. We construct an index of state capacity based on answers to the same survey in different countries. Next, we identify different explanatory variables that are freely available for all sub-national areas of Sub-Saharan Africa. These include travel time to the capital, night-time light emission, contemporary and historic population figures, terrain, and more. We use the explanatory variables to predict state capacity in a machine learning approach and extrapolate the index to all of Sub-Saharan Africa. We corroborate the validity of the extrapolated index by correlating it with empirical measures of pre-colonial centralization, contemporary political power of local ethnicities, and vaccination coverage.

Our measure is used to examine whether state capacity lowers the risk of armed conflict. Specifically, we investigate whether changes in the oil price has a different effect on the risk of armed conflict in areas with oil deposits relative to other areas. After that, we include a triple interaction term between oil deposits, oil price, and state capacity. We pursue this strategy to examine whether areas with high predicted state capacity have lower risk of armed conflict following an oil price increase relative to areas with low predicted state capacity.

The extrapolated measure of state capacity correlates positively with night-time light emissions and population figures, and negatively with travel time to the capital and other cities, travel time per kilometer to the capital and forest cover. In addition, the measure correlates with pre-colonial centralization, contemporary political power of local ethnicities, and vaccination coverage. The analysis of armed conflict suggests that areas with high predicted state capacity have lower risk of armed conflict due to an oil price increase.

Chapter 1

When Private Beats Public: A Flexible Value-Added Model with Tanzanian School Switchers

When Private Beats Public:

A Flexible Value-Added Model with Tanzanian School Switchers

By Kasper Brandt*

This paper estimates a private school learning premium in Tanzania, using unique administrative data on national exams in primary and secondary school for 635,000 students. A value-added model compares secondary school exam scores for students with the same primary school exam scores from the same primary school in the same year. The model further controls for peer effects and unobserved ability. On average, private schools improve exam scores by 0.54 standard deviations in two years. An instrumental variable model suggests the effect is causal. Subject-specific estimates for Kiswahili, English, and mathematics are 0.38, 0.48, and 0.69 standard deviations, respectively. (*JEL* I21, I24, O15)

*Department of Economics, University of Copenhagen, Øster Farimagsgade 5 DK 1353 Copenhagen K (e-mail: kasper.brandt@econ.ku.dk). A previous version of this paper is published in the WIDER working paper series under the name "Private beats public: A flexible value-added model with Tanzanian school switchers". The paper has benefited from valuable discussions with Sam Jones, Paul Glewwe, Abhijeet Singh, Elias Papaioannou, Gordon Dahl, Eric Hanushek, Margaret Raymond, Lant Pritchett, and members of the Development Economics Research Group at University of Copenhagen. Further thanks go to seminar participants at numerous conferences, and the National Examinations Council of Tanzania for providing public access to a large and useful database on student exam records.

I. INTRODUCTION

For the first time in 2008, the primary school gross enrolment rate reached 100 percent in low-income countries (UNESCO 2020). While getting children to school is arguably necessary for learning, it is not sufficient. As emphasized in the Sustainable Development Goal 4, we seek not only to provide education for children. We seek to provide *quality education*. It is estimated that 617 million children and adolescents in the primary and lower secondary school age are not achieving a proper competence level in reading and mathematics (UN DESA 2018). Finding better solutions for improving learning is therefore essential to meet the goal of quality education for all.

The educational sector generally consists of both public and private providers, and the optimal sizes of the two are often debated. While this debate continues to be raised in developed and developing countries, the potential contribution of privately run schools has been noted for low-income countries in particular. This is for two main reasons. First, in these contexts, popular models of private schools have been shown to deliver education at a similar or lower per-student cost (Psacharopoulos 1987; Jimenez, Lockheed, and Paqueo 1991; Lassibille, Tan, and Sumra 2000; Alderman, Orazem, and Paterno 2001; Andrabi, Das, and Khwaja 2008; Schirmer 2010; Tooley et al. 2011; Bold et al. 2013; Ngetich, Wambua, and Kosgei 2014; Muralidharan and Sundararaman 2015). Potential explanations include lower teacher salaries as private schools may hire from a different pool of applicants, lower prices for food and equipment due to better incentives to negotiate, and lower costs of renting buildings due to lack of regulations. Second, private schools are generally found to improve student test scores compared to public schools (Angrist et al. 2002; Andrabi et al. 2011; Muralidharan and Sundararaman 2015; Singh 2015; Barrera-Osorio et al. 2017).[1]

[1] Learning gains from attending a private school are referred to as the private school learning premium.

While a positive impact of private schools on learning outcomes is generally acknowledged in the literature, substantial caveats remain (Day Ashley et al. 2014). Most importantly, the applied methodologies often rely on cross-sectional ordinary least squares (OLS) models, thereby assuming past school inputs have no effect on test scores.[2] Secondly, the magnitude of the estimated learning premiums varies substantially between studies, highlighting the importance of context, and evidence is to a large extent based on data from South Asia. Lastly, when studying the desirability of private schools, Urquiola (2016) argues one must disentangle the potential channels through which private school students perform differently. For instance, higher achievement in private schools could merely reflect a zero-sum mechanism like peer effects.

In this paper, I set up a value-added model to estimate the learning effects of attending a private school, using a new dataset of 635,000 Tanzanian students progressing from primary to secondary school based on administrative exam records. Often due to data limitations, standard value-added models are forced to impose four assumptions concerning the learning-generating process: 1) inputs to learning are non-age varying; 2) inputs to learning are additively separable; 3) learning effects of family-supplied and school-supplied inputs decay at the same rate; and 4) learning effect of unobserved ability decay at the same rate as the supplied inputs. By comparing students from the same primary school with the same primary school exam scores in the same year, and including a proxy for unobserved ability, this paper is able to relax the latter three assumptions. Essentially, the paper estimates the learning effect of a binary treatment variable, private school enrolment, while controlling for a large set of student subgroups, peer effects, unobserved ability, and a measure of

[2] Due to the importance of lagged achievement (Andrabi et al. 2011; Deming 2014; Muralidharan and Sundararaman 2015; Elks 2016) and non-random educational investments among siblings (Dizon-Ross 2019), studies applying simple OLS models, household fixed effect models, and propensity score matching without accounting for lagged achievement are not considered in this paper.

student personality.

Despite comparing students performing similarly in primary school and including a proxy for unobserved ability, selection into private education could be an issue. For instance, two students from the same public primary school achieve the same primary school exam scores, and they progress to two different types of secondary school. If unobserved family characteristics matter more in secondary school relative to primary school, the estimated learning premium of private schooling is likely biased upward. On the other hand, if the student continuing in public school was able to achieve the same primary school exam scores with e.g. less parental support, this student likely has higher unobserved ability, which would bias the estimated learning premium of private schooling downward. To acknowledge the potential for selection into private education, various robustness analyses are performed, including estimating an instrumental variable (IV) model, testing for the risk of confounding factors, restricting the sample in different ways, and examining several dimensions of heterogeneity. The IV model exploits that students failing the primary school exams are forced into private secondary education, whereas students barely passing can enter public schools. Most students failing the primary school exams, however, exit the education system. Consequently, the failing students moving to private secondary schools could still be fundamentally different from students barely passing and moving to public secondary schools. If, however, some of those students who failed primary school are forced into private education, one would expect a correction towards zero if the baseline private school learning premium is believed to be biased upward due to selection.

The results provide evidence for a large private school learning premium in Tanzania. The preferred model demonstrates that attending a private secondary school, instead of a public school, leads to a 0.54 standard deviation increase, on average, in exam scores after two years of secondary education. An IV model further suggests the relationship is causal, at least for low-performing students who comply with the assignment to private school treatment. An analysis of potentially confounding factors shows no critical evidence of households changing behaviour when students change type of school. The only significant result is that households spend *less* on private tuition when children move from public to private education. Subject-specific analyses find that private schools, on average, increase students' exam scores by 0.38, 0.48, and 0.69 standard deviations in Kiswahili, English, and mathematics, respectively.

The paper contributes to the literature on the learning effects of private schools in developing countries. Previous studies have found positive learning effects of private schools mostly in India and Pakistan (Andrabi et al. 2011; Muralidharan and Sundararaman 2015; Singh 2015; Barrera-Osorio et al. 2017). The main points from these studies are that students in rural areas seem to benefit more from attending a private school, the learning premium tends to be larger for test scores in English, and even students switching to low-cost private schools can outperform their peers with higher socio-economic status in public schools. The private school learning premium in Africa, however, is less clear as most studies do not account for lagged test scores. Three studies from Tanzania provide mixed evidence for the private school learning premium (Psacharopoulos 1987; Jimenez, Lockheed, and Paqueo 1991; Lassibille and Tan 2001). Much has changed, however, in the education system since the 1980s and mid-1990s when the data was collected for these three studies.

In addition to estimating the private school learning premium in an understudied region, the paper adds to the general literature using value-added models by proposing a more flexible value-added model than is usually employed. It is flexible in the sense that students are compared only to other students with the same lagged test scores from the same school, whereas standard value-added models include only the lagged test scores. While the literature on validating value-added models has advanced considerably by comparing value-added estimates to lottery-based estimates (Angrist et al. 2017) and regression discontinuity estimates (Singh 2019), the current paper finds that inclusion of primary school fixed effects is critical in the Tanzanian context. The private school learning premium is substantially biased upward in both a pooled OLS model (bias of 155 percent) and a standard value-added model (bias of 84 percent). One potential explanation for the importance of the flexible value-added model in the current setting is that the primary school fixed effect captures local market effects and socio-economic status of parents.

Section II considers the theoretical framework in regard to the standard value-added model and the current paper's flexible value-added model. Section III describes the Tanzanian education system and the data at hand. Section IV outlines the empirical methods, while Section V presents the baseline regression results and discusses the robustness of the results. Section VI concludes.

II. Theoretical framework

Todd and Wolpin (2003) present a cumulative learning production function given by Equation 1. That is, a student's test score at age a is dependent on inputs at all ages up to and including age a. Under a different set of assumptions, the OLS model and the standard value-added model can be derived from this general production function.

$$T_{ia} = T_a[F_i(a), S_i(a), \mu_i, \varepsilon_{ia}]. \tag{1}$$

The outcome variable is test score for student i at age a. The right-hand side of the equation consists of family-supplied inputs, F, school-supplied inputs, S, a time-invariant parameter measuring unobserved ability, and an idiosyncratic error term. The impacts of these explanatory variables are initially allowed to vary for different ages.

The OLS model provides unbiased estimates as long as only current inputs affect test scores and all relevant inputs are available as control variables. Thus, school inputs from previous years are assumed not to influence a student's current test score. This assumption implies lagged achievement is not a relevant explanatory variable for current achievement, which is generally rejected.

When employing the standard value-added model, one needs to make four key assumptions. First, the arguments in the cumulative learning production function are additively separable. Second, while the coefficients on inputs may differ depending on time since exposure, they are independent of the actual age of the child. Applying these first two assumptions, the following expression for student i's test score at age a can be derived:

$$\begin{aligned} T_{ia} =& F_{ia}\varphi_1 + F_{i,a-1}\varphi_2 + \dots + F_{i,a=1}\varphi_a + S_{ia}\alpha_1 + S_{i,a-1}\alpha_2 + \dots + S_{i,a=1}\alpha_a \\ &+ \beta_a\mu_i + \varepsilon_{ia}. \end{aligned} \tag{2}$$

Next, the lagged test score multiplied by the rate of decay parameter is subtracted on both sides of Equation 2, which gives Equation 3:

$$\begin{aligned} T_{ia} - \gamma T_{i,a-1} =& F_{ia}\varphi_1 + F_{i,a-1}(\varphi_2 - \gamma\varphi_1) + \dots + F_{i,a=1}(\varphi_a - \gamma\varphi_{a-1}) \\ &+ S_{ia}\alpha_1 + S_{i,a-1}(\alpha_2 - \gamma\alpha_1) + \dots + S_{i,a=1}(\alpha_a - \gamma\alpha_{a-1}) \\ &+ (\beta_a - \gamma\beta_{a-1})\mu_i + \varepsilon_{ia} - \gamma\varepsilon_{i,a-1}. \end{aligned} \tag{3}$$

In order to derive the standard value-added model, one needs to impose a third and a fourth assumption. Third, learning effects from different inputs to the learning process decay at the same rate over time. For instance, the learning effects from parental involvement and a better teacher at a specific age decay at the same rate. Fourth, the impact by unobserved ability must decay at the same rate as the effects from school and family inputs. Thus, γ must meet the following condition:

$$\gamma = \frac{\varphi_{j+1}}{\varphi_j} = \frac{\alpha_{j+1}}{\alpha_j} = \frac{\beta_a}{\beta_{a-1}}\ ,\ \forall j \in \{1,2,\dots,a\} \tag{4}$$

After imposing the above-mentioned assumptions to Equation 3, the standard value-added model given by Equation 5 is derived:

$$T_{ia} = F_{ia}\varphi_1 + S_{ia}\alpha_1 + \gamma T_{i,a-1} + \eta_{ia}, \tag{5}$$

where $\eta_{ia} = \varepsilon_{ia} - \gamma\varepsilon_{i,a-1}$, φ_1 identifies the effects from current family-supplied inputs, α_1 identifies the effects from current school-supplied inputs, and γ is the persistence parameter measuring the rate of decay of learning effects from past inputs. As the arguments in the production function are assumed to be additively separable, unobserved ability has no effect on returns to inputs.

The current paper proposes a method to improve the standard value-added model by relaxing three of the four assumptions made above. Table 1 outlines the assumptions needed in the standard value-added model and in the proposed value-added model.

First, by including lagged school fixed effects into the model, previous school-supplied inputs become redundant. In Equation 3, this is equivalent to replacing previous school inputs with a lagged school fixed effect, meaning all learning effects from previous time periods are captured in the lagged school fixed effect.[3] This model assumes – conditional on test scores – students benefit

[3] When more than one cohort is examined, the lagged school fixed effect ought to be replaced by a lagged school times cohort fixed effect to account for potentially varying school inputs over time.

similarly from the same school inputs, and students in the same school at age $a-1$ received the same school inputs since beginning of school. A violation of these assumptions (e.g. due to teacher favouritism or private tutoring) will arguably lead to a negative bias on the private school learning premium if students receiving better school inputs are more likely to switch to private education in the following time period. The reasoning is that the model accounts for lagged test scores. Thus, comparing students with similar academic ability, the student who received better school inputs was either not able to make use of them or has lower unobserved ability. Consequently, the student who remains in a public school arguably has higher unobserved academic ability and should be expected to perform better in the following time period if students receive the same inputs. Equation 6 presents the lagged school fixed effect model.

$$\begin{aligned} T_{isa} =& F_{ia}\varphi_1 + F_{i,a-1}(\varphi_2 - \gamma\varphi_1) + \ldots + F_{i,a=1}(\varphi_a - \gamma\varphi_{a-1}) + S_{ia}\alpha_1 + \delta_{s,a-1} \\ & + \gamma T_{is,a-1} + (\beta_a - \gamma\beta_{a-1})\mu_i + \varepsilon_{isa} - \gamma\varepsilon_{is,a-1}, \end{aligned} \tag{6}$$

where $\delta_{s,a-1}$ is the lagged school fixed effect for school s. The coefficient associated with current school inputs, α_1, can now be identified independently of the value of the persistence parameter, and the assumption of similar rates of decay for school- and family-supplied inputs can be dropped.

As emphasized above, the effect from unobserved ability is not neutralized if $\gamma \neq \beta_a/\beta_{a-1}$. Including a proxy for unobserved ability allows the rate of decay for unobserved ability to differ from the rates of decay for the inputs. Finally, in Equation 3 it may be that the αs differ for specific types of students. For instance, high-ability students may receive more attention or they may simply be better at benefiting from the given inputs. A famously known example is Glewwe, Kremer, and Moulin (2009), showing that providing free textbooks in rural Kenya only had an effect on

Table 1: *Assumptions needed to provide unbiased coefficient estimates*

Core assumptions	Standard value-added model	Flexible value-added model
Arguments in the cumulative learning production function are additively separable.	✓	✗[1]
Coefficients on school and family inputs are non-age-varying.	✓	✓
Learning effects from school and family inputs decay at the same rate over time.	✓	✗
Impact of unobserved ability decays at the same rate as school and family inputs.	✓	✗
Students with the same test scores in the same school at age $a-1$ received the same school inputs since beginning of school.	✗	✓[2]

Notes: [1]Unobserved ability is allowed to influence returns to inputs. [2]The significance of this assumption can to some extent be tested by comparing a full sample model to a restricted sample model consisting of students who in the previous period attended a small school where all students most likely went to the same class.

Source: author's own.

high-ability students. That is, ability could influence the return to inputs, thereby rejecting the additive separability assumption. To counter this issue, the current paper proposes to include a lagged school $\times$ lagged achievement fixed effect. This ensures students are compared only to other students benefiting similarly from the same school inputs, assuming students with the same achievement from the same school receive the same school inputs. Thus, the proposed value-added model takes the following form:

$$\begin{aligned} T_{isga} =& F_{ia}\varphi_1 + F_{i,a-1}(\varphi_2 - \gamma\varphi_1) + \ldots + F_{i,a=1}(\varphi_a - \gamma\varphi_{a-1}) + S_{ia}\alpha_1 + \theta_{sg,a-1} \\ &+ \mu_i(\beta_a - \gamma\beta_{a-1}) + \eta_{isga}, \end{aligned} \tag{7}$$

where $\eta_{isga} = \varepsilon_{isga} - \gamma\varepsilon_{isg,a-1}$. Subscript g refers to achievement level, meaning $\theta_{sg,a-1}$ is a lagged school × lagged achievement fixed effect. When more than one cohort of students are available, the lagged school × lagged achievement fixed effect is replaced by a lagged school × lagged achievement × cohort fixed effect.

In the remainder of the paper, a simple OLS model, the standard value-added model (Equation 5), the school fixed effect model (Equation 6), and the flexible value-added model (Equation 7) are estimated with administrative exam records from Tanzania.

III. Context and data

Tanzanian educational context

The education system in Tanzania consists of seven years of primary education followed by four years of secondary education. Next, students have the opportunity to either continue their studies for two years of advanced secondary education or take a 2–3 year technical or vocational education. Students passing advanced secondary school may continue to university. The current paper focuses on the first two years of secondary education, called lower secondary.

The school system experienced a major reform in 2001, when primary school tuition fees were abolished. This led to a surge in the primary school gross enrolment rate, peaking at almost 107 percent in 2008. The gross enrolment rate since gradually declined, and in 2015 it reached 85 percent (World Bank 2019). Public secondary education remained partially funded by school fees until the implementation of Tanzania's Education and Training policy in 2016 (Human Rights Watch 2017). In addition to the abolition of school fees in 2016, secondary schools were no longer obliged to teach in English.[4] While primary education became universal during the 2000s, secondary school attendance only recently started to pick up. In 2017, after the abolition of secondary school fees, 71 percent of students in the final grade of primary school progressed to secondary school (World Bank 2019). An increase of 15 percentage points from 2012. This number, however, is expected to increase substantially over the next decade, meaning further investments in education is needed in the magnitude of 67 to 170 percent of the current secondary education budget (Baum and Cilliers 2018). In order to meet the demand for further investments, targeted private school vouchers are argued to provide a potentially cost-effective delivery of education to future cohorts.

Compared to Uganda, Burundi, Rwanda, and Kenya, Tanzania has the second-lowest progression rate to secondary school. The progression rate is lower in Uganda, slightly higher in Burundi and Rwanda, and almost 100 percent in Kenya. Also, the use of private schools differs among the East African countries. While the share of students attending private schools in primary education is similar to the shares in Rwanda and Burundi, almost 20 percent of primary school students attend private schools in Uganda and Kenya. At the secondary school level, the private school share is in the middle of the East African countries (World Bank 2019).

Data sources

Section II presented a strategy that relaxes three of the four key assumptions needed in standard value-added models. For this strategy, the applied data must include student lagged and present achievement, information on current schoolmates' lagged achievement, together with information on previous schoolmates' lagged achievement, present achievement, and present school enrolment.

The current paper relies on two main data sources: 1) individual exam records from the Primary

[4] Except for the exam in Kiswahili, exams are still conducted in English.

School Leaving Examination (PSLE) in 2013, 2014, and 2015; and 2) individual exam records from the Form Two National Assessment (FTNA) in 2015, 2016, and 2017. The PSLE takes place after seven years of primary education, whereas the FTNA takes place after two years of secondary education. Thus, students taking the FTNA in 2015 are expected to have taken the PSLE in 2013 or before. Neither of the exams depend on student-teacher relations.[5] The years are chosen due to availability, and only lower secondary exam scores are analysed, as names of students are not available at the exams in Form 4 and Form 6.[6] While the data tells which school a student is enrolled at by the time of the FTNA, it cannot be ruled out that private secondary schools recruit high-ability (uncaptured by the PSLE) public school students, right before the exams to boost average GPA.

The names of the students in the FTNA are merged with the names of the students in the PSLE two years before. The sample is thereby restricted to students completing lower secondary school on time. One could suspect that private schools encourage academically weaker students to repeat grades in order to boost average GPA. Doing the analysis with a three-year gap between the PSLE and the FTNA, the private school learning premium increases slightly, suggesting grade repetition is not driving the results.[7] As students cannot retake the PSLE, the delay is caused either by a gap year or repeating a grade in lower secondary school. Students who have the same names as other students in either the FTNA or the PSLE two years before are removed from the analysis. For students

[5] Both exams are administered by The National Examinations Council of Tanzania (NECTA). During transport the exam documents are guarded by police and two NECTA employees. Together with external supervisors, Heads of Schools store the exam questions in a locked room until the day of the exam. During the exam, appointed invigilators ensure students adhere to examination rules and regulations. After the exam, answers are collected and stored in regional headquarters until NECTA employees pick them up. Lastly, the exam papers are sent to independent Marking Centres, where they are assessed and marked.

[6] The FTNA is not as high stake as the exams in Form 4 and Form 6. There is, therefore, a risk that schools differ in their relative focus on different exams. For instance, if public schools give more weight to the high-stake exams compared to private schools, a positive estimate of the private school learning effect captures that public schools do not focus on the FTNA.

[7] Results are available upon request.

taking the PSLE, 2.0 percent have a name duplicate within the year of taking the exam, whereas the same number for FTNA students is 1.5 percent. Excluding students based on their names could bias the results if students with common names are distinguishable from other students. A recent study finds that people with common names have more collectivistic behaviour, whereas people with unique names are more individualistic (Knudsen 2019). This constitutes a bias problem under the assumptions that individualistic students learn faster (or slower) and are more (or less) likely to enrol in a private school compared to collectivistic students. It further constitutes a selection problem as a small fraction of students with common names are excluded from the sample. To counter this issue, an indicator for name uncommonness – being one if less than 20 students out of a million bear the name – is included as a control variable.[8]

After excluding students with a name duplicate, the names of FTNA students are merged with the names of PSLE students from two years before. Of all FTNA students with no name duplicate in the FTNA nor in the PSLE, 51.6 percent are uniquely identified in the PSLE. A falsification test is performed, in which FTNA students are merged with PSLE students from one year before. As students are not expected to complete two years of secondary education in one year, this exercise should result in substantially fewer uniquely merged students. Yet, despite having no name duplicate within a specific year, students may have name duplicates over time. Thus, there could be a low number of students being merged between the FTNA and the PSLE from one year before. In line with expectations, only 1.8 percent of FTNA students without a name duplicate are uniquely merged with a PSLE student from the year before. The merging of names results in a dataset of approximately 635,000 students observed in two time periods. The next subsection on descriptive

[8] Robustness analyses further calculate and apply sample weights to achieve a representative sample, and recode the indicator for name uncommonness to being one if maximum three students within the region bear the name. These approaches do not change the results (see Appendix Table B1).

statistics discusses how the sample students differ from the total population of students, while a robustness analysis formally accounts for non-representativeness.

Information on school ownership is obtained by the President's Office, Regional Administration and Local Government (PO-RALG). School heads complete a form asking about ownership type and send it to the Ward Education Coordinator's office (WEC) for verification. Next, the WEC sends the information to the District Executive Director's office, where the information is entered into the Basic Education Information System. Finally, PO-RALG presents the data to the public.

In addition to the exam records, the National Panel Surveys from 2011, 2013, and 2015 are utilized to analyse the risk of confounding factors and identify regional consumption growth rates.

Descriptive statistics

Table 2 presents the descriptive statistics for the variables of interest. Columns (1) and (2) present population means and standard deviations, respectively. The variables *GPA PSLE*, *GPA PSLE other*, and *Private primary* are based on exam records from 2,330,649 primary school students taking the PSLE in 2013, 2014, or 2015. The remaining variables are based on exam records from 1,245,860 secondary school students taking the FTNA in 2015, 2016, or 2017. Next, columns (3) and (4) present the means and standard deviations for the applied sample. All variables are based on the same 635,162 students. Finally, columns (5) and (6) split the applied sample into private secondary school students and public secondary school students. In addition to the descriptive statistics presented in Table 2, the regional distribution of the sample and the distribution of subject-specific grades can be seen in Appendix Table A1 and Appendix Figure A1, respectively.

In all specifications tested, the dependent variable *GPA FTNA* refers to student achievement at the FTNA after two years of secondary education. It is defined as the grade point average (GPA)

Table 2: *Descriptive statistics for all students and the applied sample*

	Pop. mean	Pop. std.	Sample mean	Sample std.	Private mean	Public mean
GPA FTNA	1.308	0.881	1.318	0.886	2.407	1.198
GPA PSLE	1.664	0.833	2.229	0.696	2.758	2.170
GPA PSLE other	1.713	0.780	2.229	0.605	2.434	2.206
Private Primary	0.035	0.184	0.067	0.250	0.465	0.023
Private Secondary	0.182	0.386	0.100	0.299	1.000	0.000
Female	0.516	0.500	0.524	0.499	0.549	0.521
Uncommon name	0.148	0.355	0.137	0.344	0.139	0.137
Peers PSLE	2.227	0.425	2.228	0.416	2.758	2.170
Peers failed (PSLE)	0.029	0.099	0.013	0.049	0.068	0.007
Peers with A (PSLE)	0.177	0.212	0.172	0.210	0.490	0.137
Cohort 2016	0.325	0.468	0.312	0.463	0.324	0.311
Cohort 2017	0.388	0.487	0.410	0.492	0.367	0.415
Religious courses	0.282	0.450	0.264	0.441	0.423	0.247
Bible course	0.085	0.279	0.070	0.255	0.319	0.042
Islam course	0.230	0.421	0.228	0.420	0.120	0.240
Girls Only Secondary	0.034	0.182	0.031	0.173	0.203	0.012
Boys Only Secondary	0.017	0.131	0.016	0.125	0.083	0.009
N	See notes		635,162		63,206	571,956

Notes: 'Pop. mean' refers to the population mean. Population means of *GPA PSLE*, *GPA PSLE other*, and *Private primary* are based on 2,330,649 primary school students. The population means of the remaining variables are based on 1,245,860 secondary school students. 'Pop. std.' refers to the standard deviation of all students. Sample mean refers to the applied sample. The last two columns provide mean values for sample students attending private and public secondary schools, separately. Regional distribution of the sample and distribution of subject-specific grades can be seen in Appendix Table A1 and Appendix Figure A1, respectively. *Religious courses* is an indicator for whether the student's secondary school provides elective courses in either Bible knowledge or Islamic knowledge.

Source: National Examination Council of Tanzania, and author's calculations.

based on exam scores in Kiswahili, English, and mathematics.[9] This approach is taken as these core subjects are the only subjects that can be directly linked between primary and secondary school. Grade 'A' is given four points, 'B' is given three points, 'C' is given two points, 'D' is given one point, and 'F' is given zero points.[10] The average GPA in the core subjects in secondary school is similar for the applied sample and the full population of students taking the FTNA. It is further noticed that private school students perform substantially better in the core subjects in secondary school. Conditional on lagged achievement, Figure 1 illustrates how both low-achieving and high-achieving students perform better in private schools compared to public schools. The difference, however, is substantially larger for students performing better in primary school.

GPA PSLE is calculated the same way as *GPA FTNA*. The GPA is larger for the applied sample compared to the full population of primary school students. The reason is that the applied sample includes only students progressing to secondary school, whereas the full population includes students not progressing to secondary school. The gap in *GPA PSLE* between public and private secondary school students suggests private schools attract better students. The variable *GPA PSLE other* measures the primary school GPA based on the subjects 'community knowledge' and 'science'. Similar to *GPA PSLE*, the applied sample has a higher average of *GPA PSLE other* than the full population, and private school students tend to perform better. In the empirical analysis, variables related to exam scores are standardized by the sample means and sample standard deviations.

The main explanatory variable is the private secondary school dummy taking the value 1 if a student was enrolled in a private secondary school when taking the FTNA. Around 10 percent of

[9] The official GPA is calculated by taking the average of the seven highest graded subjects for a student (Government of Tanzania 2015).

[10] In secondary school in 2015, there is also the possibility of getting 'B+' and 'E'. 'B+' is given the same points as a 'B', while 'E' corresponds to zero points. For more on the grading system, see Government of Tanzania (2015).

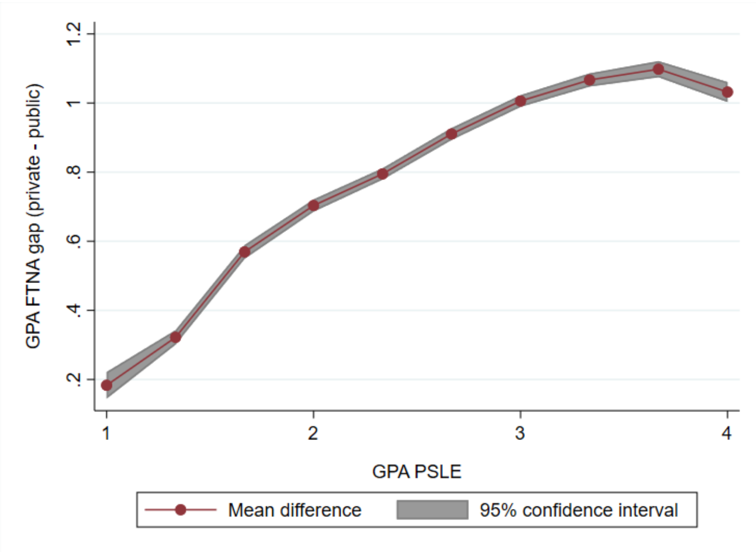


Figure 1: *Average GPA gap conditional on lagged achievement*

Notes: The y-axis refers to the difference in average GPA of English, Kiswahili, and mathematics in secondary school between private school students and public school students. *GPA PSLE* is the average of primary school exam scores in English, Kiswahili, and mathematics. The figure is based on 633,533 sample students. Students with a GPA PSLE below 1 are excluded due to large confidence bounds caused by few observations.

Source: National Examination Council of Tanzania, and author's calculations.

the sample students attend a private secondary school, meaning there is an under-representation of private secondary school students. In the robustness section, two approaches are pursued to test whether the private school learning premium is driven by a non-representative sample: 1) estimating a model using sample weights to get a representative sample in regard to private school enrolment, student gender, uncommonness of name, year of exam, secondary school GPA, ability of peers, and districts; and 2) employ a multiple imputation method to retrieve synthetic values of primary school exam scores for non-merged secondary school students. Over time, 33,354 students switch from public to private, while 13,068 students switch from private to public. A total of 549,679 students are always in public school, while 28,949 students are always in private school.[11]

[11]For 10,112 students, the type of primary school could not be determined.

Additional variables include peer effects, gender of the student, indicator for whether the student has an uncommon first name, and student cohort. To proxy for peer effects, *Peers PSLE* takes the average primary school GPA of a student's schoolmates in secondary school. By using primary school GPA instead of secondary school GPA of secondary school peers, peer effects are ensured not to capture the effect from the private school learning premium. To further acknowledge potential non-linearities of peer effects, the share of peers who failed the primary school exams and the share of peers who achieved at least one A in the three core subjects in primary school are included as control variables. These two dimensions of a student's peers are included as agglomeration of failing students could create a learning environment where failing becomes more acceptable, and the real ability of some peers could be constrained by the upper exam score. A first name is defined as uncommon if 20 or fewer students, out of more than one million, have been given the name.[12]

IV. METHODOLOGY

A. Baseline specifications

Six baseline specifications are estimated in the current paper, including the standard value-added model and a flexible value-added model derived in Section II. The first five specifications are included to illustrate different sources of bias, whereas the sixth is the preferred flexible value-added model. In all specifications, standard errors are clustered at the secondary school level.

The specifications presented do not include family-supplied inputs nor information on extra private tuition, which could potentially be an issue for identification. If, however, lagged student

[12]Recoding the uncommon name indicator to one if three or fewer students within the region bear the name, the coefficient estimate in the analysis increases but the remaining variables remain unaffected. Results are available upon request.

achievement and peer effects are accounted for, controlling for socio-economic variables is found to have a very limited and insignificant impact on school value-added estimates (Andrabi et al. 2011; Deming 2014; Muralidharan and Sundararaman 2015; Elks 2016) and teacher value-added estimates (Ballou, Sanders, and Wright 2004; Aaronson, Barrow, and Sander 2007; Kane and Staiger 2008; Chetty, Friedman, and Rockoff 2014). As the preferred model accounts for lagged achievement in the *same* primary school in the *same* cohort, the potential concern is mitigated even further. Furthermore, utilizing household survey data from Tanzania National Panel Survey reveals students switching to private school are simultaneously reducing expenditures to private tuition.[13] In a robustness analysis, sensitivity towards unobserved heterogeneity is further examined.

The first model applied is the simple cross-sectional OLS model given by Equation 8:

$$GPA_{i,s,c} = \beta_0 + \beta_1 Private_s + \beta_2 Female_i + \beta_3 Individualistic_i + \delta_c + \varepsilon_{i,s,c}. \tag{8}$$

Subscripts i, s, and c represent students, secondary schools, and cohorts, respectively. The dependent variable is standardized GPA based on lower secondary school exam scores in Kiswahili, English, and mathematics for student i in secondary school s, and cohort c. The sample consists of student cohorts taking the FTNA in 2015, 2016, or 2017. The parameter of main interest is β_1, measuring the private school learning premium. Other explanatory variables include a female indicator to control for potential gender discrimination or differences in outside options, a proxy for individualistic behaviour based on uncommonness of the name, and cohort fixed effects to account for e.g. varying difficulty of exams. The model provides unbiased coefficient estimates if the error term is uncorrelated with the explanatory variables. This is likely not true, however, as students in private

[13]See Appendix Table A3.

schools tend to outperform public school students before they start secondary school.

One may account for private school students being academically stronger students by including GPA from the PSLE as an explanatory variable. This gives the standard value-added model:

$$GPA_{i,s,p,c} = \beta_0 + \beta_1 Private_s + \beta_2 Female_i + \beta_3 Individualistic_i + \beta_4 GPA_{i,p} + \delta_c + \varepsilon_{i,s,p,c}. \tag{9}$$

Subscript p represents primary schools, meaning $GPA_{i,p}$ is standardized GPA based on primary school exam scores in Kiswahili, English, and mathematics for student i in primary school p. This variable is called the persistence parameter as it measures how well previous exam scores (in primary school) explain current exam scores (in lower secondary school). Again, for this model to provide unbiased coefficient estimates, the error term has to be uncorrelated with the explanatory variables.

Better primary schools may have taught students other things than what is needed for the PSLE. For instance, by focusing more on English to make the students understand the teaching in secondary school or preparing them for what is expected in secondary school. Given such unobserved primary school heterogeneity exists and private school students attend better primary schools, the private school learning premium is biased upward in Equation 9. Consequently, a model including primary school times cohort fixed effects is proposed in Equation 10:

$$GPA_{i,s,p,c} = \beta_0 + \beta_1 Private_s + \beta_2 Female_i + \beta_3 Individualistic_i + \beta_4 GPA_{i,p} + \delta_{p,c} + \varepsilon_{i,s,p,c}. \tag{10}$$

It is well-established in the literature that peer effects exist (Epple and Romano 2011; Sacerdote 2011). Since private school students' peers tend to perform better than public school students' peers, the private school learning premium is biased upward in Equation 10. Thus, in order to disentangle the true private school learning premium from peer effects, one needs to account for the latter.

Peer effects can be accounted for by controlling for the primary school performance of secondary school peers. While the performance of schoolmates is generally highly correlated with individual performance, Angrist (2014) finds causal peer effects to be less satisfactorily explored. The author argues the most compelling results are found when manipulating peer characteristics in a way that is unrelated to the characteristics of the individual student, but it is also in these cases the literature is least likely to find peer effects. Manski (1993) stresses three problems with standard approaches to measuring peer effects: (1) students self-select into specific groups; (2) students could simultaneously affect each other; and (3) peer effects may arise due to students' background characteristics or students' behaviour, and it is difficult to distinguish between these two effects. While *lagged* achievement of current peers addresses the second issue of simultaneity, it could be that students self-select into the same secondary school as their ambitious primary school peers.[14] Higher ambitions, however, will to some extent be captured by the lagged student achievement.[15] The model accounting for peer effects is given by Equation 11:

$$\begin{aligned} GPA_{i,s,p,c} = & \beta_0 + \beta_1 Private_s + \beta_2 Female_i + \beta_3 Individualistic_i + \beta_4 GPA_{i,p} \\ & + \beta_5 Peers\ PSLE_{i,s,c} + \beta_6 Peers\ failed_{i,s,c} + \beta_7 Peers\ with\ A_{i,s,c} + \delta_{p,c} + \varepsilon_{i,s,p,c}, \end{aligned} \tag{11}$$

where $Peers\ PSLE_{i,s,c}$, $Peers\ failed_{i,s,c}$, and $Peers\ with\ A_{i,s,c}$ are the average PSLE score, the share who failed primary school, and the share with at least one A in the core primary school subjects, respectively, for student i's schoolmates in secondary school s.[16] The applied formula for calculating

[14]Further manipulating the peer effect variable to include only current peers from another primary school does not change the results.

[15]Since the peer effects variable also acts as a control variable to capture socio-economic characteristics, the current paper refrains from claiming a causal effect from peer effects to achievement.

[16]It could be that private schools are better at capitalizing on peer effects. Following this hypothesis, the

the average PSLE score for peers is given by Equation 12, where $N_{s,c}$ represents the number of students in student i's cohort of secondary school s:

$$Peers\ PSLE_{i,s,c} = \frac{\sum_{j \neq i}^{N_{s,c}} GPA_{j,p}}{N_{s,c} - 1} \ , \ \ j \in \{1, 2, ..., N_{s,c}\} \tag{12}$$

While linear relationships between exam scores and the peer effect variables are assumed, the relationship could very well be more complex. That is, students do not interact similarly with all schoolmates. As an example, Carrell, Sacerdote, and West (2013) study the impact of placing low-achieving students into high-achieving peer groups. They find that the high-achieving students avoid the low-achieving students, thereby emphasizing the complexity of peer effects as they are dependent on social interactions. Similarly, Garlick (2018) finds that peer effects largely operate within race groups, and argues that students interact with socially proximate peers. In Appendix Table B4, the primary school exam scores for current peers are separated by own and opposite gender as students are thought of being more likely to interact with students of their own gender.

One critique of the model specified in Equation 11 is students could have different returns to school inputs, such as high-ability students benefiting more from a given level of school inputs. Consequently, a 'primary school times lagged achievement times cohort' fixed effect model is proposed in Equation 13. Thus, students with the same school inputs who have benefited similarly from these inputs after seven years of primary education are compared. As emphasized in Section II, this model assumes students taking the primary school exams together have received the same inputs during primary education. Appendix Table B9 restricts the sample to students who are even

primary school exam scores of secondary school peers are interacted with the private school dummy in Appendix Table B4.

more likely to have received the same inputs.[17]

$$
\begin{aligned}
GPA_{i,s,p,g,c} = & \beta_0 + \beta_1 Private_s + \beta_2 Female_i + \beta_3 Individualistic_i + \beta_4 Peers\ PSLE_{i,s,c} \\
& + \beta_5 Peers\ failed_{i,s,c} + \beta_6 Peers\ with\ A_{i,s,c} + \delta_{p,g,c} + \varepsilon_{i,s,p,g,c}.
\end{aligned}
\tag{13}
$$

Subscript g represents primary school GPA, and $\delta_{p,g,c}$ represents primary school times primary school GPA times cohort fixed effects. With 17,407 primary schools, three cohorts of students, and 13 primary school GPA possibilities, this model allows for 678,873 dummy variables. Of these, 482,577 are omitted as some schools have no students achieving a specific GPA in a specific year.

Despite comparing students from the same primary school achieving the same GPA based on Kiswahili, English, and mathematics, there could be differences in unobserved ability if the three exam scores inadequately capture true ability. In the standard value-added model, it is assumed the effect from unobserved ability decays at the same rate as inputs to the cumulative learning process. A violation of this assumption will bias the private school learning premium if unobserved ability is correlated with both exam scores and private education. The direction of the bias, however, is ambiguous. To account for this potential bias, two additional exam scores from primary school are utilized as students with higher unobserved ability are expected to perform better in other subjects. Thus, as a proxy for unobserved ability, exam scores in ‘community knowledge’ and ‘science’ from

[17] In order to ensure students went to the same primary school class, Appendix Table B9 restricts the sample to students who went to a primary school with 55 or less and 40 or less exam takers. These numbers are the average number of grade 7 students per class and the national class size target, respectively. The private school premium remains insignificantly different from the baseline results.

the PSLE are used. The preferred specification to the current paper is given by Equation 14:

$$\begin{aligned} GPA_{i,s,p,g,c} = &\beta_0 + \beta_1 Private_s + \beta_2 Female_i + \beta_3 Individualistic_i + \beta_4 Peers\ PSLE_{i,s,c} \\ &+ \beta_5 Peers\ failed_{i,s,c} + \beta_6 Peers\ with\ A_{i,s,c} + \beta_7 GPA\ other_{i,p} + \delta_{p,g,c} + \varepsilon_{i,s,p,g,c}, \end{aligned} \tag{14}$$

where $GPA\ other_{i,p}$ is the average exam score of 'community knowledge' and 'science' in the PSLE for student i.

B. Threats to identification and robustness analyses

Despite comparing secondary school exam scores for students from the same primary school achieving the same primary school GPA in the same year, one may still think of explanations for why the preferred specification in Equation 14 does not capture a genuine private school learning premium. The current subsection briefly describes how these concerns can be tested. The exact robustness specifications are presented and discussed either in Section V part B or in Appendix B. Specifically, six potential threats will be examined.

First, one may fear the coefficient estimate is not causal if, for instance, a student's future ability to learn is different from past ability to learn, the future ability to learn is known by either the student or the family, and the future ability has an impact on school choice. An IV model is proposed to test if this concern is justified for a subgroup of the student population. Whether or not a student fails the PSLE is used as an instrument for enrolling in private secondary school as students failing primary school are eligible only to private secondary school.[18] The potential causality issue, however, could still be intact if all failing students – who progress to secondary school – would have enrolled in

[18] Students fail if they score an average of D or E. This information is readily available from NECTA without further coding.

private secondary school independent of primary school exam scores. These are referred to as always-takers. Still, if some students *are* genuinely forced into private secondary school (referred to as compliers), a correction towards zero should be expected in the IV model if the private school learning premium is biased upward in the preferred baseline model.

Second, although the inclusion of primary school times primary school GPA times cohort fixed effects is likely to result in a high R^2 value, some unobserved heterogeneity will inevitably remain. Given this unobserved heterogeneity is correlated with both private schooling and exam scores in a similar way as the observed variables, the private school learning premium could be biased upward in the preferred baseline model. Oster (2019) proposes a method to examine what would hypothetically happen to the coefficient estimates given all variation could be explained and selection on observables is informative about selection on unobservables.

Third, switching from a public to a private school may happen together with other changes in the household. Thus, if a change in school type is correlated with changes in private tuition, child labour, child health, or household's general satisfaction with life, the identification of the private school learning premium could suffer from confounding factors. The National Panel Surveys from 2010-2011 and 2012-2013 are utilized to explore if changes in school types are correlated with changes in other household and child characteristics.

Fourth, the dependent variable consists of exam scores in three different subjects. Thus, one cannot tell from the preferred model whether private school students perform well across subjects or if the results are purely driven by one subject. To shed light on this issue a subject-specific analysis is proposed. Naturally, instead of using the primary school GPA as an explanatory variable, the primary school subject-specific exam score is used. This exam score is instrumented by the exam scores of the remaining subjects in order to account for measurement errors.

Fifth, the sample covers slightly more than half of all secondary school students, leaving almost half of the students out of the sample. This could for obvious reasons constitute a problem if the sample students respond differently to private schooling than the students outside the sample. To counter this issue, sample weights are computed to give representative estimates in regard to private school enrolment, student gender, uncommonness of name, year of exam, secondary school GPA, ability of peers, and districts. In addition to sample weighting, a multiple imputation approach is pursued to fill up primary school information for secondary school students outside the sample.

Sixth, private schools serve a very heterogeneous group of students. Consequently, the average private school learning premium may not tell us much about the general effect of private schools if the estimated effect is driven by expensive elite schools. To examine this issue, the top 1, 15, and 50 percent of private schools are excluded from the sample. This method will create a mechanically negative bias on the average effect. Similarly for public schools, some schools are perceived as 'elite public schools' taking in only the highest achieving students from primary school. In Kenya, Lucas and Mbiti (2014) find little evidence of these schools improving test scores relative to non-elite schools. An analysis of students progressing to the top 1 percent of public schools investigates whether these students achieve similar or higher exam scores relative to their peers from primary school progressing to a private secondary school.

Finally, a set of heterogeneity tests are performed in Appendix B. While these tests are not critical threats to identification, they provide insight into whether the private school learning premium is a general phenomenon or driven by specific types of students. They examine whether the learning premium differs in regard to urbanization, regional income growth, primary school size, primary school type, academic quality of peers, religion, same-gender secondary schools, and cohorts.

V. Results

A. Baseline specifications

The results from estimating the baseline models specified in Equations 8–14 are presented in Table 3. The model of main interest is presented in column (6), whereas columns (1)–(5) illustrate different sources of bias. Following the discussion on choice of significance levels (Hubbard and Bayarri 2003; Benjamin et al. 2018), the traditionally applied significance levels are replaced by 0.01, 0.005, and 0.001 when evaluating the results. This approach is taken due to high test power and low risk of type 2 errors. In result tables, however, only standard errors are provided.

The model in column (1) accounts for student gender, uncommonness of student name, and the year of taking the FTNA. The coefficient representing the private school learning premium suggests that attending a private secondary school instead of a public secondary school increases a student's GPA in Kiswahili, English, and mathematics by 1.38 standard deviations after two years. Descriptive statistics in Section III, however, demonstrated private secondary school students were already performing better than public secondary school students at the PSLE. Thus, the private school learning premium estimated in column (1) is highly susceptible to omitted variable bias.

In column (2), the private school coefficient estimate drops by around 28 percent when including lagged achievement, demonstrating private schools attract better students. The private school learning premium could still be biased upward if private school students went to better primary schools. The intuition is that good primary schools teach students things that are useful in secondary school, but they cannot be perfectly captured by primary school exam scores. The private school coefficient estimate drops by 36 percent after including primary school times cohort fixed effects,

Table 3: *Results from estimating the baseline specifications*

Dependent variable: GPA_s $(FTNA)$	(1)	(2)	(3)	(4)	(5)	(6)
$Private_s$	1.376	0.991	0.637	0.528	0.523	0.540
	(0.037)	(0.024)	(0.017)	(0.017)	(0.018)	(0.017)
$Female$	-0.160	-0.099	-0.035	-0.040	-0.009	0.059
	(0.011)	(0.007)	(0.005)	(0.004)	(0.005)	(0.005)
Uncommon name	0.016	0.009	0.016	0.014	0.016	0.015
	(0.005)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
GPA_p *(PSLE)*		0.447	0.628	0.583		
		(0.008)	(0.006)	(0.003)		
Peers $PSLE_s$				0.113	0.101	0.070
				(0.010)	(0.011)	(0.011)
Peers $failed_s$				-0.535	-0.615	-0.603
				(0.058)	(0.062)	(0.061)
Peers with A_s				0.321	0.269	0.302
				(0.047)	(0.049)	(0.048)
GPA $other_p$ *(PSLE)*						0.223
						(0.002)
Cohort FE	Yes	Yes	No	No	No	No
Primary school × *Cohort* FE	No	No	Yes	Yes	No	No
'*Primary school* × *Primary school GPA* × *Cohort*' FE	No	No	No	No	Yes	Yes
N	635,162	635,162	635,162	635,162	635,162	635,162
R^2	0.181	0.365	0.584	0.595	0.725	0.738

Notes: Column (6) is the preferred model. GPA_s and GPA_p are the GPA of the subjects Kiswahili, English, and mathematics in secondary and primary school, respectively. *Peers* $PSLE_s$ is the average GPA of the subjects Kiswahili, English, and mathematics in primary school for secondary school schoolmates. *Peers* $failed_s$ and *Peers with* A_s are the share of secondary school peers who failed primary school and the share of peers with at least one A from the three core subjects in primary school, respectively. *GPA* $other_p$ is the GPA of the subjects 'community knowledge' and 'science' in primary school. GPA_s, GPA_p, *Peer* $effects_s$, and *GPA* $other_p$ are standardized by their sample means and standard deviations. Standard errors are clustered at the secondary school level.

Source: National Examination Council of Tanzania, and author's calculations.

highlighting the importance of comparing students from the same primary school.

Another potential issue is peer effects. If well-performing students end up in the same secondary schools, and the ability of schoolmates influences a student's FTNA exam scores, the private school coefficient estimate is still biased upward. The model in column (4) accounts for peer effects by including the average PSLE score for a student's secondary school schoolmates, the share of schoolmates who failed the PSLE, and the share of schoolmates who achieved at least one A at the PSLE. The estimated private school learning premium drops to 0.53 of a standard deviation, and peer effects are highly significant. Moving from the 20th to the 80th percentile for all three dimensions of peers performance is associated with an improvement in secondary school exam scores of 0.18 of a standard deviation.[19] This tells a story of two mechanisms at play causing private school students to perform better: 1) students attending private schools have high-ability peers, potentially resulting in spillover effects;[20] and 2) a true private school learning premium, which could be caused by a number of different factors, such as efficient management, more focus on student achievement, and better facilities, among others. Appendix Table B4 further illustrates that exam scores of peers of the same gender are driving the positive relationship.

Section II argued that students could have different returns to school inputs dependent on their ability. By including a '*primary school* $\times$ *primary school GPA* $\times$ *cohort*' fixed effect, these different returns can be taken into account as students are compared to other students from the same primary school achieving the same primary school GPA. These students are likely to have received the same

[19] A one standard deviation improvement in all three dimensions of peers performance is associated with an improvement in secondary school exam scores of 0.16 of a standard deviation.

[20] It is important to stress that the peer effect variables could be endogenous even after controlling for lagged achievement. This is because peer effects and lagged achievement, according to the literature, capture the effects from family-supplied inputs.

school inputs and benefited similarly.[21] Column (5) shows that with this specification, the private school learning premium is slightly reduced to 0.52 of a standard deviation.

Finally, column (6) of Table 3 further includes GPA based on the subjects 'community knowledge' and 'science' from primary school. Section IV argued a significant effect from this variable could be due to unobserved ability. Another option is complementarities between subjects. If a student is good at science, it may be easier to learn math. The coefficient associated with $GPA\ other_p$ is positive and highly significant, despite accounting for lagged achievement in core subjects. Noteworthy is that the private school learning premium increases slightly, indicating a negative conditional correlation between private school enrolment and unobserved ability. The preferred coefficient estimate suggests that attending private secondary school is, on average, expected to improve a student's FTNA GPA in Kiswahili, English, and mathematics by 0.54 of a standard deviation after two years of secondary education. In Tanzanian primary schools, Mbiti et al. (2019) find that giving school grants combined with financial incentives for teachers improve student test scores by 0.36 of a standard deviation after two years. This comparison reveals that the coefficient estimate on private education is high when comparing to other interventions. Appendix Figure A2 illustrates the relative contributions of the private school learning premium and the other explanatory variables.

Almost half of the student subgroups (the fixed effect categories) in columns (5) and (6) have no variation in private secondary school treatment. The baseline regressions are further analysed on an effective sample, where student subgroups from columns (5) and (6) without private school treatment variation are excluded. This procedure does not change any of the overall conclusions.[22]

While the current paper seeks to estimate the magnitude of the private school learning premium,

[21] Restricting the sample to students even more likely to have received the same school inputs in primary school does not significantly change the private school learning premium (see Appendix Table B9).

[22] Results are available upon request.

it does not offer formal tests for the drivers of the learning premium. Potential avenues to explore include teacher quality and quantity, school facilities, learning incentives for students, and teaching incentives for teachers. School characteristics, however, presented by PO-RALG suggest private schools in 2016 had slightly more students per teacher, and a smaller proportion of teachers were considered to be qualified. On the other hand, private schools had better school facilities in regard to toilet facilities and books per student (Basic Statistics working group 2019). In regard to teaching incentives, Mbiti et al. (2019) show that these matter a great deal at the primary school level. As bonus pay is more common in private schools, the private school learning premium could, at least partially, be driven by these teaching incentives. Another common perception is that private schools are better managed. Crawfurd (2017) demonstrates, however, that quality of school management does not explain the private school learning premium in Ugandan secondary schools.

B. Robustness analyses

The following robustness analyses elucidate different threats to identification and types of heterogeneity. Threats to identification include endogeneity of the private school learning premium, omitted unobserved heterogeneity, and confounding factors. It is further examined whether the private school learning premium is equally driven by all three core subjects. The methodology and results of each robustness analysis are commented upon, while the results themselves are presented in Appendix A. In Appendix B, supplementary robustness and heterogeneity analyses are discussed in detail. These include application of sample weights and a multiple imputation approach to reflect on representativeness of the sample, and interactions of the private school learning premium with different geographical areas, peer effects, a religious school dummy, a same-gender school dummy, and cohort dummies. The preferred baseline model is further estimated on subsamples excluding students from the best performing private schools, and excluding students enrolled in large primary schools. A summary of the robustness and heterogeneity analyses are reported in Table 4.

Endogeneity of the learning premium

Despite comparing students with similar achievement in primary school and including a proxy for unobserved ability, the preferred baseline model in Table 3 may not perfectly account for selection into private schools. Consequently, the estimated private school learning premium could be endogenous given it is a specific type of student enrolling in private secondary school.

In order to assess the threat of unobserved selection, an IV model considers students around the cut-off point of passing the PSLE. The information on whether a student passes or fails is readily available from NECTA. Students failing primary school are in principle not eligible to enter a public secondary school, leading to a substantially larger share of students entering a private secondary school. Hence, failing is seen as an assignment to private school treatment. Students who are assigned to treatment without getting it (drop-outs) are referred to as never-takers. Students who will only get the treatment when assigned are referred to as compliers, whereas students who will get the treatment independent of assignment are referred to as always-takers.

The key issue of this IV model is selective attrition, as never-takers who are assigned to treatment will drop out of the education system and have missing outcomes at the secondary school exams. This can have a grave impact on the coefficient estimates from an IV model if never-takers have unobserved characteristics explaining secondary school exam scores. The reason for this is that the group of students assigned to treatment will differ in an unobserved dimension from the group of students not assigned to treatment. The former group will consist only of compliers and always-takers, whereas the latter group will also consist of never-takers. As an example, Engberg et al. (2014) show how ignoring selective attrition leads to biased estimates in a simple IV model when estimating the impact of a magnet program on student achievement.

Even if never-takers do not possess unobserved characteristics influencing secondary school exam scores, bias arise if always-takers or compliers hold such unobserved characteristics. Again, the assignment group will differ from the non-assignment group, leading to biased IV estimates.

One may examine whether unobserved characteristics differ between compliers and always-takers by comparing the coefficient estimates from the baseline value-added model with the coefficient estimates from the IV model. Given the private school learning premiums in the two models are equal, two explanations exist: 1) always-takers do not possess unobserved characteristics improving their secondary school exam scores relative to compliers; or 2) all private school students who failed the primary school exams are always-takers. Comparing student peers with similar achievement in primary school, it is found that some students are indeed compliers to the private school treatment.[23] Consequently, if private school students do not possess unobserved characteristics improving their secondary school exam scores, we should expect a similar private school learning premium in the IV model as in the preferred baseline value-added model.

While one cannot explicitly test the validity of the instrument, column (2) demonstrates that including the instrument in the baseline model yields an insignificant coefficient estimate. It is reassuring that the variable does not correlate significantly with secondary school exam scores,

[23]Comparing students from the same primary school and cohort who achieved the exact same primary school exam scores, failing is associated with a 34 percent higher probability of being present in a private secondary school two years later. Despite achieving the same exam scores, it is possible to either fail or pass as letter grades are used as achievement outcomes. The students, however, receive an (to the author unobserved) exam score between 0-100, from which the letter grades are determined. Thus, students may achieve a "big" or "small" A, B, C, D, and E, thereby explaining why students with the same letter grades can both fail and pass primary school as a whole.

which could indicate a direct effect of failing caused by, for instance, a desire not to experience failure again or decreased confidence in oneself. This is not to say, however, that the instrument indeed has no direct effect. Another potential channel through which the instrument could have a direct effect stems from the fact that students are compared to others with the same letter grades. Variation within letter grades, however, is not accounted for. Arguably, the instrument captures some of the within letter grade variation, suggesting failing the primary school exams could be negatively correlated with secondary school exam scores. The first-stage regression in column (3) shows that the instrument is positively correlated with private school enrolment, and the relevance of the instrument is supported by the Cragg–Donald Wald F statistic.[24]

In the IV model in column (4), the private school learning premium is almost identical to the baseline estimate in column (1). This suggests there are no differing unobserved characteristics affecting both private school enrolment and secondary school exam scores for always-takers and compliers within the subgroup of low-performing students. Following the baseline results in Table 3, this was expected as even including a proxy for unobserved ability did not lower the private school learning premium. The result is also in line with previous studies, finding that socio-economic variables have a very limited and insignificant impact on school value-added estimates when lagged test scores and peer effects are accounted for (Andrabi et al. 2011; Deming 2014; Muralidharan and Sundararaman 2015; Elks 2016). Important to emphasize is that this analysis does not test whether never-takers hold relevant unobserved characteristics relative to compliers and always-takers.

The results are further robust to controlling for the share of secondary schools being private within the region to account for supply-side constraints, and excluding one region at a time.[25]

[24] The critical value for the Stock-Yogo weak ID test with a maximum IV bias of 10 percent is 16.38.
[25] Results are available upon request.

Following these results, there are strong indications of a causal private school learning premium, at least for low-performing students who comply with the assignment to private school treatment.

Sensitivity to unobserved heterogeneity

Oster (2019) proposes a method to examine what happens to a specific coefficient estimate assuming one is able to explain all variation in the dependent variable, and selection on observables is informative about selection on unobservables. The idea is that the difference between the coefficient estimates of the explanatory variable of interest from an uncontrolled model and the preferred model is informative about what would happen to the coefficient estimate could we account for unobserved controls. This adjustment is a harsh test of unobserved heterogeneity, and the coefficient should be interpreted as an extreme lower bound estimate.

In the restricted estimator case, it is assumed that selection on unobservables is equal to selection on observables. It is further assumed that the coefficient estimate to the explanatory variable of interest is independent of controlling for the true effects from the observed control variables. That is, including an index equal to the observed control variables multiplied by their coefficient estimates from a regression where unobserved heterogeneity is also controlled for. Under these conditions, the following β^* is a consistent estimator of true β:

$$\beta^* = \widetilde{\beta} - \left(\mathring{\beta} - \widetilde{\beta}\right) \frac{R^2_{\max} - \widetilde{R^2}}{\widetilde{R^2} - \mathring{R}^2}, \tag{15}$$

where $\widetilde{\beta}$ is the coefficient estimate of the private school dummy in the preferred model, $\mathring{\beta}$ is the coefficient estimate of the private school dummy in the uncontrolled model, $R^2_{\max}$ is R^2 from a hypothetical model accounting for unobserved controls, $\widetilde{R^2}$ is R^2 from the preferred model, and $\mathring{R}^2$ is

R^2 from the uncontrolled model. Based on a sample of randomized results from top journal articles, it is suggested that $R^2_{\max} = 1.3\widetilde{R^2}$. This level of $R^2_{\max}$ ensures that 90 percent of the randomized results survive the sensitivity test.

Following the method laid out above, the private school learning premium in the current paper is robust to unobserved controls, but the magnitude of the coefficient drops from 0.54 to 0.22.[26] In a sample of non-randomized results from top journal articles, only 45 percent survive the proposed sensitivity test without changing sign (Oster 2019).

Since variables such as peer effects, lagged exam scores, and quality of primary school are highly correlated with both a student's present exam scores and private school enrolment, including them as control variables leads to a lower estimated private school learning premium. Assuming unobservable controls are similarly correlated with private school enrolment seems unrealistic. The author of the proposed sensitivity test further argues that setting the relative degree of selection on unobservables equal to selection on observables may serve as an upper bound. In a wage data simulation exercise, the author finds an average relative degree of selection on unobservables equal to 0.545 of the selection on observables. The relative degree of selection can also take negative values if there are other variables like $GPA\ other_p$ having a positive impact on GPA_s and a negative conditional correlation with private secondary education.

Confounding factors

Switching from public to private school is certainly not a random choice taken by households. There could be numerous reasons for switching to a private school when progressing from primary to

[26]In the case of the unrestricted estimator, the preferred private school learning premium does *not* survive the sensitivity test. The relative degree of selection on unobservables compared to selection on observables has to be less than 0.36 in order for the private school learning premium to be positive.

secondary school. Many of these are by nature difficult – or even impossible – to observe. While it remains beyond the scope of this paper to account for all potential reasons, a few dimensions are identified as important for further examination. These dimensions include private tuition, child labour, child health, and general life satisfaction of the household.

The Tanzania National Panel Surveys in 2010-2011 and 2012-2013 are utilized to explore if changes in school types are correlated with changes in the dimensions stated above. The applied sample consists of 3,305 children enrolled in primary and secondary school in both survey rounds.[27] Private tuition is proxied by log of household expenditures to private tuition. Child labour is examined both at the extensive margin and the intensive margin by using log of hours worked. Child health is proxied by hospitalizations and visits to traditional healers. Finally, household satisfaction with life is based on average satisfaction with own life for adult household members.

Appendix Table A3 presents the results from regressing changes in the above-mentioned variables on changes in type of school. The models illustrate that changes in the outcome variables of interest – except for private tuition – are insignificantly correlated with changes in type of school. Shifting to a private school is associated with a reduction in money spent on private tuition, suggesting the preferred private school learning premium of 0.54 of a standard deviation could be biased downward. One potential explanation for this negative correlation between private school enrolment and private tuition expenditures is that students receive a better education in private schools, and hence they do not need the extra private tuition.

[27] Primary school students are included to increase the power of the results. Restricting the sample to children in lower secondary school, however, does not change the main points.

Subject-specific learning premiums

Appendix Table A4 seeks to shed light on whether it is a specific subject that is driving the overall private school learning premium. To study these subject-specific private school learning premiums, the GPA based on Kiswahili, English, and mathematics exam scores is replaced by the exam scores in each subject. Columns (1), (3), and (5) apply the standard value-added model for the subjects Kiswahili, English, and mathematics, respectively.

In the value-added literature, however, measurement errors are found to be of significant importance (Andrabi et al. 2011; Lockwood and McCaffrey 2014; Koedel, Mihaly, and Rockoff 2015). Not accounting for measurement errors will result in a bias towards zero for the persistence parameter. This bias may also contaminate the private school learning premium if enrolment in private secondary school is correlated with achievement at the PSLE after conditioning on the other explanatory variables. The direction of the bias is positive (negative) if there is a positive (negative) conditional correlation between the two variables. Worth emphasizing is that measurement errors are problematic particularly when studying specific exam scores. Studying an overall GPA, however, would still be problematic in the case of correlated measurement errors across subjects (Andrabi et al. 2011). The exam scores used as a proxy for unobserved ability could mitigate any potentially correlated measurement errors across the three core subjects in the baseline regressions.

As a consequence of measurement errors, columns (2), (4), and (6) follow Andrabi et al. (2011) by instrumenting primary school exam scores in Kiswahili, English, and mathematics, respectively, with the remaining exam scores. The last three columns of Appendix Table A4 include a *primary school* $\times$ *instrumented subject-specific primary school exam score* $\times$ *cohort* fixed effect to ensure students benefit similarly from the same school inputs. The primary school exam score of interest is predicted based on primary school exam scores in the other subjects, gender, year fixed effects, and primary school school fixed effects. Next, students with a similarly predicted exam score from the same primary school in the same cohort are compared by including the aforementioned fixed effect.[28]

The first six columns demonstrate that: 1) the private school learning premiums are significantly positive for all subjects; 2) in line with theory, accounting for measurement error increases the persistence parameter; and 3) the private school learning premiums change slightly in both directions when accounting for measurement error, indicating a weak and unsystematic conditional correlation between lagged achievement and private secondary school enrolment. In columns (7)–(9), the private school learning premiums decrease, which is in line with the baseline results in Section V. Attending private secondary school is associated with significant increases in Kiswahili, English, and mathematics FTNA exam scores of 0.38, 0.48, and 0.69 standard deviations, respectively.[29]

Further robustness and heterogeneity analyses

In addition to the robustness analyses above, examination of the representativeness of the sample along with a set of heterogeneity tests are presented and discussed in Appendix B. Specifically, sample weights and a multiple imputation approach are applied to reflect on representativeness, and the private school enrolment dummy is interacted with different geographical areas, primary school type, peer effects, a religious school dummy, a same-gender school dummy, and cohort dummies. Three final models restrict the sample to students outside the best-performing private

[28]Predicted subject-specific exam scores are split by each 0.1 point (for example, students getting a predicted exam score of 1.30 to 1.39 are compared).

[29]Several other subjects are taught in secondary school. Kiswahili, English, and mathematics, however, are the only ones with a clear link between primary school and secondary school.

schools, students progressing the to best-performing public schools and their primary school peers progressing to private school, and students from small primary schools.

The results suggest the following: 1) neither applying sample weights nor using a multiple imputation approach significantly change the private school learning premium from the baseline model; 2) in line with Singh (2015), the private school learning premium is larger in rural areas; 3) high and low income growth regions have similar private school learning premiums; 4) in line with Andrabi et al. (2011), the private school learning premium is larger for students who went to public primary school; 5) peer effects are more influential in private schools; 6) while public schools offering the course Islamic knowledge are as good as other public schools, private schools offering Islamic knowledge perform worse than other private schools; 7) there is no effect of schools offering Bible knowledge; 8) same-gender school students perform better than mixed-gender school students, and private girls-only school students perform better than other private school students; 9) the private school learning premium was similar for the 2015 and 2016 cohorts. In 2017, however, students had been fully affected by the free secondary school reform, and the private school learning premium increased by an estimated 0.10 of a standard deviation compared to 2015; 10) the results are highly robust to excluding the best-performing private schools from the sample. Even excluding the top 50 percent of private schools leads to a private school learning premium of 0.30 of a standard deviation; 11) students progressing to an 'elite public school' perform significantly worse than their primary school peers progressing to a private school. This also holds after excluding the best-performing private schools; and 12) restricting the sample to students from small primary schools – to ensure they attended the same class – does not significantly change the private school learning premium.

Table 4: *Summary of robustness and heterogeneity analyses*

Robustness	Table	PSLP (baseline)	PSLP (robustness)	Main point
Potential endogeneity	A2	0.259	0.259	Always-takers are similar to compliers.
Oster adjustment	-	0.540	0.220	The positive PSLP survives a harsh test of unobserved heterogeneity.
Confounding factors	A3	-	-	Private tuition drops when moving to private school. Other outcomes remain unchanged.
Kiswahili score English score Math score	A4	0.378 0.475 0.686	- - -	PSLP is significantly positive for all subjects, largest for math, and lowest for Kiswahili.
Sample weights	B1	0.540	0.508	No significant effect of using weights to attain representativeness in key dimensions.
Multiple imputation	B1	0.540	0.558	No significant effect of using multiple imputation for non-merged students.
Heterogeneity and subsamples	**Table**	**PSLP**	**PSLP × heterogeneity**	**Main point**
Urban secondary	B2	0.576	-0.126	Higher PSLP in rural areas.
Income growth	B2	0.558	0.002	No effect of regional income growth.
Private primary	B3	0.609	-0.280	Higher PSLP for public primary students.
Performance of peers	B4	0.470	0.067	Peers matter more in private schools.
Religion in school	B5	0.550	-0.025	No effect of school offering religious course.
Same-gender school	B6	0.519	0.047	Only higher PSLP for girls-only schools.
Cohort 2015 Cohort 2016 Cohort 2017	B7	0.484 0.482 0.626	- - -	Similar PSLP for the 2015 and 2016 cohorts. The 2017 cohort, fully affected by the free secondary school reform, has a higher PSLP.
Worst private schools	B8	0.301	-	Significant PSLP despite excluding 50 percent best private schools
Elite public schools	B8	0.548	-	Similar PSLP as baseline for students progressing to elite public schools.
Small primary schools	B9	0.520	-	No effect of students from primary schools with one expected class per cohort only

Notes: PSLP refers to private school learning premium.

Source: National Examination Council of Tanzania, and author's calculations.

VI. CONCLUSION

The literature on the private school learning premium in developing countries focuses mostly on South Asia, while evidence from African countries is lacking. Most studies apply cross-sectional data sources, leaving only a minority of studies applying value-added models or lottery-based experiments. The magnitude of the learning premiums differ substantially between studies, highlighting the importance of context.

This paper sets up a value-added model to estimate the learning effects of attending a private secondary school in Tanzania, using a new dataset of exam records for 635,000 students progressing from primary to secondary school. The model compares students with the same primary school GPA from the same primary school in the same year, and further controls for peer effects, unobserved ability, and a measure of student personality. Various robustness analyses are performed, including estimating an IV model and testing for the risk of confounding factors.

The preferred result on the private school learning premium suggests that attending private secondary school is expected to increase the FTNA exam scores by 0.54 of a standard deviation after two years of secondary schooling. An IV model suggests the private school learning premium is causal, at least for a subgroup of students. An OLS model and a standard value-added model bias the private school learning premium upward by 155 and 84 percent, respectively. The difference between a standard value-added model and this paper's flexible value-added model could be driven by lagged school fixed effects capturing local market effects and socio-economic status of parents. Robustness analyses show positive subject-specific learning premiums of 0.38, 0.48, and 0.69 standard deviations for Kiswahili, English, and mathematics, respectively. Despite a considerable drop, the private school learning premium survives a harsh test for unobserved heterogeneity.

The current paper contributes to the existing literature in two dimensions. First, while the use of private schools have increased rapidly in sub-Saharan Africa, the evidence on learning effects has not followed. By estimating a private school learning premium in Tanzania, evidence is added to an understudied region of developing countries. Second, despite recent progress in validating value-added models, the current paper adds to the literature on the general use of value-added models by proposing a flexible value-added model. It is flexible in the sense that students are compared only to other students with the same lagged exam scores from the same school. This approach is argued to be useful, in particular, when socio-economic variables are missing.

Private secondary schools in Tanzania deliver in terms of improving student exam scores compared to public schools. The next steps are to figure out why, compare costs of private and public schools, and examine inequality effects of expanding the private school sector. Until we know more about how the public sector can close the learning gap, policy recommendations emerging from this paper's analysis include offering private school vouchers to students from low-income households, or supporting private schools in areas where demand is low due to prices. Vouchers have the potential to both improve learning for treated students and narrow the peer effect channel.

REFERENCES

Aaronson, Daniel, Lisa Barrow, and William Sander. 2007. "Teachers and Student Achievement in the Chicago Public High Schools." *Journal of Labor Economics* 25 (1): 95–135.

Alderman, Harold, Peter F. Orazem, and Elizabeth M. Paterno. 2001. "School Quality, School Cost, and the Public/Private School Choices of Low-Income Households in Pakistan." *The Journal of Human Resources* 36 (2): 304–326.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2008. "A Dime a Day: The Possibilities and Limits of Private Schooling in Pakistan." *Comparative Education Review* 52 (3): 329–355.

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc. 2011. "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics." *American Economic Journal: Applied Economics* 3 (3): 29–54.

Angrist, Joshua D. 2014. "The Perils of Peer Effects." *Labour Economics* 30:98–108.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters. 2017. "Leveraging Lotteries for School Value-Added: Testing and Estimation." *The Quarterly Journal of Economics* 132 (2): 871–919.

Angrist, Joshua D., Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." *American Economic Review* 92 (5): 1535–1558.

Ballou, Dale, William Sanders, and Paul Wright. 2004. "Controlling for Student Background in Value-Added Assessment of Teachers." *Journal of Educational and Behavioral Statistics* 29 (1): 37–65.

Barrera-Osorio, Felipe, David S Blakeslee, Matthew Hoover, Leigh Linden, Dhushyanth Raju, and Stephen P Ryan. 2017. *Delivering Education to the Underserved Through a Public-Private Partnership Program in Pakistan.* Working Paper 23870. Cambridge, MA: National Bureau of Economic Research.

Basic Statistics working group. 2019. *Education Statistics.* http://opendata.go.tz/dataset?groups=education-group. Accessed May 5, 2019.

Baum, Donald R., and Jacobus Cilliers. 2018. "Private School Vouchers for Expanding Secondary School Access? The Case of Tanzania." *International Journal of Educational Management* 32 (7): 1307–1318.

Benjamin, Daniel J., James O. Berger, Magnus Johannesson, Brian A. Nosek, E.-J. Wagenmakers, Richard Berk, Kenneth A. Bollen, et al. 2018. "Redefine Statistical Significance." *Nature Human Behaviour* 2 (1): 6.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, and Justin Sandefur. 2013. *The High Return to Low-Cost Private Schooling in a Developing Country.* IGC WORKING PAPER. Oxford, UK: International Growth Centre.

Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications.* USA: Cambridge University Press.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81 (3): 855–882.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9): 2593–2632.

Crawfurd, Lee. 2017. "School Management and Public–Private Partnerships in Uganda." *Journal of African Economies* 26 (5): 539–560.

Day Ashley, Laura, Claire Mcloughlin, Monazza Aslam, Jakob Engel, Joseph Wales, Shenila Rawal, Richard Batley, Geeta Kingdon, Susan Nicolai, and Pauline Rose. 2014. *The Role and Impact of Private Schools in Developing Countries: A Rigorous Review of the Evidence. Final Report.* Technical Report. London, UK: Department for International Development.

Deming, David J. 2014. "Using School Choice Lotteries to Test Measures of School Effectiveness." *American Economic Review* 104 (5): 406–411.

Dizon-Ross, Rebecca. 2019. "Parents' Beliefs About Their Children's Academic Ability: Implications for Educational Investments." *American Economic Review* 109 (8): 2728–2765.

Elks, Phil. 2016. *Lessons Learned from Introducing Value Added Performance Measures in Uganda.* DFID Think Piece. Department for International Development.

Engberg, John, Dennis Epple, Jason Imbrogno, Holger Sieg, and Ron Zimmer. 2014. "Evaluating Education Programs That Have Lotteried Admission and Selective Attrition." *Journal of Labor Economics* 32 (1): 27–63.

Epple, Dennis, and Richard E. Romano. 2011. "Peer Effects in Education: A Survey of the Theory and Evidence." In *Handbook of Social Economics,* vol. 1B, 1053–1163. Amsterdam: Elsevier.

Garlick, Robert. 2018. "Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment." *American Economic Journal: Applied Economics* 10 (3): 345–369.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics* 1 (1): 112–135.

Government of Tanzania. 2015. *The National Examinations Council Act Cap. 107.* Government Notice 509.

Heyneman, Stephen P., and Jonathan M. B. Stern. 2014. "Low Cost Private Schools for the Poor: What Public Policy Is Appropriate?" *International Journal of Educational Development* 35:3–15.

Hubbard, Raymond, and M. J. Bayarri. 2003. "Confusion Over Measures of Evidence (p's) Versus Errors (α's) in Classical Statistical Testing." *The American Statistician* 57 (3): 171–178.

Human Rights Watch. 2017. *"I Had a Dream to Finish School": - Barriers to Secondary Education in Tanzania.* Technical Report. New York: Human Rights Watch.

Jimenez, Emmanuel, Marlaine E. Lockheed, and Vicente Paqueo. 1991. “The Relative Efficiency of Private and Public Schools in Developing Countries.” *The World Bank Research Observer* 6 (2): 205–218.

Kane, Thomas J., and Douglas O. Staiger. 2008. *Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation.* Working Paper 14607. Cambridge, MA: National Bureau of Economic Research.

Knudsen, Anne Sofie Beck. 2019. *Those Who Stayed: Individualism, Self-Selection and Cultural Change During the Age of Mass Migration.* SSRN SCHOLARLY PAPER ID 3321790. Rochester, NY: Social Science Research Network.

Koedel, Cory, Kata Mihaly, and Jonah E. Rockoff. 2015. “Value-Added Modeling: A Review.” *Economics of Education Review* 47:180–195.

Lassibille, Gérard, and Jee-Peng Tan. 2001. “Are Private Schools More Efficient Than Public Schools? Evidence from Tanzania.” *Education Economics* 9 (2): 145–172.

Lassibille, Gérard, Jee-Peng Tan, and Suleman Sumra. 2000. “Expansion of Private Secondary Education: Lessons from Recent Experience in Tanzania.” *Comparative Education Review* 44 (1): 1–28.

Lockwood, J. R., and Daniel F. McCaffrey. 2014. “Correcting for Test Score Measurement Error in ANCOVA Models for Estimating Treatment Effects.” *Journal of Educational and Behavioral Statistics* 39 (1): 22–52.

Lucas, Adrienne M., and Isaac M. Mbiti. 2014. “Effects of School Quality on Student Achievement: Discontinuity Evidence from Kenya.” *American Economic Journal: Applied Economics* 6 (3): 234–263.

Manski, Charles F. 1993. “Identification of Endogenous Social Effects: The Reflection Problem.” *The Review of Economic Studies* 60 (3): 531–542.

Mbiti, Isaac, Karthik Muralidharan, Mauricio Romero, Youdi Schipper, Constantine Manda, and Rakesh Rajani. 2019. “Inputs, Incentives, and Complementarities in Education: Experimental Evidence from Tanzania.” *The Quarterly Journal of Economics* 134 (3): 1627–1673.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2015. “The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India.” *The Quarterly Journal of Economics* 130 (3): 1011–1066.

Ngetich, Solomon Kipyego, Benjamin Kyalo Wambua, and Zachariah Kiptoo Kosgei. 2014. “Determination of Unit Cost Among Secondary Schools in Kenya: A Case of Nandi North District, Kenya.” *European Scientific Journal* 10 (16): 211–224.

Oster, Emily. 2019. “Unobservable Selection and Coefficient Stability: Theory and Evidence.” *Journal of Business & Economic Statistics* 37 (2): 187–204.

Psacharopoulos, George. 1987. “Public versus Private Schools in Developing Countries: Evidence from Colombia and Tanzania.” *International Journal of Educational Development* 7 (1): 59–67.

Sacerdote, Bruce. 2011. “Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?” In *Handbook of the Economics of Education,* edited by Eric A. Hanushek, Stephen Machin, and Ludger Woesmann, 3:249–277. Amsterdam: Elsevier.

Schafer, J. L. 1997. *Analysis of Incomplete Multivariate Data.* 1 edition. London, UK: Chapman and Hall.

Schirmer, Stefan. 2010. *Hidden Assets: South Africa’s Low-Fee Private Schools.* CDE IN DEPTH 10. Johannesburg: The Centre for Development and Enterprise.

Singh, Abhijeet. 2019. “Learning More with Every Year: School Year Productivity and International Learning Divergence.” *Journal of the European Economic Association.*

———. 2015. “Private School Effects in Urban and Rural India: Panel Estimates at Primary and Secondary School Ages.” *Journal of Development Economics* 113:16–32.

Todd, Petra E., and Kenneth I. Wolpin. 2003. “On the Specification and Estimation of the Production Function for Cognitive Achievement*.” *The Economic Journal* 113 (485): F3–F33.

Tooley, James, Yong Bao, Pauline Dixon, and John Merrifield. 2011. “School Choice and Academic Performance: Some Evidence From Developing Countries.” *Journal of School Choice* 5 (1): 1–39.

Tukey, John W. 1977. *Exploratory Data Analysis.* 1 edition. Reading, MA: Addison-Wesley.

UN DESA. 2018. *The Sustainable Development Goals Report 2018.* Technical report. New York: UN.

UNESCO. 2020. *Education Statistics.* http://data.uis.unesco.org. Accessed January 25, 2020.

UNICEF. 2016. *Education Budget Brief: FY 2011/12 - FY 2015/16.* Technical report. UNICEF.

Urquiola, M. 2016. "Chapter 4 - Competition Among Schools: Traditional Public and Private Schools." In *Handbook of the Economics of Education,* edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 5:209–237. Amsterdam: Elsevier.

World Bank. 2019. *World Development Indicators.* http://databank.worldbank.org/data/reports.aspx?source=2&cour Accessed November 6, 2019.

Appendix A: Sample distributions and robustness analyses

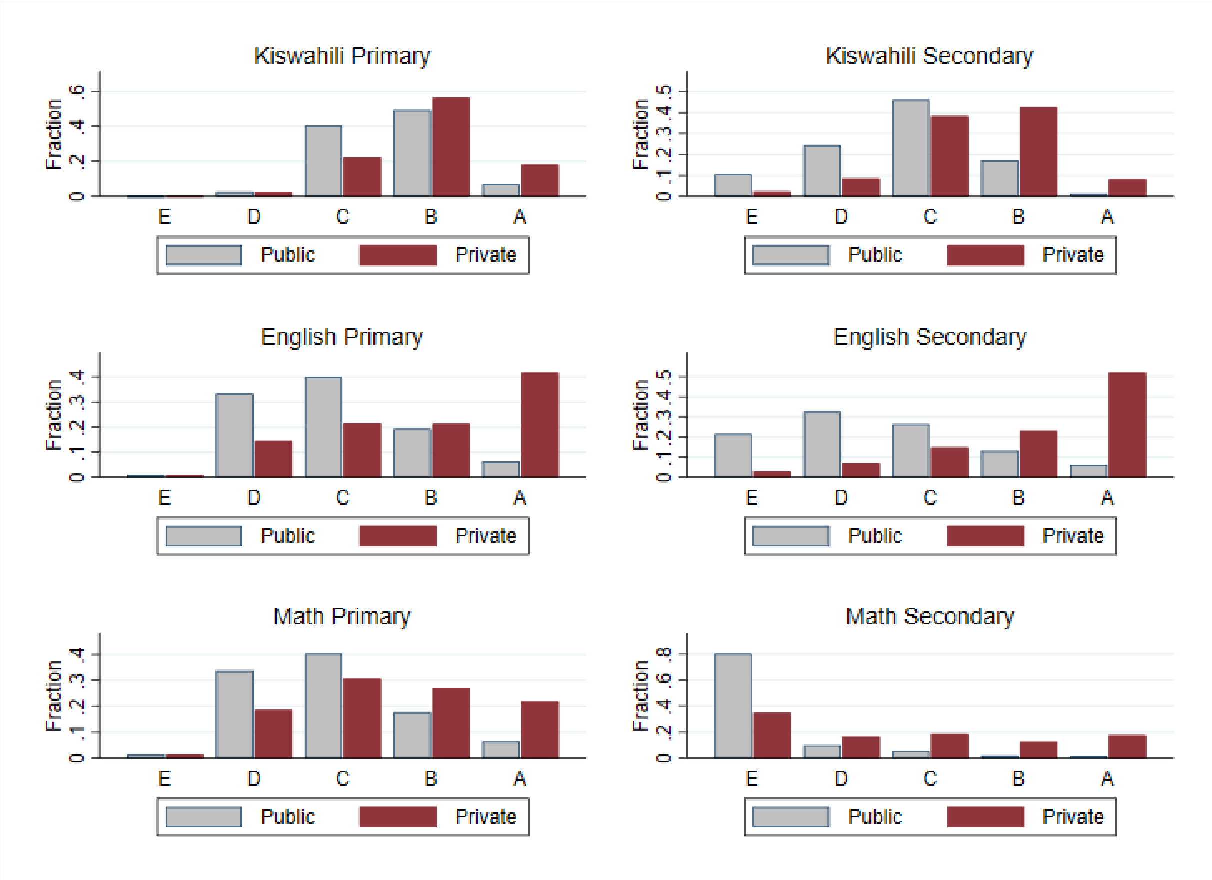


Figure A1: *Distribution of subject-specific exam scores*

Notes: The distinction between public and private school students are at the secondary school level. Thus, the figures illustrate the fractions of public secondary school students getting a specific grade in primary and secondary school for different subjects, and the fractions of private secondary school students getting a specific grade in primary and secondary school for different subjects. The figures are based on the 635,162 sample students.

Source: National Examination Council of Tanzania, and author's calculations.

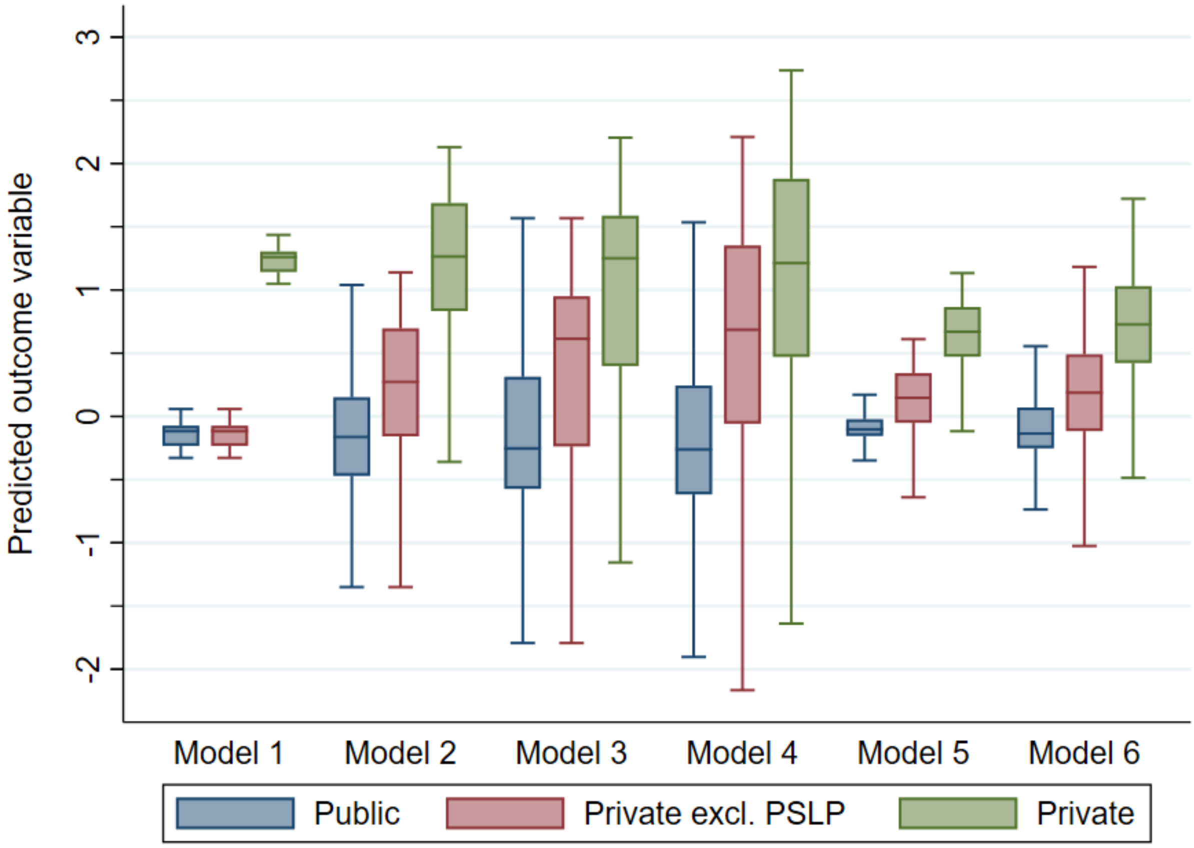


Figure A2: *Box plots of the predicted outcome variable*

Notes: The y-axis refers to the predicted outcome variable (standardized secondary school GPA based on Kiswahili, English, and mathematics) from the six models in Table 3. The first category (blue) is public secondary school students, the second category (red) is private secondary school students excluding the private school learning premium (PSLP), and the third category (green) is private secondary school students including all explanatory variables. The lower and upper adjacent values follow the recommendation by Tukey (1977).

Source: National Examination Council of Tanzania, and author's calculations.

Table A1: *Regional distribution of secondary school students in the sample*

	Population	Sample	Private	Public
Arusha	0.056	0.051	0.072	0.049
Dar Es Salaam	0.106	0.098	0.168	0.091
Dodoma	0.035	0.037	0.036	0.037
Geita	0.032	0.038	0.009	0.041
Iringa	0.035	0.032	0.035	0.032
Kagera	0.048	0.044	0.032	0.045
Katavi	0.008	0.008	0.002	0.009
Kigoma	0.031	0.023	0.031	0.022
Kilimanjaro	0.071	0.072	0.130	0.066
Lindi	0.016	0.017	0.004	0.019
Manyara	0.027	0.023	0.012	0.025
Mara	0.043	0.048	0.018	0.051
Mbeya	0.075	0.083	0.083	0.083
Morogoro	0.050	0.045	0.053	0.044
Mtwara	0.027	0.031	0.012	0.033
Mwanza	0.077	0.081	0.064	0.083
Njombe	0.022	0.023	0.024	0.022
Pwani	0.034	0.035	0.078	0.031
Rukwa	0.016	0.013	0.007	0.014
Ruvuma	0.032	0.026	0.026	0.026
Shinyanga	0.029	0.033	0.022	0.035
Simiyu	0.023	0.030	0.005	0.033
Singida	0.023	0.023	0.013	0.024
Songwe	0.001	0.000	0.004	0.000
Tabora	0.028	0.027	0.024	0.027
Tanga	0.055	0.056	0.037	0.058
Total	100%	100%	100%	100%
N	1,245,860	635,162	63,206	571,956

Source: National Examination Council of Tanzania, and author's calculations.

Table A2: *Value-added model versus IV model*

	Value-added model (1)	Value-added model (2)	IV model first stage (3)	IV model second stage (4)
$Private_s$	0.259	0.259		0.259
	(0.029)	(0.031)		(0.065)
Female	0.110	0.110	0.005	0.110
	(0.010)	(0.010)	(0.003)	(0.010)
Uncommon name	0.016	0.016	0.008	0.016
	(0.015)	(0.015)	(0.005)	(0.015)
Peers $PSLE_s$	0.087	0.087	0.042	0.087
	(0.020)	(0.020)	(0.015)	(0.021)
Peers failed $(PSLE)_s$	-0.293	-0.293	1.568	-0.294
	(0.074)	(0.074)	(0.064)	(0.137)
Peers with A $(PSLE)_s$	0.412	0.412	0.903	0.412
	(0.118)	(0.118)	(0.080)	(0.135)
GPA $other_p$ (PSLE)	0.110	0.110	0.003	0.110
	(0.008)	(0.008)	(0.003)	(0.008)
Failing PSLE		0.000	0.274	
		(0.019)	(0.012)	
'*Primary school × Primary school GPA × Cohort*' FE	Yes	Yes	Yes	Yes
Cragg-Donald Wald F statistic			2,368.2	
N	19,373	19,373	19,373	19,373
R^2	.361	.361	.744	.354

Notes: The dependent variable in columns (1), (2), and (4) is GPA of the subjects Kiswahili, English, and mathematics in secondary school. The dependent variable in column (3) is private secondary school enrolment. The sample consists of 'primary school × primary school GPA × Cohort' cells with variation in failing primary school. *Failing PSLE* states, based on NECTA classification, whether a student failed the primary school exams, and it is used as the instrument for private school enrolment. *Peers $PSLE_s$* is the average GPA of the subjects Kiswahili, English, and mathematics in primary school for secondary school schoolmates. *Peers $failed_s$* and *Peers with A_s* are the share of secondary school peers who failed primary school and the share of peers with at least one A from the three core subjects in primary school, respectively. *GPA $other_p$* is the GPA of the subjects 'community knowledge' and 'science' in primary school. *GPA_s*, *Peer $effects_s$*, and *GPA $other_p$* are standardized by their sample means and standard deviations. Standard errors are clustered at the secondary school level.
Source: National Examination Council of Tanzania, and author's calculations.

Table A3: *Potentially confounding factors when switching type of school*

Dependent variable:	Δ Private tuition	Δ Child labour	Δ Child labour hours	Δ Child health	Δ Life satisfaction
	(1)	(2)	(3)	(4)	(5)
$\Delta Private$	-1.990	-0.028	-0.085	-0.001	-0.120
	(0.505)	(0.036)	(0.089)	(0.022)	(0.150)
Sample mean	1.731	.212	.502	.025	4.323
Sample std. dev.	3.831	.409	1.026	.156	1.676
N	3,305	3,305	3,305	3,305	3,305
R^2	0.0098	0.0002	0.0002	0.0000	0.0002

Notes: Simple OLS models are employed. The dependent variables are changes in log of household expenditures to private tuition, work status of the child, log of work hours plus one, whether the child has been hospitalized or visited a traditional healer in the past 12 months, and the average satisfaction with life for adult household members, respectively. *Sample mean* and *Sample std. dev.* are the mean and standard deviation, respectively, of the dependent variable without differencing. Robust Huber-White standard errors are in parentheses.

Source: National Panel Surveys 2010-2011 and 2012-2013, and author's calculations.

Table A4: *Analysis of subject-specific exam scores (continued on next page)*

Dependent variable:	Kiswahili FTNA score (1)	Kiswahili FTNA score (2)	English FTNA score (3)	English FTNA score (4)	Math FTNA score (5)	Math FTNA score (6)
Private$_s$	0.460	0.412	0.566	0.597	0.775	0.768
	(0.015)	(0.014)	(0.026)	(0.020)	(0.024)	(0.024)
PSLE score	0.316	0.398	0.331	0.426	0.397	0.425
	(0.002)	(0.003)	(0.004)	(0.003)	(0.005)	(0.005)
Female	0.056	0.075	-0.108	-0.085	-0.148	-0.086
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.005)
Uncommon name	0.008	0.001	0.017	0.012	0.022	0.023
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Peers PSLE$_s$	0.043	0.034	0.011	-0.002	-0.063	-0.033
	(0.010)	(0.010)	(0.011)	(0.009)	(0.009)	(0.008)
Peers failed$_s$	-0.775	-0.632	-0.935	-0.862	-1.406	-1.309
	(0.075)	(0.073)	(0.073)	(0.064)	(0.078)	(0.077)
Peers with A$_s$	0.354	0.069	0.525	0.260	1.716	1.489
	(0.088)	(0.079)	(0.062)	(0.053)	(0.088)	(0.081)
PSLE score instrumented	No	Yes	No	Yes	No	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes
Primary school FE	Yes	Yes	Yes	Yes	Yes	Yes
‘*Primary school* × *PSLE score* × *Cohort*’ FE	No	No	No	No	No	No
N	635,162	635,162	635,162	635,162	635,162	635,162
R^2	0.276	0.289	0.424	0.480	0.453	0.445

Analysis of subject-specific exam scores (continued from previous page)

Dependent variable:	Kiswahili FTNA score (7)	English FTNA score (8)	Math FTNA score (9)
Private$_s$	0.378	0.475	0.686
	(0.016)	(0.019)	(0.025)
Female	0.129	-0.036	-0.054
	(0.007)	(0.006)	(0.005)
Uncommon name	0.000	0.011	0.019
	(0.005)	(0.005)	(0.005)
Peers PSLE$_s$	0.027	0.110	0.027
	(0.011)	(0.010)	(0.011)
Peers failed$_s$	-0.660	-0.475	-1.296
	(0.083)	(0.068)	(0.086)
Peers with A$_s$	0.410	-0.007	1.208
	(0.084)	(0.048)	(0.091)
PSLE score instrumented	Yes	Yes	Yes
'*Primary school* × *PSLE score* × *Cohort*' fixed effects	Yes	Yes	Yes
N	635,162	635,162	635,162
R^2	0.640	0.751	0.767

Notes: The first six columns apply a standard value-added specification, whereas the last three columns apply a flexible value-added model. *PSLE score* is the subject-specific exam score in primary school. When applying the IV approach, *PSLE score* is instrumented by the exam scores in all other primary school subjects. In addition to the subjects mentioned, exam scores in 'community knowledge' and 'science' are used as instruments. *Peers PSLE$_s$* is the average primary school subject-specific exam score for a student's secondary school schoolmates. *Peers failed$_s$* and *Peers with A$_s$* are the share of secondary school peers who failed primary school and the share of peers who achieved an A in the specific subject in primary school, respectively. *PSLE score*, *Peer effects*, and the dependent variables are standardized by their sample means and standard deviations. Standard errors are clustered at the secondary school level. Source: National Examination Council of Tanzania, and author's calculations.

Appendix B: Supplementary robustness and heterogeneity analyses

The following robustness analyses include reflections on the representativeness of the sample, and different heterogeneity tests interacting the private secondary school enrolment dummy with different geographical areas, primary school type, peer effects, a religious school dummy, a same-gender school dummy, and cohort dummies. The preferred baseline model is further estimated on subsamples excluding students from the best-performing private schools, and excluding students enrolled in large primary school

Representativeness of results

The descriptive statistics in Section III illustrated how the sample at hand differs from the total population of students. Naturally, this makes one wonder whether the estimated private school learning premium is representative to the entire student population in Tanzania. To counter this selection issue, two approaches are pursued. First, sample weights are computed and applied to the baseline model. Second, multiple imputation is employed to predict the values of the missing primary school exam scores. This method is considered superior to single imputation as the imputed values are drawn from a set of potential outcomes, and hence contain more variation.

To counter non-representativeness, sample weights are calculated to give representative estimates in regard to private school enrolment, student gender, uncommonness of name, year of exam, secondary school GPA, ability of peers, and districts. The results from this approach are representative given observations are missing completely at random (MCAR) within weighting classes. A probit model is used to calculate the probability of a student being in the sample. Next, the inverse of the predicted probability of being in the sample is used as sample weights in the preferred baseline model of Section V. The results from including these sample weights are presented in Table B1, column (1). There is a marginal drop in the private school learning premium compared to the baseline model, but the change is insignificantly different from zero.

In order to interpret the baseline results from Section V as representative for the entire student population, one has to assume missing observations are MCAR. A less restrictive assumption is to say observations are missing at random (MAR). This means that conditional on the other variables of the model the sample is random. After imposing the MAR assumption, one may use different forms of imputation to arrive at representative estimates. The conventional single imputation approach is not pursued as it does not adequately account for missing-data uncertainty (Cameron and Trivedi 2005). That is, single imputation leads to an underestimation of standard errors due to the imputed values being perfectly determined by a model based on observed data. Instead, the multiple imputation approach is pursued, where missing data are imputed m times to create m different and complete datasets. The question is how large should m be? Following Schafer (1997) and Cameron and Trivedi (2005), the number of imputations does not need to be high, and a value of ten should be adequate.

The results from estimating the preferred baseline model on the multiply imputed data are reported in Table B1, column (2). The private school learning premium increases by two percentage points when including the missing observations. This suggests that non-imputed students may benefit slightly less from attending private school compared to imputed students. The other explanatory variables – except for the uncommon name indicator – remain significant at the 0.1 percent level. The coefficient estimate associated with $PeersPSLE$, however, changes sign. At the same time, the coefficient estimate associated with $PeerswithA$ increases considerably, suggesting it is mostly

Table B1: *Sample weighting and multiple imputation*

Dependent variable: $GPA_s\ (FTNA)$	(1)	(2)
$Private_s$	0.508	0.558
	(0.016)	(0.016)
$Female$	0.072	0.020
	(0.005)	(0.004)
Uncommon name	0.012	0.002
	(0.003)	(0.003)
Peers $PSLE_s$	0.078	-0.055
	(0.011)	(0.011)
Peers $failed_s$	-0.514	-0.507
	(0.050)	(0.049)
Peers with A_s	0.314	0.744
	(0.048)	(0.048)
GPA $other_p$ *(PSLE)*	0.218	0.183
	(0.002)	(0.002)
'*Primary school* × *Primary school GPA* × *Cohort*' FE	Yes	Yes
N	635,162	1,245,860
R^2	0.745	0.639

Notes: In column (1), sample weights are applied to get a representative sample in regard to private school enrolment, student gender, uncommonness of name, year of exam, secondary school GPA, ability of peers, and districts. Column (2) presents the results from a multiple imputation approach with ten imputations, and where primary school exam scores and school are imputed. R^2 in column (2) is the average R^2 from the ten different models. GPA_s is the GPA of the subjects Kiswahili, English, and mathematics in secondary school. *Peers* $PSLE_s$ is the average GPA of the subjects Kiswahili, English, and mathematics in primary school for secondary school schoolmates. *Peers* $failed_s$ and *Peers with* A_s are the share of secondary school peers who failed primary school and the share of peers with at least one A from the three core subjects in primary school, respectively. *GPA* $other_p$ is the GPA of the subjects 'community knowledge' and 'science' in primary school. GPA_s, *Peer* $effects_s$, and *GPA* $other_p$ are standardized by their sample means and standard deviations. Standard errors are clustered at the secondary school level. Source: National Examination Council of Tanzania, and author's calculations.

interactions with very high-ability peers that makes a difference. Thus, based on the multiple imputation method, the baseline sample in Section V is conceivably not perfectly representative. The potential bias on the private school learning premium from omitting missing students, however, is found to be negative, suggesting a slightly larger private school learning premium for excluded students.

Geographical heterogeneity

The literature finds that lagged achievement captures the majority of relevant student and family characteristics. One may, however, still worry that students attending private secondary schools belong to families who suddenly became wealthier and then were able to afford other educational inputs like private schooling and private tuition. As income changes may differ in urban settings compared to smaller towns and villages, one robustness test investigates heterogeneity in the private school learning premium in regard to urbanization.[30] In addition, data from Tanzania National Panel Surveys in 2013 and 2015 are used to identify regions with the most severe consumption changes. Despite the survey data being only nationally – and not regionally – representative, two additional models extending the preferred baseline model are estimated: 1) including interactions between the private school dummy and two dummies indicating whether a student lives in a top or bottom region in terms of consumption growth; and 2) including an interaction between the private school dummy and the consumption growth rate in the region where the student lives.[31] There are substantial breaks between both the regions with the second lowest and third lowest consumption

[30]Urbanization is defined as cities with more than 100,000 inhabitants, which include Dar es Salaam, Arusha, Mwanza, Mbeya, Morogoro, Tanga, Kigoma, Dodoma, and Songea.

[31]The top regions in terms of consumption growth are Singida and Mbeya, while the bottom two regions are Iringa (including Njombe Region) and Kilimanjaro.

growth rate, and the regions with the second highest and third highest consumption growth rate. This makes them natural breaking points for the robustness analysis.

Table B2 presents the results from the above-mentioned models. From column (1) it is clear that urban areas have a smaller private school learning premium compared to other areas. This is in line with Singh (2015), finding that rural areas in India have a higher private school learning premium. Column (2) shows that high-growth and low-growth regions have similar learning premiums as other regions. Similarly in column (3), the interaction term between private schooling and consumption growth is insignificant. Thus, despite urban areas having a dissimilar learning premium relative to the rest of the country, the difference does not seem to be driven by income changes.

One-off impact of private schools

Andrabi et al. (2011) examined the impact of Pakistani primary school students switching between public and private schools. In a difference-in-differences model, they compare test score gains for students between third and fourth grade, and between third and fifth grade. The one-year gains (comparing test scores between third and fourth grade) from switching to a private school are 0.26–0.31 of a standard deviation, whereas the two-year gains are only slightly higher. The authors argue this is due to low persistence in test scores. While low persistence may explain some of the lower gains in the second year of private schooling, it seems implausible that low persistence should cause as small academic gains in the second year as the authors find. Another explanation for the lower academic gains in the second year of private schooling could be that private schools pick the low-hanging fruits in terms of improving students' test scores in the first year. Irrespective of the exact reasons for lower academic gains in the second year, the results suggests that one should try to distinguish between students coming from private and public schools when estimating the private

Table B2: *Analysis of geographical heterogeneity*

Dependent variable: $GPA_s\ (FTNA)$	(1)	(2)	(3)
$Private_s$	0.576	0.525	0.558
	(0.019)	(0.019)	(0.021)
$Urban_s$	0.018		
	(0.012)		
$Private_s \times Urban_s$	-0.126		
	(0.024)		
$High\ growth_s$		0.032	
		(0.019)	
$Low\ growth_s$		0.061	
		(0.016)	
$Private_s \times High\ growth_s$		0.058	
		(0.047)	
$Private_s \times Low\ growth_s$		0.019	
		(0.029)	
$Region\ growth_s$			-0.002
			(0.001)
$Private_s \times Region\ growth_s$			0.002
			(0.001)
'*Primary school* $\times$ *Primary school GPA* $\times$ *Cohort*' FE	Yes	Yes	Yes
N	635,162	635,162	635,162
R^2	0.739	0.738	0.738

Notes: GPA_s is the GPA of the subjects Kiswahili, English, and mathematics. GPA_s is standardized by its sample mean and standard deviation. $High\ growth_s$ is a dummy taking the value 1 if the school is located in Singida or Mbeya, whereas $Low\ growth_s$ is a dummy taking the value 1 if the school is located in Iringa, Njombe, or Kilimanjaro. $Region\ growth_s$ is a continuous variable measuring consumption growth between 2013 and 2015 for the region in which the school is located. The models further account for student gender, uncommonness of student name, the average GPA of the subjects Kiswahili, English, and mathematics in primary school for secondary school schoolmates, share of peers who failed primary school, share of peers who achieved at least one A in primary school, and the GPA of the subjects 'community knowledge' and 'science' in primary school. Standard errors are clustered at the secondary school level.

Source: National Examination Council of Tanzania, National Panel Surveys 2010-2011 and 2012-2013, and author's calculations.

Table B3: *Analysis of primary school type heterogeneity*

Dependent variable: $GPA_s\ (FTNA)$	(1)
$Private_s$	0.609 (0.019)
$Private_s \times Private_p$	-0.280 (0.020)
'*Primary school* $\times$ *Primary school GPA* $\times$ *Cohort*' FE	Yes
N	625,050
R^2	0.738

Notes: Primary school type could not be determined for 10,112 students. GPA_s is the GPA of the subjects Kiswahili, English, and mathematics. It is standardized by the sample mean and standard deviation. The model further accounts for student gender, uncommonness of student name, the average GPA of the subjects Kiswahili, English, and mathematics in primary school for secondary school schoolmates, the share of peers who failed primary school, the share of peers who achieved at least one A in primary school, and the GPA of the subjects 'community knowledge' and 'science' in primary school. Standard errors are clustered at the secondary school level.
Source: National Examination Council of Tanzania, and author's calculations.

school learning premium.

Table B3 estimates the private school learning premium separately for students attending public primary school and students attending private primary school. The results suggest the impact of private schools is larger for students who went to public primary school. The coefficient estimate is around twice the size for public primary school students compared to private primary school students. These results are in line with Andrabi et al. (2011), finding the private school learning premium is larger in the first year of private schooling.

Private schools capitalizing on peer effects

The current paper has endeavoured to distinguish any private school learning premium from peer effects. The premise has been that private schools are in essence not better than public schools if shifting the student populations would make public schools perform better than private schools. While private schools may improve students' exam scores through both peer effects and a more genuine private school learning premium, the two mechanisms could also jointly affect students' exam scores. Given either public schools or private schools are better at capitalizing on peer effects, the interaction term between the two mechanisms is relevant to include in the model.

As emphasized in Section IV, the impact of peer effects may well be more complex than the linear relationship assumed in the baseline results. While the current paper does not seek to explore every avenue of peer effects, it is useful to separate the peer effect variable into two main categories: 1) peer effects from students with the same gender; and 2) peer effects from students with the opposite gender. Attending the same school as high-ability peers may have limited impact if one does not interact with these high-ability peers. Assuming students interact the most with other students of the same gender, one may expect the peer effects from students with the same gender to be of higher importance than peer effects from students with the opposite gender.

Table B4 presents the results from a model including the interaction between private schooling and the primary school exam scores of current peers in column (1). Part of the private school learning premium seems to be driven by private school students benefiting more from peer effects than public school students. As discussed in the methodology section, this does not necessarily mean that private schools are better at capitalizing on peer effects as the peer effect variable could be capturing family-supplied inputs. The peer effects variable ranges between -2.8 and 4.2. Consequently, the

Table B4: *Analysis with private school and peer effects interaction*

Dependent variable: GPA_s $(FTNA)$	(1)	(2)
$Private_s$	0.470	0.432
	(0.021)	(0.021)
$Peers\ PSLE_s$	0.078	
	(0.011)	
$Private_s \times Peers\ PSLE_s$	0.067	
	(0.013)	
$Peers\ PSLE\ own_s$		0.047
		(0.007)
$Private_s \times Peers\ PSLE\ own_s$		0.028
		(0.010)
$Peers\ PSLE\ opposite_s$		-0.005
		(0.006)
$Private_s \times Peers\ PSLE\ opposite_s$		0.052
		(0.009)
'*Primary school* $\times$ *Primary school GPA* $\times$ *Cohort*' FE	Yes	Yes
N	635,162	605,428
R^2	0.739	0.708

Notes: GPA_s is the GPA of the subjects Kiswahili, English, and mathematics. $Peers\ PSLE_s$ is the average GPA of the subjects Kiswahili, English, and mathematics in primary school for all secondary school schoolmates. $Peers\ PSLE\ own_s$ and $Peers\ PSLE\ opposite_s$ are the average GPA of the subjects Kiswahili, English, and mathematics in primary school for secondary school same gender schoolmates and opposite gender schoolmates, respectively. $Peers\ PSLE_s$, $Peers\ PSLE\ own_s$, $Peers\ PSLE\ opposite_s$, and GPA_s are standardized by their sample means and standard deviations. Column (2) considers only mixed-gender schools. The models further account for student gender, uncommonness of student name, the share of peers who failed primary school, the share of peers who achieved at least one A in primary school, and the GPA of the subjects 'community knowledge' and 'science' in primary school. Standard errors are clustered at the secondary school level.
Source: National Examination Council of Tanzania, and author's calculations.

private school learning premium is positive independent of current peers.

The sample in column (2) is limited to schools with both boys and girls. The results demonstrate that the overall peer effects found in the baseline results are driven by same gender peer effects. This result is in line with the expectation of students interacting mostly with students of the same gender. For private school students, however, the primary school achievement of peers of the opposite gender is positively correlated with secondary school exam scores, suggesting private school students are more likely to interact with peers of the opposite gender.

Heterogeneity on other observables

The current robustness analyses test whether schools that offer religious courses have a different private school learning premium, whether same-gender schools vary from mixed-gender schools, and whether cohorts are distinctively different. In the religion analysis, the private school dummy is interacted with a dummy indicating whether or not a school offers elective courses in either Bible knowledge or Islamic knowledge. This 'religious courses' dummy is further divided into a 'Bible course' dummy and an 'Islamic course' dummy. Similar to the religion analysis, the same-gender analysis interacts the private school dummy with a dummy stating whether a school is a same-gender school. This same-gender school dummy is further divided into girls-only schools and boys-only schools. Finally, due to a major secondary school reform taking place at the beginning of 2016, the three cohorts are examined separately. The 2015 cohort took the FTNA before the implementation of the reform, whereas the 2016 cohort had attended secondary school for one year under the new reform, and the 2017 cohort had only attended secondary school under the new reform.

Table B5 shows the impact of attending a school offering a religious course compared to a school not offering a religious course. There seems to be no overall effect of offering religious courses neither in the public school sector nor in the private school sector. Examining schools offering Bible knowledge and Islamic knowledge separately, however, it is evident that schools offering Islamic

Table B5: *Analysis of secondary schools offering religious courses*

Dependent variable: GPA_s $(FTNA)$	(1)	(2)	(3)
$Private_s$	0.550	0.536	0.550
	(0.020)	(0.019)	(0.018)
$Religious\ courses_s$	-0.013		
	(0.009)		
$Private_s \times Religious\ courses_s$	-0.025		
	(0.023)		
$Bible\ course_s$		-0.001	
		(0.017)	
$Private_s \times Bible\ course_s$		0.013	
		(0.028)	
$Islamic\ course_s$			-0.018
			(0.010)
$Private_s \times Islamic\ course_s$			-0.129
			(0.029)
'*Primary school* × *Primary school GPA* × *Cohort*' FE	Yes	Yes	Yes
N	635,162	635,162	635,162
R^2	0.738	0.738	0.739

Notes: *Religious courses* is an indicator for whether the school offers either Bible knowledge or Islamic knowledge as elective courses. GPA_s is the GPA of the subjects Kiswahili, English, and mathematics. GPA_s is standardized by its sample mean and standard deviation. The models further account for student gender, uncommonness of student name, the average GPA of the subjects Kiswahili, English, and mathematics in primary school for secondary school schoolmates, the share of peers who failed primary school, the share of peers who achieved at least one A in primary school, and the GPA of the subjects 'community knowledge' and 'science' in primary school. Standard errors are clustered at the secondary school level.

Source: National Examination Council of Tanzania, and author's calculations.

knowledge have a lower private school learning premium. This result is independent of whether the religion dummy is based on at least one student, at least 25 percent of students, at least 50 percent of students, or at least 75 percent of students attending the religious course. Despite the negative coefficient estimate on the interaction between private schooling and Islamic course, private schools

Table B6: *Analysis of same-gender secondary schools*

Dependent variable: GPA_s $(FTNA)$	(1)	(2)	(3)	(4)
$Private_s$	0.519	0.548	0.512	0.519
	(0.017)	(0.017)	(0.018)	(0.017)
$Same\ gender\ school_s$	0.176			
	(0.032)			
$Private_s \times Same\ gender\ school_s$	0.047			
	(0.038)			
$Boys\ school_s$		0.150		0.200
		(0.065)		(0.066)
$Private_s \times Boys\ school_s$		-0.046		-0.016
		(0.073)		(0.074)
$Girls\ school_s$			0.119	0.158
			(0.031)	(0.029)
$Private_s \times Girls\ school_s$			0.089	0.080
			(0.039)	(0.037)
'*Primary school* $\times$ *Primary school GPA* $\times$ *Cohort*' FE	Yes	Yes	Yes	Yes
N	635,162	635,162	635,162	635,162
R^2	0.739	0.739	0.739	0.739

Notes: GPA_s is the GPA of the subjects Kiswahili, English, and mathematics. GPA_s is standardized by its sample mean and standard deviation. The models further account for student gender, uncommonness of student name, the average GPA of the subjects Kiswahili, English, and mathematics in primary school for secondary school schoolmates, the share of peers who failed primary school, the share of peers who achieved at least one A in primary school, and the GPA of the subjects 'community knowledge' and 'science' in primary school. Standard errors are clustered at the secondary school level.
Source: National Examination Council of Tanzania, and author's calculations.

that offer Islamic knowledge still outperform public schools.

The effect of attending a same-gender school is presented in Table B6. Column (1) demonstrates that students attending same-gender secondary schools perform better than their mixed-gender counterparts. The private school learning premium, however, is insignificantly different from the learning premium for students attending mixed-gender schools. Columns (2), (3), and (4) illustrate that students in both boys-only and girls-only schools perform better than students in mixed-gender schools. Further, girls-only school students have a higher (borderline significant) private school learning premium compared to students in mixed-gender and boys-only schools.

Table B7 presents the results from studying the cohorts in separate regression models. There are only small deviations in the results between the three years. While the coefficient estimate on private schooling drops from 2015 to 2016, the change is insignificantly different from zero. The 2017 cohort was fully affected by the new reform in 2016. This led to a substantial increase in enrolment into public secondary schools. Assuming this increase in enrolment had negative effects on quality of education, as seemed to be the case when Tanzania introduced free primary school, one would expect the private school learning premium to increase in 2017. In line with expectations, the private school learning premium is significantly higher in 2017 compared to the two previous years.

Accounting for public and private elite schools

It is well-established that the private sector serves a very heterogeneous group of students. For instance, the private sector may serve both children from the slums of big cities, children in remote areas without any public school, and children of wealthy parents enrolled in elite international schools. Heyneman and Stern (2014), however, state there is anecdotal evidence of low-cost private schools actually being of poor quality. Consequently, one may suspect the private school learning premium estimated in baseline model is purely driven by children enrolled in expensive elite schools.

To examine whether the learning premium is driven by elite schools, students enrolled in the best performing private schools are excluded from the sample. Assuming a positive correlation between absolute and value-added performance, this procedure could create a mechanically lower

Table B7: *Cohorts analysed separately*

Dependent variable: $GPA_s\ (FTNA)$	(1)	(2)	(3)	(4)
$Private_s$	0.484	0.482	0.626	0.505
	(0.025)	(0.021)	(0.021)	(0.021)
$Private_s \times Cohort16$				-0.012
				(0.019)
$Private_s \times Cohort17$				0.098
				(0.020)
'*Primary school* $\times$ *Primary school GPA*' FE	Yes	Yes	Yes	Yes
FTNA cohort	2015	2016	2017	All
N	176,206	198,259	260,697	635,162
R^2	0.752	0.750	0.722	0.738

Notes: GPA_s is the GPA of the subjects Kiswahili, English, and mathematics. GPA_s is standardized by its sample mean and standard deviation. The models further account for student gender, uncommonness of student name, the average GPA of the subjects Kiswahili, English, and mathematics in primary school for secondary school schoolmates, the share of peers who failed primary school, the share of peers who achieved at least one A in primary school, and the GPA of the subjects 'community knowledge' and 'science' in primary school. Standard errors are clustered at the secondary school level.
Source: National Examination Council of Tanzania, and author's calculations.

private school learning premium. First, the top one percent of private schools are excluded. Next, household panel survey data from 2015 are employed to calculate average private secondary school fees. Removing the top 15 percent gives an average of 601,000 Tanzanian shillings paid in school fees per year.[32] This number is only slightly larger than the average spending of 565,000 per student in public secondary schools in the financial year 2015/2016.[33] Thus, as a second robustness model,

[32]On January 1 2015, 601,000 Tanzanian shillings were equivalent to 346 US dollars.

[33]Total educational expenditures were budgeted at 4.0 trillion Tanzanian shillings in the financial year 2015/2016, and approximately 18 percent are spent on secondary education (UNICEF 2016). Further 5 percent are spent on administration, and it is assumed 1 percentage point is attributable to secondary education. In 2015, 1,774,383 students were enrolled in secondary school (World Bank 2019). While a

students enrolled in the top 15 percent private schools are excluded from the sample. Lastly, making an even harsher test for whether low-performing private schools are of poorer quality than public schools, the 50 percent best performing private schools are excluded from the sample.

It is generally perceived that some of the best-performing secondary schools are public schools, often referred to as 'elite public schools'. Given these schools are indeed 'elite schools' and on a par with privately operated schools, the private school learning premium should be indifferent from null. To investigate this hypothesis, a sample of students progressing to the 1 percent best-performing public schools are compared to their primary school peers obtaining the same primary school exam scores and progressing to a private school.

The results from estimating the preferred baseline model without students from the best performing private schools are reported in Table B8. In column (1) it is very few private school students that have been excluded, and the private school learning premium is identical to the full sample model. The second column excludes substantially more private school students, and the private school learning premium drops to 0.50 of a standard deviation. Finally, removing students from the best half of all private schools naturally yields a significantly lower private school learning premium. Surprisingly, however, even for the worst half of private schools, there is still a substantial private school learning premium of 0.30 of a standard deviation. This result is in contrast to the anecdotal evidence claimed by Heyneman and Stern (2014). In regard to students progressing to the best-performing public schools, they achieve lower exam scores at the FTNA relative to their primary school peers who progressed to a private school. The magnitude is similar to the baseline

private and public school distinction is only available two years before, it shows 78.6 percent of secondary school students were enrolled in public schools. In addition, public school students paid a school fee of 20,000 shillings per year before 2016. Thus, the average spending per public school student is calculated as: $4,000,000,000,000 \times (0.18+0.01)/(1,774,383 \times 0.786)+20,000$.

Table B8: *Examining elite schools*

Dependent variable: $GPA_s\ (FTNA)$	(1)	(2)	(3)	(4)
$Private_s$	0.540	0.498	0.301	0.548
	(0.017)	(0.017)	(0.018)	(0.038)
$Female$	0.058	0.058	0.055	0.061
	(0.005)	(0.005)	(0.005)	(0.028)
Uncommon name	0.014	0.016	0.016	-0.009
	(0.003)	(0.003)	(0.003)	(0.019)
Peers $PSLE_s$	0.069	0.077	0.057	0.344
	(0.011)	(0.010)	(0.010)	(0.043)
Peers $failed_s$	-0.604	-0.450	-0.091	0.251
	(0.062)	(0.059)	(0.057)	(0.233)
Peers with A_s	0.295	0.157	0.218	-0.352
	(0.048)	(0.047)	(0.051)	(0.154)
GPA $other_p$ *(PSLE)*	0.223	0.225	0.226	0.200
	(0.002)	(0.002)	(0.002)	(0.059)
'*Primary school* × *Primary school GPA* × *Cohort*' FE	Yes	Yes	Yes	Yes
N	633,747	617,024	591,916	15,774
R^2	0.734	0.704	0.684	0.702

Notes: Columns (1), (2), and (3) exclude students from the top 1, 15, and 50 percent of private schools in terms of average exam scores, respectively. Column (4) includes only students progressing to the 1 percent best-performing public schools and their primary school peers progressing to a private school. GPA_s and GPA_p are the GPA of the subjects Kiswahili, English, and mathematics in secondary and primary school, respectively. *Peer* $effects_s$ is the average GPA of the subjects Kiswahili, English, and mathematics in primary school for secondary school schoolmates. *Peers* $failed_s$ and *Peers with* A_s are the share of secondary school peers who failed primary school and the share of peers with at least one A from the three core subjects in primary school, respectively. *GPA* $other_p$ is the GPA of the subjects 'community knowledge' and 'science' in primary school. GPA_s, GPA_p, *Peer* $effects_s$, and *GPA* $other_p$ are standardized by their sample means and standard deviations. Standard errors are clustered at the secondary school level.

Source: National Examination Council of Tanzania, and author's calculations.

model, and the result holds even after excluding the best-performing private schools. This finding is in line with Lucas and Mbiti (2014) demonstrating in a regression discontinuity framework that public schools perceived as ‘elite schools’ in Kenya do not live up to their reputation.

Students from small primary schools

Section II on theoretical considerations assumes primary school students achieving the same GPA in the same primary school and the same year receive the same school inputs. In large primary schools, however, students may not attend the same classes and therefore they may be subject to varying teacher inputs and peers. One may account for this issue by considering only students that were enrolled in primary schools where there most likely was only one class. Students from primary schools with less than 56 exam takers are considered in Table B9 column (1). This cut-off point is chosen as the average number of students per class in grade 7 in Tanzania was 55 in 2013 (World Bank 2019). The model is further tested on a sample of students enrolled in primary schools with 40 or fewer students, which is the national class size target.

The results in Table B9 demonstrates that excluding students enrolled in large primary schools does not significantly impact the private school learning premium. This is evident both for students from primary schools with less than the average students per class, and the more conservative sample of schools with less than 41 exam takers. When limiting the sample to schools with less than 41 exam takers, the learning premium drops by five percentage points. This change, however, is insignificantly different from zero (p-value of 0.108). Thus, increasing the probability of students being matched with other students in the same primary school class is not significantly affecting the baseline results.

Table B9: *Students from small primary schools*

Dependent variable: $GPA_s\ (FTNA)$	(1)	(2)
$Private_s$	0.520	0.489
	(0.023)	(0.026)
$Female$	-0.021	-0.035
	(0.007)	(0.009)
Uncommon name	0.013	0.010
	(0.006)	(0.008)
Peers PSLE$_s$	0.070	0.069
	(0.016)	(0.020)
Peers failed$_s$	-0.470	-0.461
	(0.091)	(0.114)
Peers with A$_s$	0.285	0.265
	(0.066)	(0.083)
GPA other$_p$ (PSLE)	0.222	0.223
	(0.004)	(0.005)
'*Primary school* $\times$ *Primary school GPA* $\times$ *Cohort*' FE	Yes	Yes
N	262,560	145,490
R^2	0.786	0.812

Notes: The samples in columns (1) and (2) consist of students enrolled in primary schools with no more than 55 and 40 students, respectively, taking the primary school exams. GPA_s and GPA_p are the GPA of the subjects Kiswahili, English, and mathematics in secondary and primary school, respectively. *Peer effects$_s$* is the average GPA of the subjects Kiswahili, English, and mathematics in primary school for secondary school schoolmates. *GPA other$_p$* is the GPA of the subjects 'community knowledge' and 'science' in primary school. GPA_s, GPA_p, *Peer effects$_s$*, and *GPA other$_p$* are standardized by their sample means and standard deviations. Standard errors are clustered at the secondary school level.
Source: National Examination Council of Tanzania, and author's calculations.

Chapter 2

Private School Competition in Kenya: Do Students Learn More?

Private School Competition in Kenya: Do Students Learn More?*

Kasper Brandt[1] and Sam Jones[1]

[1]*Department of Economics, University of Copenhagen*

The impacts of private school enrolment is studied predominantly at the individual level. This approach, however, limits the ability to provide solid policy recommendations accounting for potential spillover effects. The current paper aggregates individual level information to the district-cohort-year level, and employs a system generalized method of moments estimator to study the effects of private school enrolment on average latent ability. The results suggest a substantial positive impact on learning from private school enrolment. A 10 percentage point increase in the proportion of children enrolled in private schools is expected to improve average latent ability in a district-cohort by 0.11 of a standard deviation. (*JEL* H44, I21, O15)

*We wish to thank Simon Quinn, Steve Bond, and Henrik Hansen for their valuable feedback and discussions about the paper.

I. Introduction

Funding for education remains a major policy concern, in particular in developing countries. Since 2000, the primary school net enrolment rate for low-income countries has increased from 56 to 81 percent, translating into more than 60 million additional students from a baseline of 54 millions (UNESCO 2020). While efforts to lower user costs of education to get children into public-funded schools are arguably a main driver for this development, the number of primary school students in a private institution in low-income countries has also increased by 9 million students. These private schools create a dilemma to policy makers. On one hand, these schools take over a funding responsibility that the state should otherwise take and they may improve learning relative to public schools. On the other hand, the government loses control over school operation and inequality may widen given it is mostly wealthy families sending their children to private schools. Independent of whether one favours public or private education, however, it is important to keep in mind that getting children to school – public or private – is only an intermediate goal (Pritchett 2013). The end goal is for them to learn.

Whether private schools improve learning relative to public schools is an empirical question, which has usually been found to be true in developing countries (Angrist et al. 2002; Andrabi et al. 2011; Muralidharan and Sundararaman 2015; Singh 2015; Brandt 2018; Wamalwa and Burns 2018). As Urquiola (2016) argues, two other conditions must be met before concluding on the desirability of private schools. First, learning improvements cannot be driven solely by financial superiority. It is generally, however, found that private schools are as cheap or cheaper to operate in terms of per-pupil costs (Psacharopoulos 1987; Jimenez, Lockheed, and Paqueo 1991; Lassibille, Tan, and Sumra 2000; Alderman, Orazem, and Paterno 2001; Andrabi, Das, and Khwaja

2008; Schirmer 2010; Tooley et al. 2011; Bold et al. 2013; Ngetich, Wambua, and Kosgei 2014; Muralidharan and Sundararaman 2015). Second, learning improvements cannot come at the expense of learning in public schools. As most studies are based on individual observations, it is difficult to consider spillover effects.[1] One alternative is to move the unit of observation to local education markets (Urquiola 2016). Hsieh and Urquiola (2006) does so for Chile, finding worsening effects on repetition rates and years of schooling from private school expansion. For India and Kenya, Pal and Kingdon (2010) and Bold et al. (2013), respectively, find that districts with rising private school shares also experience improvements in literacy and test scores. As acknowledged by the authors, however, concerns about reverse causality arise in the simple ordinary least squares (OLS) framework.

The present paper shows that average learning in a district improves by an expected 0.11 standard deviation when expanding the share of students enrolled in private schools by 10 percentage points. We use item response theory (IRT) to derive a latent ability measure based on 235,000 individual cognitive tests taken over a five-year period for 6–10 year old children in Kenya. Model variables are aggregated to the district-cohort-year level to account for spillover effects. We include the lagged dependent variable to account for dynamics of learning. Utilizing the system generalized method of moments (GMM) estimator, we use internal instruments for potentially predetermined and endogenous variables. For instance, private school shares in period *t-2* and before are used as instruments for the contemporary change in the private school share, and the change from *t-2* to *t-1* and prior changes are used as instruments for the level in period *t*. The significantly positive effect of

[1] Muralidharan and Sundararaman (2015) are able to estimate spillover effects due to the design of the voucher programme they consider. In some villages, a share of students were offered a private school voucher, whereas other villages had no students being offered a private school voucher. Comparing public school students in these two types of villages yielded little to no evidence of spillover effects on public school students.

private education is robust to: 1) restricting the sample to district-cohorts present in four consecutive time periods instead of an unbalanced sample; 2) excluding all endogenous variables except private schooling; 3) disaggregating the unit of observation from district-cohort to district-cohort-gender; 4) using the difference GMM estimator; and 5) using average test scores instead of item response theory.

By way of structure, the remainder of the paper goes as follows: Section II discusses some of the primary challenges involved in identifying a causal private education premium. Section III describes the context and the applied data. Section IV develops the empirical identification strategy. Section V presents the results. Section VI concludes.

II. Challenges in Identifying the private school premium

Estimation of a causal impact due to private education is notoriously problematic (e.g., Vandenberghe and Robin 2004). At least three sets of problems prevail. The first refers to endogenous selection. Factors that are likely to determine use of private provision include a range of (often) unobserved elements including individual ability, household conditions (e.g., total wealth, intra-household allocation of resources), local public school quality and perceived idiosyncratic returns to different types of education. Controlling for these factors via a selection-on-observables strategy puts heavy requirements on to the applied data, which has motivated use of exogenous instrumental variables or experimentally imposed random variation in access to private schooling. However, these strategies also have been criticized, in part due to the problem of essential heterogeneity. As Ravallion (2015) clarifies, if differences in expected gains from private education generate selective behavioural responses to the option of take-up, estimates based on an experimentally randomized instrumental variable may be seriously biased.

A second challenge relates to peer effects – do gains from private schools reflect an innate contribution of these schools regardless of the students, or do they simply capture a spillover effect from learning with a particular set of peers?[2] This is important as it raises the question of the appropriate counter-factual, as well as potential general equilibrium effects associated with expansion of private schooling options. Indeed, following Hsieh and Urquiola (2006), concerns have been raised that large-scale reforms that promote school choice may lead to greater sorting (segregation) between students, which may particularly disadvantage lower-ability students remaining in government schools.

The third challenge, which is connected to the second, refers to competition effects on public schools from growth of private schools. Establishment of a single private school in a particular school district would be unlikely to have a material effect on any single school or classroom (public school peer group). However, where private schooling options become widespread, public schools may face substantial competition to attract students. Assuming public school funding is associated with student numbers, the net effect may be to raise productivity in the government school sector (for discussion see Hoxby 2003).

The relevant econometric implication of these concerns is that private education may impact learning beyond the direct effect of receiving different school inputs. More specifically, learning may depend on the existing private education take-up rate and ability of peers. To incorporate these ideas in a simple framework, consider a generic education production function for child i from

[2] For elaboration and discussion of evidence on peer effects in schools see Sacerdote (2011).

household j in location k at time t:

$$y_{ijkt} = \alpha_i + \beta_1 p_{it} + \beta_2 \bar{p}_{(-i)kt} + \beta_3 \bar{y}_{(-i)k(t-1)} + C'_{it}\gamma + H'_{jt}\delta + L'_{kt}\lambda + \varepsilon_{ijkt} \tag{1}$$

where y is a metric of cognitive achievement; α_i is (unobserved) individual ability; p_{it} takes a value of one if the child uses private education services; $\bar{p}_{(-i)kt}$ is the average of the same measure for her local age cohort (the peer group, excluding i); $\bar{y}_{(-i)k(t-1)}$ is the lagged average cognitive achievement for child i's local age cohort; C, H, L are sets of vectors capturing time-variant child-, household- and location-specific factors; and ε is white noise error. The coefficients β_1 and β_2 capture the direct effect of private education and contextual peer effects associated with private education, respectively. The coefficient β_3 captures the impact of having peers with higher cognitive achievement. The peer effects term is lagged to avoid contemporary school effects simultaneously influencing child i's achievement and her peers' achievement.[3] Estimates in the literature often set $\beta_2 = \beta_3 = 0$ by assumption. This may be convenient, but there is also a risk that channels of learning remain uncaptured and potentially biasing β_1.

How might estimates of Equation 1 proceed? Despite the presence of control variables, endogeneity remains a concern. For instance, innate ability is plausibly correlated with take-up of private education, inducing sorting bias (also known as cream-skimming). One strategy for addressing sorting bias is to aggregate the data into district-level cells (Hsieh and Urquiola 2006; Pal and

[3] While endogenous peer effects are mitigated by considering the lagged achievement of peers, one can still imagine a case where child i impacts her peers in time period *t-1*, and the peers impact child i in time period t. These endogenous peer effects are excluded such that the coefficient associated with the peer effects term can be understood as a reduced form.

Kingdon 2010; Bold et al. 2013), yielding:

$$\bar{y}_{kt} \approx \bar{\alpha}_k + \bar{p}_{kt}(\beta_1 + \beta_2) + \beta_3 \bar{y}_{k(t-1)} + \bar{Z}'_{kt}\phi + \varepsilon_{kt} \tag{2}$$

where $\bar{Z} = (\bar{C}, \bar{H}, \bar{L})$. While the lagged outcome variable represents the impact of having peers with higher achievement, it may also capture different inputs to learning until time period *t-1*. That is, if one cannot account for all historic learning inputs to child *i*, a second best is to control for lagged achievement.[4] The above specification also combines different impact channels, preventing an analysis of the distinction between individual versus wider spillover effects. In addition, aggregating from the individual to the district level entails limitations on its own. For instance, a misspecified model is likely to cause an exacerbated bias from omitted variables (Hanushek, Rivkin, and Taylor 1996). As estimating Equation 2 based on a district fixed effect model causes dynamic panel bias (see Section IV), the empirical analysis applies internal instruments to counter endogeneity issues.

III. Context and Data

A. *Context*

The private school sector in Kenya consists of a wide variety of providers: religious faith-based schools, local community schools, low-fee schools concentrated in urban slums, and elite schools for children of rich parents. The total size of these private providers is profound in primary education. The most recent data from UNESCO in 2014 shows that around 16 percent of primary school

[4] Value-added models have been found to produce little to no bias relative to lottery-based estimates (Angrist et al. 2017) and regression discontinuity estimates (Singh 2019). In addition, the inclusion of a lagged term substantially mitigates the potential bias from omitted socio-economic variables (Andrabi et al. 2011; Deming 2014; Muralidharan and Sundararaman 2015; Elks 2016).

students attend private schools, which is an increase of 5.4 percentage points since 2009 (UNESCO 2020). Compared to other countries in the region, Kenya and Uganda have similar rates of private school enrolment in primary education, whereas the rates of private provision in Tanzania, Rwanda, and Burundi are substantially lower at around 2-3 percent.

Part of the reason for the increase in private school enrolment can be found in 2003 when free public primary education was re-introduced. While it is natural to think the private school share declines when the relative cost increases, the opposite happened. Enrolment into private schools soared. Potential explanations include extended support from the government to low-fee private schools (Edwards Jr., J. Klees, and L. Wildish 2017), and public school students switching to private schools due to perceived declining quality of public schools and academically weaker peers (Oketch et al. 2010; Bold et al. 2011; Lucas and Mbiti 2012; Nishimura and Yamano 2013).

The financial strength of schools differs considerably within the private school sector. Bold et al. (2013) find the median cost per pupil is twice as large in public primary schools compared to private primary schools in 2005-2006. The mean cost per pupil, however, is larger in private primary schools, revealing the large cost variation in the private sector. Comparing similar points in the cost distributions for public and private schools, two thirds of schools in the private sector are cheaper to operate than public schools. Similarly for secondary schools in 2009 in Nandi North District, Ngetich, Wambua, and Kosgei (2014) show that per pupil costs in private schools were lower than per pupil costs in public schools.[5] While funding structures may have changed between 2006 and the end of our sample period in 2015, we cannot reject that private primary schools operate at approximately similar or lower costs per pupil than public primary schools.

[5] There were only two private schools and 39 public schools in the sample. The highest cost private school operated at marginally higher costs per pupil than the public school average, whereas the lowest cost private school operated at costs per pupil 34 percent under the public school average.

B. Data

The current paper relies on household surveys conducted by Twaweza in Kenya from 2011 to 2015 (also known as the 'Uwezo data').[6] Children in the age group 6–16 years are tested in their reading abilities in Swahili and English, math abilities, and general knowledge. To avoid high risk of measurement error due to most older children achieving a perfect score, the sample is limited to the age group 6-10 years, corresponding to approximately grade 1-4.[7] This leaves us with 235,000 students tested between 2011 and 2015. In order to account for some questions being more difficult than others, item response theory (IRT), discussed below, is used to generate an underlying latent ability variable, which will serve as the dependent variable. As the Uwezo data is based on a repeated cross-sectional design, children enter only once in the dataset. To be able to account for dynamics of learning, local market fixed effects, and spillover effects, a district panel is created by aggregating individual observations to the district-cohort-year level. This also implies, however, that we can only study a composite of learning within the district.

Item Response Theory

In general, two common methods for measuring student ability, based on tests, are pursued: 1) classical test theory (CTT); and 2) item response theory (IRT).[8] CTT is most commonly applied. This theory assumes that a student's observed test score reflects the student's true ability together with an error term. It is assumed that each test question (or each placement on an ordinal scale) is equally good at measuring ability for all students. While we find IRT superior to CTT in our setting,

[6] The survey started in 2009, but we do not utilize the first round due to variable inconsistency.
[7] The test was adapted to the expected level of a grade 2 student. 16 and 32 percent of all sample children and the oldest cohort (10-year-old children), respectively, achieve the highest score.
[8] For a detailed explanation of IRT, we refer to van der Linden (2016).

a robustness analysis estimates the baseline model with a dependent variable based on CTT.

IRT is different from CTT as it acknowledges some questions (or placements on an ordinal scale) are more difficult than others, and some questions are better at discriminating between different levels of true ability. That is, answering a tough question correctly may say more about latent ability than answering an easy question correctly, or moving from very bad to bad in a test may have a different impact on latent ability compared to moving from good to very good. Furthermore, some questions may either be so difficult that it is pure luck for the students who answer them correctly or the knowledge required to give a correct answer is random and undetermined by latent ability. In these cases, the questions are bad at discriminating between different levels of latent ability, and IRT will account for this by giving them a low discrimination parameter.

Frequently, models based on IRT are called Rasch models named after the Danish mathematician Georg Rasch. The simplest model is the dichotomous Rasch model, where respondents can answer either yes or no.[9] The Polytomous Rasch model, on the other hand, is a generalization of the simple dichotomous model, allowing for questions with an ordinal scale. That is, several answers (or outcomes) are allowed and there is a natural order in terms of correctness/ability. The distance in correctness, however, between two answers is unknown. One famous polytomous Rasch model is the partial credit model developed by Masters (1982). This model is preferred as it allows for different scales of the test questions.

For the estimation of a latent ability variable, four aspects of a student's academic ability are used: (1) reading ability in Swahili, where the student can either read nothing, letters, words, a paragraph, or the entire text given; (2) reading ability in English, where the student can either read nothing, letters, words, a paragraph, or the entire text given; (3) numeric ability, where the

[9] In education studies one would usually have correct/incorrect instead of yes/no.

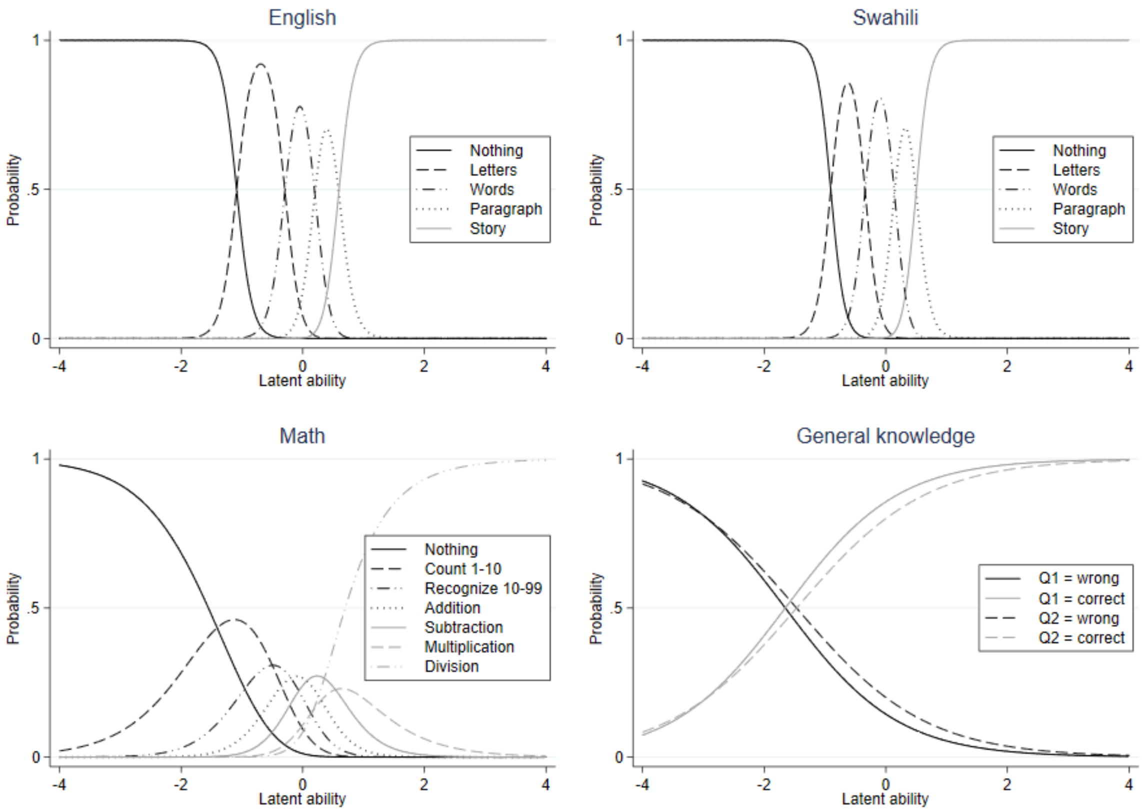
student can either do nothing, count and match numbers to pictures with a given number of objects, recognize numbers between 10–99, do addition, do subtraction, do multiplication, or do division; and (4) general knowledge, where the student is given two correct/incorrect questions about everyday life.

Estimations of latent ability are run separately for each year. Category characteristic curves from estimating latent ability for children in 2011 are presented in Figure 1. Similar category characteristic curves for 2012–2015 can be seen in Appendix Figures A1–A4. Each category is, as expected, able to explain different parts of the latent ability spectrum. That is, if a student can only read letters, the student is most likely in the lower part of the latent ability spectrum, and if doing division is mastered, the student is most likely in the upper part of the latent ability spectrum. The fourth graph shows that the difficulty level of the two general knowledge questions is similar. Furthermore, the steepness of the curves illustrate how well the categories are at discriminating different latent ability levels from each other. The categories in English and Swahili are better at discriminating than the categories in math and general knowledge.

Summary Statistics

Table 1 presents sample means of the variables applied in the current paper for each year, together with the means and standard deviations from the pooled dataset. The unit of observation is district-cohort-year cells. The cohort is determined by the year of the survey subtracted the age of the child. Each child is given a population weight to create a representative sample. When averaging over districts, cohorts, and years, the population weights are summed. The summary statistics presented in this section are weighted by these summed population weights multiplied by the square root of

Figure 1: *Category characteristic curves for students in 2011*



Notes: The y-axis refers to the probability of being able to master whatever category in the different tests. The x-axis refers to the estimated latent ability. The figure is based on children in the age group 6–10 years in 2011.
Source: authors' calculations.

the number of children within the district-cohort-year cells.[10] Cohort means in 2011 are presented despite being excluded from the sample, as lagged ability is not available in 2011.

In order to ease interpretation, the dependent variable is standardized to mean zero and a standard deviation of one separately for each year. Next, the standardized ability is averaged for each cohort in each district in each year. The averages of the dependent variable may therefore deviate from zero as cohorts – instead of individuals – are considered, and the sample consists only of cohorts

[10]The summed weights are multiplied by the square root of the number of children to account for greater uncertainty when the average is based on fewer children.

Table 1: *Summary statistics for district-cohorts in the sample period*

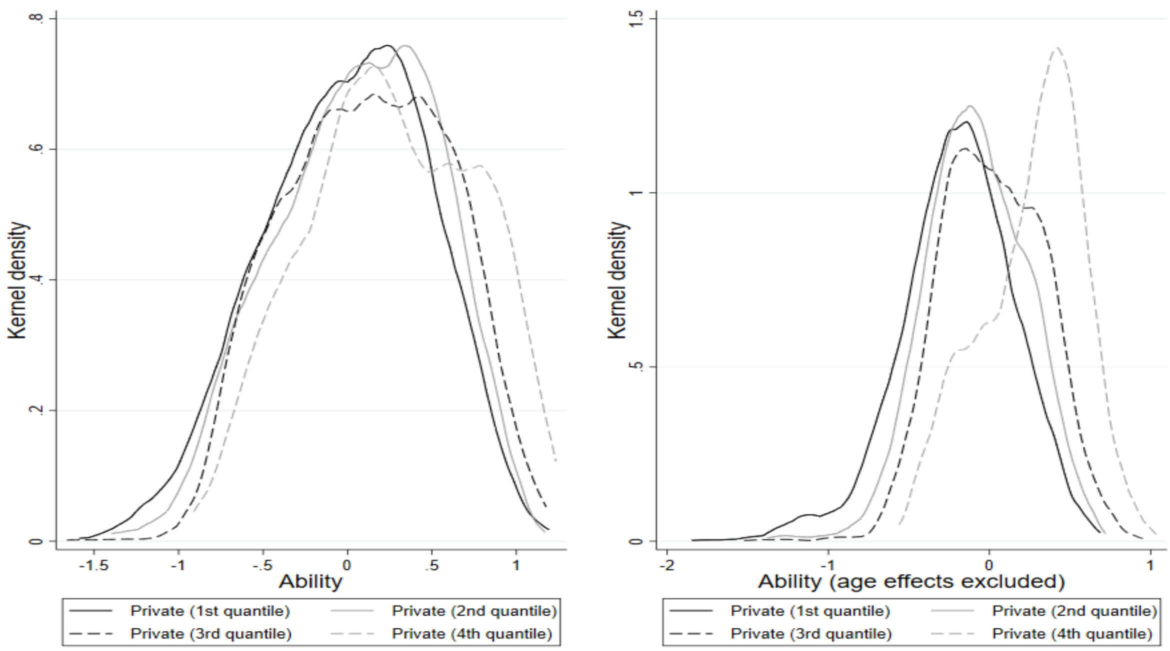
	2011 mean	2012 mean	2013 mean	2014 mean	2015 mean	Pooled mean	Pooled std.
Ability	-0.058	0.124	0.131	0.080	0.109	0.110	0.489
Ability (lagged)	.	-0.202	-0.192	-0.215	-0.225	-0.209	0.521
Private	0.168	0.174	0.195	0.231	0.243	0.213	0.167
Female	0.493	0.494	0.500	0.493	0.508	0.499	0.0.072
Age	8.059	8.559	8.563	8.510	8.527	8.539	1.146
Not in shool	0.130	0.071	0.081	0.104	0.074	0.083	0.094
Household size	6.402	6.264	6.080	6.027	6.031	6.092	0.825
Assets	0.274	0.356	0.441	0.445	0.396	0.412	0.978
Mother secondary school	0.212	0.237	0.224	0.253	0.236	0.238	0.162
Children per district-cohort	110.8	99.3	91.6	91.9	93.0	93.7	29.0
N	615	484	616	615	604	2,319	2,319

Notes: District-cohort groups from 2011 are not part of the pooled dataset as the sample consists only of groups where information on lagged ability is available. *Ability* is our dependent variable, and it is calculated by taking the average yearly standardized latent ability for each district-cohort in each year. *Private* is the share of students enrolled in private schools. *Female* is the share girls. *Age* is the age of children in the district-cohort groups. *Not in school* is the share of children not in school. *Household size* is the average number of household members. *Assets* is the average of an asset index calculated by using a principal component analysis based on various household assets and housing characteristics. *Mother secondary school* is the proportion of children having a mother with some secondary schooling. *Children per district-cohort* is the average number of children in the district-cohort groups.
Source: authors' calculations.

with lagged ability. Figures 2 and 3 further present kernel densities of the dependent variable for the four quantiles of private school enrolment and the simple relationship between the dependent variable and private school enrolment shares, respectively. The graph on the left side of Figure 2 shows raw kernel densities, whereas the graph to the right shows kennel densities after controlling for age effects. Most evident from the right side graph, district-cohorts-year cells with higher private enrolment shares have higher average ability.[11] The same picture emerges from Figure 3. Older cohorts perform better than younger cohorts, and for each age group there is a clear positive

[11]The distributional differences between the left and right side graphs are driven by private school students being younger than public school students.

Figure 2: *Kernel densities of the dependent variable for different subgroups of the sample*



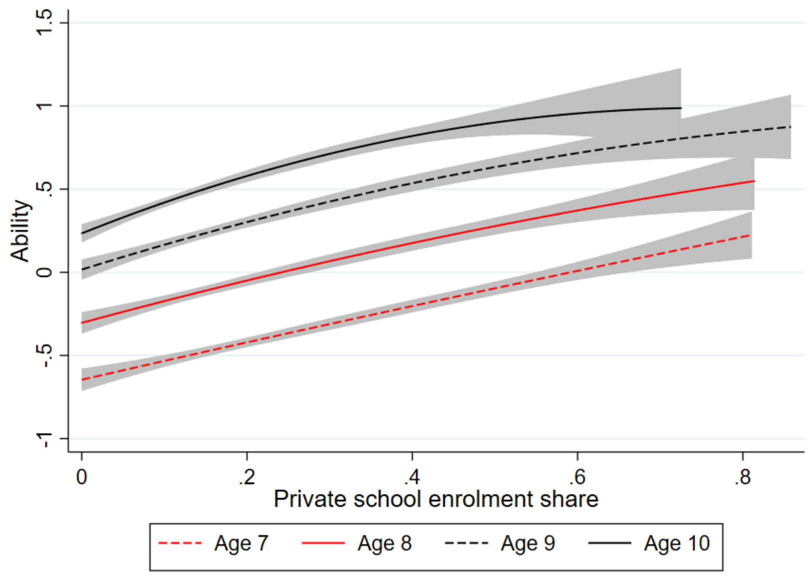
Notes: The left graph presents kernel densities of raw ability, whereas the right graph presents kernel densities of ability where age effects have been neutralized. The kernel densities are split by each quantile of the sample based on private school enrolment shares. The figure is based on average standardized ability for 2,319 district-cohort-year observations.
Source: authors' calculations.

relationship between the dependent variable and the share of children attending private school.

The group averages in 2011 may slightly deviate from the group averages in 2012–2015 as all groups of children are included in 2011, whereas only groups with lagged ability are included in 2012-2015. Specifically, the district-cohorts in 2012–2015 will be older as it is required that they were also available and at least six years old in the previous year. In line with expectations, the share of children that are not in school – predominantly children who have never enrolled – is higher in 2011, and the average age is lower in 2011.

Table 1 further demonstrates that private school enrolment shares have increased over the considered time period from an average of 16.8 percent in 2011 to 24.3 percent in 2015. This overall

Figure 3: *Relationship between the dependent variable and private school enrolment shares*



Notes: The grey areas represent the 95 percent confidence intervals. The figure is based on average standardized ability and private school enrolment shares for 2,319 district-cohort-year observations. Source: authors' calculations.

increase hides considerable within country variation. Of all districts, 15 percent has a negative development in the share of students enrolled in private school and nearly one in four has a positive development of at least ten percentage points. The asset index, *Assets*, is derived from the first component of a principal component analysis calculated separately for each year, and it is based on various household assets and housing characteristics.[12] The approach of calculating the asset index by year is pursued as different assets may change value from year to year. For instance, owning a phone may not tell as much about wealth in 2015 as it would in 2011. Finally, we notice that the share of mothers with some secondary education is relatively stable throughout the considered time period.

[12]The variables used to calculate the asset index are: type of walls, ownership of a TV, radio, phone, bicycle, motorbike, and car, direct access to clean water, access to electricity, and number of meals per day.

IV. METHODOLOGY

A. Baseline specification

A two-step system generalized method of moments (GMM) estimator is used to identify the effects of private schooling on learning (Arellano and Bover 1995; Blundell and Bond 1998). This estimator is an expansion of the difference GMM estimator developed by Holtz-Eakin, Newey, and Rosen (1988) and popularized by Arellano and Bond (1991). The applied methodology leans on the introduction to difference and system GMM by Roodman (2009b). We seek to estimate the model given by Equation 3:

$$y_{c,d,t} = \alpha y_{c,d,t-1} + \beta' \mathbf{X}_{c,d,t} + \mu_{c,d} + v_{c,d,t} \tag{3}$$

where $y_{c,d,t}$ is the average standardized latent ability for cohort c in district d in time period t. The explanatory variables included in **X** are private school enrolment share, share of children not enrolled in school, gender composition, household assets, household size, information on mothers' education, age of children, and time dummies. Finally, we suspect there could be district-cohort fixed effects, which are represented by $\mu_{c,d}$.

The naive approach would be to estimate Equation 3 by OLS. This is problematic as the lagged dependent variable will be correlated with the fixed effect, referred to as dynamic panel bias (Nickell 1981). One may get rid of the fixed effect by employing a fixed effect model. Employing this within-group transformation, however, the transformed lagged dependent variable, together with all predetermined and endogenous explanatory variables, are negatively correlated with the transformed fixed effect.[13] While first differences may be instrumented by longer lags, this is not possible for the fixed effect model as all lags – to some extent – will be correlated with the error term. This is of particular concern with short time panels. In line with Jones and Tarp (2016), one may bias-correct the coefficient estimate associated with the lagged dependent variable.[14] This procedure, however, does not account for the potential endogeneity of the other explanatory variables. Instead, the GMM estimator is especially suited for dynamic models with a large number of observations and a low number of time periods, where some variables are suspected for being endogenous.

The identifying assumption for the system GMM estimator is that the explanatory variables are not correlated with future realizations of the error term, but they are relevant for explaining future realizations of themselves. For instance, start by assuming private school enrolment is endogenous to latent ability. Under this assumption, the proportion of private school students in time period t should be uncorrelated with changes in the error term between time period $t+1$ to $t+2$, and changes in the proportion of private school students between time period t to $t+1$ should be uncorrelated with the fixed effect and shocks to latent ability in time period $t+2$. At the same time, both temporal and spatial variation in the explanatory variable of interest is necessary, though not sufficient, for the internal instruments to be relevant. That is, we must have that the proportion of private school students in time period t is indicative of changes in the proportion of private school students from time period $t+1$ to $t+2$, and changes in the proportion of private school students between time period t to $t+1$ are indicative of the proportion of private school students in time period $t+2$.

[13]This is seen as $y_{c,d,t-1}$ is positively correlated with the $v_{c,d,t-1}$ term in the transformed error term $v^*_{c,d,t} = v_{c,d,t} - (1/(T-1))(v_{c,d,2} + ... + v_{c,d,T})$

[14]The bias-correction works in three steps. First, the fixed effect model is estimated. Second, the coefficient estimate associated with the lagged dependent variable is corrected using the formula $\alpha_{BC} = [1 + T + \alpha_{FE}(T^2 + T)](T^2 - T + 1)^{-1}$. Third, the fixed effect model is re-estimated where the coefficient estimate associated with the lagged dependent variable is constrained at being equal to the bias-corrected estimate.

Before estimating the model, we need to make a few assumptions in regard to the data generating process and take a few decisions on how to estimate the model. This include: 1) classifying explanatory variables as exogenous, predetermined, or endogenous; 2) determining how many lags to use as instruments; 3) deciding whether instruments are separated for each time period or collapsed; and 4) deciding on the level of significance when testing whether the moment conditions are different from zero. Appendix Table A1 provides an overview of the decisions taken in regard to estimating the baseline model.

B. Robustness tests

While the decisions made in regard to the applied estimator reflect what we consider to be the optimal approach to identify a true aggregate effect of private schools on learning, several dimensions could potentially be modified. These include: 1) restricting the sample to district-cohorts present in four consecutive time periods instead of using an unbalanced sample; 2) excluding all endogenous variables except for private school enrolment shares, assuming the instruments are valid even without conditioning on the other explanatory variables; 3) disaggregating the unit of observation from district-cohort to district-cohort-gender; 4) using difference GMM instead of system GMM; and 5) using average test scores instead of IRT.

The first difference transformation may suffer from a smaller sample size if the sample is unbalanced. In the extreme case, a panel dataset can disappear completely in first differences when none of the observations have a lagged dependent variable. As some district-cohorts were not represented each year, the applied sample has gaps in the panel structure, and it is therefore relevant to consider only district-cohorts without gaps. One could also use forward orthogonal deviations instead of first-differences, which in some cases have shown to result in stronger instruments (Hayakawa and Qi 2019).[15]

In the baseline model, the identifying assumption is that the instruments are valid conditional on the other explanatory variables. It is not given, however, that the conditionality is needed. If private school enrolment increases for district-cohort *d-c* from period *t-1* to period *t*, it is straightforward to think that this could be driven by more capable students joining the group in period *t*. That is, in period *t*, by coincidence, the district-cohort *d-c* has both higher unobserved ability and a higher private school enrolment share. Less clear is it to think of a case where the change from period *t-2* to *t-1* is correlated with unobserved ability in period *t*. One case resulting in a correlation is when the error term is autocorrelated.[16]

The preferred unit of observation is district-cohorts, which is in line with previous literature. To increase precision, we divide the district between different cohorts as one cannot move between cohorts, and potential spillover effects are most likely to happen within a cohort rather than between them. One may, however, disaggregate the unit of observation further by changing to district-cohort-gender cells. While the number of observations almost doubles, the disadvantage is that we may have uncaptured spillover effects within the cohort between gender. Empirically, however, Brandt (2018) demonstrates how the correlation between exam scores and academic quality of peers is predominantly driven by peers of a student's own gender for secondary school students in Tanzania. Given this result holds for Kenyan primary school students, the disadvantage of disaggregating to district-cohort-gender cells is modest.

[15]The forward orthogonal transformation subtracts the average of all future observations instead of the lagged observation. Consequently, the orthogonal deviation transformation is less sensitive to gaps in the panel dataset.

[16]We test for second-order autocorrelation to guide the determined lag structure.

Finally, as argued in Section III, IRT holds an advantage over classical test theory as it accounts for difficulty and discrimination parameters, and it generates a cardinal scale. As some might still have preferences for classical test theory due to its easy interpretation, we change the dependent variable from latent ability to actual average test score. Each learning outcome is given one point. For instance, learning to do multiplication rather than only being able to do subtraction has the same impact on the composite test score as learning to read words rather than only being able to read letters.

V. RESULTS

A. Baseline results

Table 2 presents the results from estimating Equation 3 using different estimators. The preferred estimator is the GMM estimator in column (4), whereas columns (1) to (3) demonstrate how the choice of estimator matters a great deal when estimating the impact of private schools. As a rule of thumb, the coefficient estimate on the lagged outcome variable in the GMM model should be between the OLS estimate and the panel fixed effects estimate. While this is the case in 2, the GMM estimate is relatively close to the OLS estimate. One potential explanation when the GMM estimate is larger than the OLS estimate is that one of the control variables contaminates the coefficient estimate associated with the lagged outcome variable in OLS. Inspecting the differences between the OLS model and the GMM model, the variable *Assets* comes out as an outlier. Appendix Table A2 presents the OLS model, the panel fixed effects model, and the bias-corrected panel fixed effects model when constraining the coefficient estimate on the variable *Assets* to be equal to the GMM estimate. In line with expectations, the coefficient estimates on both the lagged outcome variable

Table 2: *The effect of private schooling on average latent ability*

	OLS	Panel FE	Bias-corrected panel FE	GMM
Ability (lagged)	0.280***	-0.179***	0.031	0.236***
	(0.020)	(0.046)	(.)	(0.060)
Private	0.143***	0.267*	0.209	1.051***
	(0.037)	(0.158)	(0.148)	(0.326)
Not in school	-0.778***	-0.727***	-0.770***	-0.962***
	(0.074)	(0.145)	(0.147)	(0.330)
Female	-0.021	0.099	0.075	-0.550
	(0.061)	(0.099)	(0.104)	(0.472)
Age	0.188***	0.337***	0.410***	0.228***
	(0.008)	(0.018)	(0.021)	(0.023)
Household size	-0.079***	-0.057**	-0.064***	-0.248**
	(0.009)	(0.023)	(0.022)	(0.104)
Assets	0.120***	0.097***	0.116***	-0.038
	(0.010)	(0.034)	(0.034)	(0.094)
Mothers secondary school	0.078	0.056	0.076	-0.444
	(0.049)	(0.115)	(0.115)	(0.547)
N	2,319	2,319	2,319	2,319

Notes: The dependent variable is the average standardized latent ability for cohort c in district d at time period t. Each model further includes year dummies and a constant. The sample is weighted by the sum of child population weights within the district-cohort-year multiplied by the square root of the number of children within the group. Standard errors clustered at the district-cohort level are presented in parentheses in columns (1) to (3). Robust Windmeijer-corrected standard errors, clustered at the district-cohort level, are presented in parentheses in column (4). In the GMM model, we have that: 1) the moment conditions for *Mothers secondary school* in the levels equation and the second lag of *Not in school* in the difference equation are excluded, as the p-values for the difference-in-Hansen tests are below 0.1; 2) age and year dummies are exogenous; 3) the lagged dependent variable is predetermined; 4) the remaining variables are endogenous; 5) the instrument count is 31; the joint Hansen test of over-identifying restrictions has a p-value of 0.57; and 6) the test for no second-order autocorrelation has a p-value of 0.64. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Source: authors' calculations.

and private schooling increase in the constrained OLS model compared to the unconstrained model.

A few insights from column (4) are worth highlighting. First, the coefficient estimate on the lagged outcome variable, also known as the persistence parameter, is largely in line with previous studies at the individual level (Glewwe, Ilias, and Kremer 2010; Andrabi et al. 2011). If students had perfect memory and all new learning was built on top of existing knowledge, the persistence parameter would be one. As the parameter is only 0.24, much of the knowledge generated today is forgotten next year. Second, the proportion enrolled in private schools has a strong positive effect on average latent ability. Increasing the share of students enrolled in private schools by 10 percentage points is expected to cause an increase in latent ability of 0.11 standard deviations.[17] This is interpreted as a large effect since the average is for all children in the group, including the children that are not switching to private school. Worth noticing, however, is the large standard errors, suggesting a 95 percent confidence interval between 0.41 and 1.69 for the coefficient estimate. Even the low end of the 95 percent confidence interval is substantially higher than the OLS estimate, potentially caused by contaminating control variables.[18]

A third insight from Table 2 is that the proportion of students not enrolled in school has a strong negative impact on latent ability. This is a comforting result, suggesting there *is* a learning premium from school enrolment. Finally, the absence of an effect from *Assets* and *Mothers secondary school* is puzzling. One would expect an improvement in assets and education of mothers to have a positive impact through more resources for education and better parental support. In particular for *Assets*, however, the effect could also go the opposite direction if the reason for more resources is that parents work more or rely on the children to work as well. Empirical evidence from Tanzania suggests that agricultural-related assets and ownership of livestock have adverse effects on children's educational performance (Kafle, Jolliffe, and Winter-Nelson 2018). In particular for *Mothers secondary school*, the development over time is relatively stable, making the GMM

[17] The standard deviation of private school enrolment is 0.167 (see Table 1), meaning a one standard deviation increase in the private school enrolment share is expected to lead to a 0.18 standard deviation increase in average latent ability.

[18] Constraining both *Assets* and *Mothers secondary school* to the GMM estimates increases the coefficient estimate on private school enrolment shares in the OLS model fourfold.

estimator unsuitable for identifying an effect.

The coefficient estimate associated with age provides a convenient comparison for the coefficient estimate associated with private school enrolment shares. Instead of interpreting the effect of private schools on latent ability, one may compare the effect relative to the age effect. An increase of 10 percentage points in the proportion of children enrolled in private schools is expected to increase average latent ability by what is equivalent to everyone being approximately half of a year older. This large relative effect is a result of both higher private school productivity and low average learning per year relative to other developing countries (Singh 2019).[19]

We present robustness analyses in Appendix Table A3, containing the following: 1) restricting the sample to district-cohorts present in four consecutive time periods; 2) excluding all endogenous variables except for private school enrolment shares; 3) disaggregating the unit of observation to district-cohort-gender cells; 4) using difference GMM instead of system GMM; and 5) using average test scores instead of IRT. While the magnitude of the effect from increasing the private school enrolment share differs among the different models, the effect remains significantly positive and insignificantly different from the baseline result in all analyses. In addition to the presented robustness analyses, one may further investigate the sensitivity towards decreasing the number of instruments, increasing the number of instruments, and using forward orthogonal deviations instead of first differences. These additional tests are all very similar to the baseline results.[20]

[19] Back of the envelope calculation shows that children learn to master approximately one additional ‘competence’ when becoming one year older. For instance, they learn to read a paragraph instead of words in Swahili, while not learning anything in English, math, and general knowledge, or they learn to do subtraction instead of addition only, while not learning anything in Swahili, English, and general knowledge.

[20] Results are available upon request.

VI. Conclusion

Most of the literature examining the effects of private schools in developing countries focuses on individual level estimates. This is a shame as aggregating to a higher order level could be beneficial for studying the impact of private schools as a whole. For instance, spillover effects and peer effects could play a crucial role in determining the desirability of expanding the private school sector. While the aggregated analysis does not distinguish between the different mechanisms, it provides a composite estimate of the effect on learning from increasing or decreasing the private school sector.

The current paper contributes to the literature by estimating a composite learning effect in local markets of private education by accounting for dynamics of learning. The existing literature estimating the learning effect on local markets does not have the opportunity to follow cohorts over a longer time period. Utilizing five rounds of the Uwezo data, we are able to create a pseudo-panel of 6–10 year old children in specific districts. Due to the five different time periods, we can employ a system generalized method of moments (GMM) estimator to account for dynamics of learning and the potential for endogenous explanatory variables.

The baseline results suggest that expanding the private school sector by 10 percentage points in a district-cohort leads to an increase in average latent ability of 0.11 standard deviations. This is considered to be a large effect as the effect is an average for all students in the district-cohort. Importantly, however, the estimated standard errors are large and the 95 percent confidence interval for the effect of a 10 percentage points expansion of the private sector ranges between 0.04 and 0.17 standard deviations. The baseline positive effect of private education is robust to: 1) restricting the sample to district-cohorts present in four consecutive time periods instead of an unbalanced sample; 2) excluding all endogenous variables except private schooling; 3) disaggregating the unit of observation from district-cohort to district-cohort-gender; 4) using difference GMM instead of system GMM; and 5) using average test scores instead of item response theory (IRT).

While the results provide an argument for expanding the private school sector, the baseline estimate is relatively uncertain. Furthermore, the variation in private school shares stem mostly from changes at the bottom. That is, the information utilized in the paper does *not* come from district-cohort with an initially high share of private school students increasing the share even further. To avoid issues of inequality, policy makers could introduce scholarships to low-income children rather than supporting the private school sector in general.

REFERENCES

Alderman, Harold, Peter F. Orazem, and Elizabeth M. Paterno. 2001. “School Quality, School Cost, and the Public/Private School Choices of Low-Income Households in Pakistan.” *The Journal of Human Resources* 36 (2): 304–326.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2008. “A Dime a Day: The Possibilities and Limits of Private Schooling in Pakistan.” *Comparative Education Review* 52 (3): 329–355.

Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc. 2011. “Do Value-Added Estimates Add Value? Accounting for Learning Dynamics.” *American Economic Journal: Applied Economics* 3 (3): 29–54.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters. 2017. “Leveraging Lotteries for School Value-Added: Testing and Estimation.” *The Quarterly Journal of Economics* 132 (2): 871–919.

Angrist, Joshua D., Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. “Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment.” *American Economic Review* 92 (5): 1535–1558.

Arellano, Manuel, and Stephen Bond. 1991. “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.” *The Review of Economic Studies* 58 (2): 277–297.

Arellano, Manuel, and Olympia Bover. 1995. “Another Look at the Instrumental Variable Estimation of Error-Components Models.” *Journal of Econometrics* 68 (1): 29–51.

Blundell, Richard, and Stephen Bond. 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics* 87 (1): 115–143.

Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, and Justin Sandefur. 2013. *The High Return to Low-Cost Private Schooling in a Developing Country.* IGC WORKING PAPER. Oxford, UK: International Growth Centre.

———. 2011. *Why Did Abolishing Fees Not Increase Public School Enrollment in Kenya?* Working Paper 271. Center for Global Development.

Brandt, Kasper. 2018. *Private Beats Public: A Flexible Value-Added Model With Tanzanian School Switchers.* WIDER WORKING PAPER 2018/81.

Deming, David J. 2014. "Using School Choice Lotteries to Test Measures of School Effectiveness." *American Economic Review* 104 (5): 406–411.

Edwards Jr., D. Brent, Steven J. Klees, and Janet L. Wildish. 2017. "Dynamics of Low-Fee Private Schools in Kenya: Governmental Legitimation, Schools-Community Dependence, and Resource Uncertainty." *Teachers College Record* 119 (7): 1–42.

Elks, Phil. 2016. *Lessons Learned from Introducing Value Added Performance Measures in Uganda.* DFID Think Piece. Department for International Development.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. "Teacher Incentives." *American Economic Journal: Applied Economics* 2 (3): 205–227.

Hanushek, Eric A., Steven G. Rivkin, and Lori L. Taylor. 1996. "Aggregation and the Estimated Effects of School Resources." *The Review of Economics and Statistics* 78 (4): 611–627.

Hayakawa, Kazuhiko, and Meng Qi. 2019. "Further Results on the Weak Instruments Problem of the System GMM Estimator in Dynamic Panel Data Models." *Oxford Bulletin of Economics and Statistics.*

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen. 1988. "Estimating Vector Autoregressions with Panel Data." *Econometrica* 56 (6): 1371–1395.

Hoxby, Caroline. 2003. "School Choice and School Competition: Evidence from the United States." *Swedish Economic Policy Review* 10.

Hsieh, Chang-Tai, and Miguel Urquiola. 2006. "The Effects of Generalized School Choice on Achievement and Stratification: Evidence from Chile's Voucher Program." *Journal of Public Economics* 90 (8): 1477–1503.

Jimenez, Emmanuel, Marlaine E. Lockheed, and Vicente Paqueo. 1991. "The Relative Efficiency of Private and Public Schools in Developing Countries." *The World Bank Research Observer* 6 (2): 205–218.

Jones, Sam, and Finn Tarp. 2016. "Does Foreign Aid Harm Political Institutions?" *Journal of Development Economics* 118:266–281.

Kafle, Kashi, Dean Jolliffe, and Alex Winter-Nelson. 2018. "Do Different Types of Assets Have Differential Effects on Child Education? Evidence from Tanzania." *World Development* 109:14–28.

Lassibille, Gérard, Jee-Peng Tan, and Suleman Sumra. 2000. "Expansion of Private Secondary Education: Lessons from Recent Experience in Tanzania." *Comparative Education Review* 44 (1): 1–28.

Lucas, Adrienne M., and Isaac M. Mbiti. 2012. “Access, Sorting, and Achievement: The Short-Run Effects of Free Primary Education in Kenya.” *American Economic Journal: Applied Economics* 4 (4): 226–253.

Masters, Geoff N. 1982. “A Rasch Model for Partial Credit Scoring.” *Psychometrika* 47 (2): 149–174.

Muralidharan, Karthik, and Venkatesh Sundararaman. 2015. “The Aggregate Effect of School Choice: Evidence from a Two-Stage Experiment in India.” *The Quarterly Journal of Economics* 130 (3): 1011–1066.

Ngetich, Solomon Kipyego, Benjamin Kyalo Wambua, and Zachariah Kiptoo Kosgei. 2014. “Determination of Unit Cost Among Secondary Schools in Kenya: A Case of Nandi North District, Kenya.” *European Scientific Journal* 10 (16): 211–224.

Nickell, Stephen. 1981. “Biases in Dynamic Models with Fixed Effects.” *Econometrica* 49 (6): 1417–1426.

Nishimura, Mikiko, and Takashi Yamano. 2013. “Emerging Private Education in Africa: Determinants of School Choice in Rural Kenya.” *World Development* 43:266–275.

Oketch, Moses, Maurice Mutisya, Moses Ngware, and Alex C. Ezeh. 2010. “Why Are There Proportionately More Poor Pupils Enrolled in Non-State Schools in Urban Kenya in Spite of FPE Policy?” *International Journal of Educational Development* 30 (1): 23–32.

Pal, Sarmistha, and Geeta G. Kingdon. 2010. *Can Private School Growth Foster Universal Literacy? Panel Evidence from Indian Districts.* IZA DISCUSSION PAPERS 5274. Institute of Labor Economics (IZA).

Pritchett, Lant. 2013. *The Rebirth of Education: Schooling Ain't Learning.* Brookings Institution Press.

Psacharopoulos, George. 1987. "Public versus Private Schools in Developing Countries: Evidence from Colombia and Tanzania." *International Journal of Educational Development* 7 (1): 59–67.

Ravallion, Martin. 2015. "On the Implications of Essential Heterogeneity for Estimating Causal Impacts Using Social Experiments." *Journal of Econometric Methods* 4 (1): 145–151.

Roodman, David. 2009a. "A Note on the Theme of Too Many Instruments." *Oxford Bulletin of Economics and Statistics* 71 (1): 135–158.

———. 2009b. "How to Do Xtabond2: An Introduction to Difference and System GMM in Stata:" *The Stata Journal.*

Sacerdote, Bruce. 2011. "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In *Handbook of the Economics of Education,* edited by Eric A. Hanushek, Stephen Machin, and Ludger Woesmann, 3:249–277. Amsterdam: Elsevier.

Schirmer, Stefan. 2010. *Hidden Assets: South Africa's Low-Fee Private Schools.* CDE IN DEPTH 10. Johannesburg: The Centre for Development and Enterprise.

Singh, Abhijeet. 2019. "Learning More with Every Year: School Year Productivity and International Learning Divergence." *Journal of the European Economic Association.*

———. 2015. "Private School Effects in Urban and Rural India: Panel Estimates at Primary and Secondary School Ages." *Journal of Development Economics* 113:16–32.

Tooley, James, Yong Bao, Pauline Dixon, and John Merrifield. 2011. "School Choice and Academic Performance: Some Evidence From Developing Countries." *Journal of School Choice* 5 (1): 1–39.

UNESCO. 2020. *Education Statistics.* http://data.uis.unesco.org. Accessed January 25, 2020.

Urquiola, M. 2016. "Chapter 4 - Competition Among Schools: Traditional Public and Private Schools." In *Handbook of the Economics of Education,* edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 5:209–237. Amsterdam: Elsevier.

van der Linden, Wim J. 2016. *Handbook of Item Response Theory.* First. Chapman and Hall.

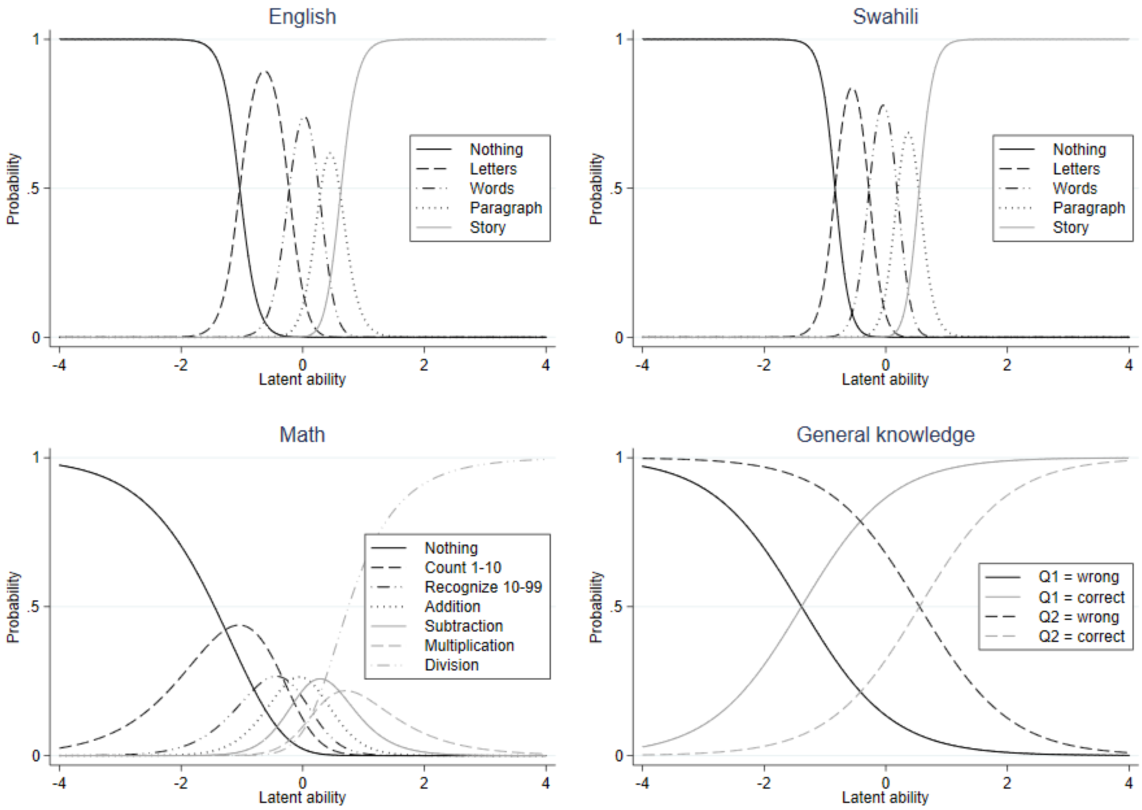
Vandenberghe, V., and S. Robin. 2004. "Evaluating the Effectiveness of Private Education across Countries: A Comparison of Methods." *Labour Economics,* European Association of Labour Economists 15th Annual Conference, Universidad Pablo de Olavide, Seville, 18-21 September 2003, 11 (4): 487–506.

Wamalwa, Fredrick M., and Justine Burns. 2018. "Private Schools and Student Learning Achievements in Kenya." *Economics of Education Review* 66:114–124.

Windmeijer, Frank. 2005. "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators." *Journal of Econometrics* 126 (1): 25–51.

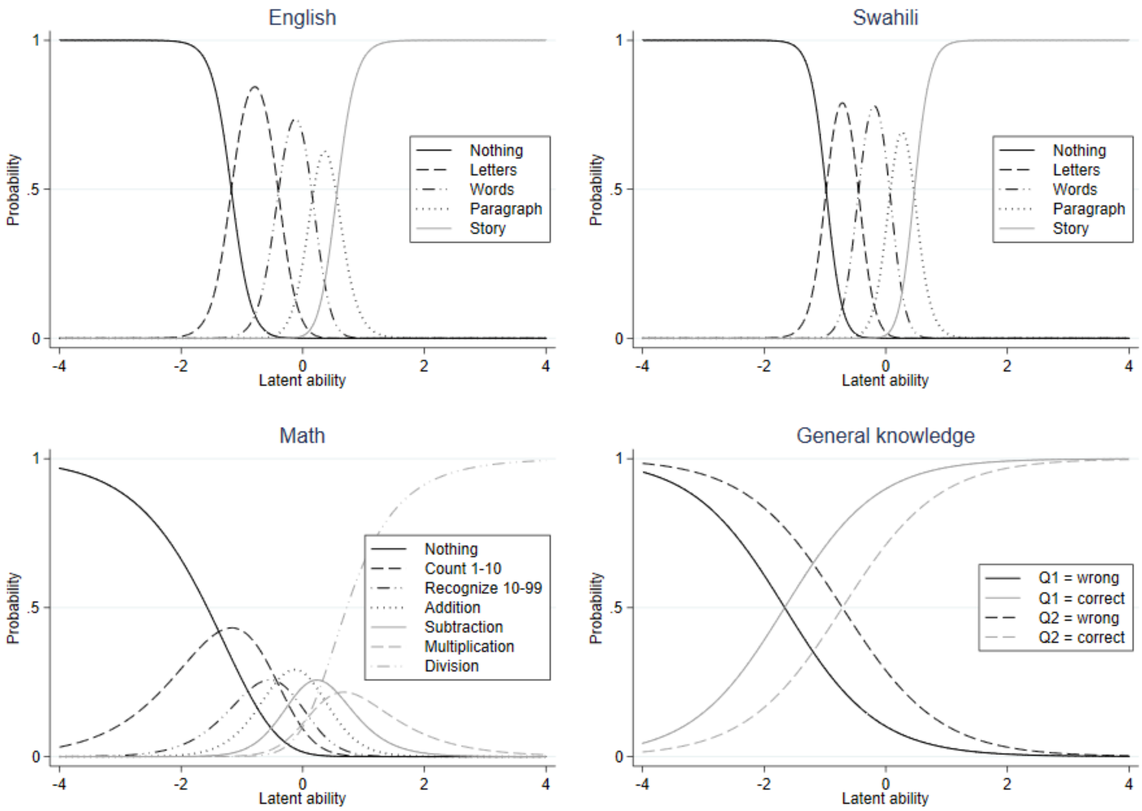
APPENDIX

Figure A1: *Category characteristic curves for students in 2012*



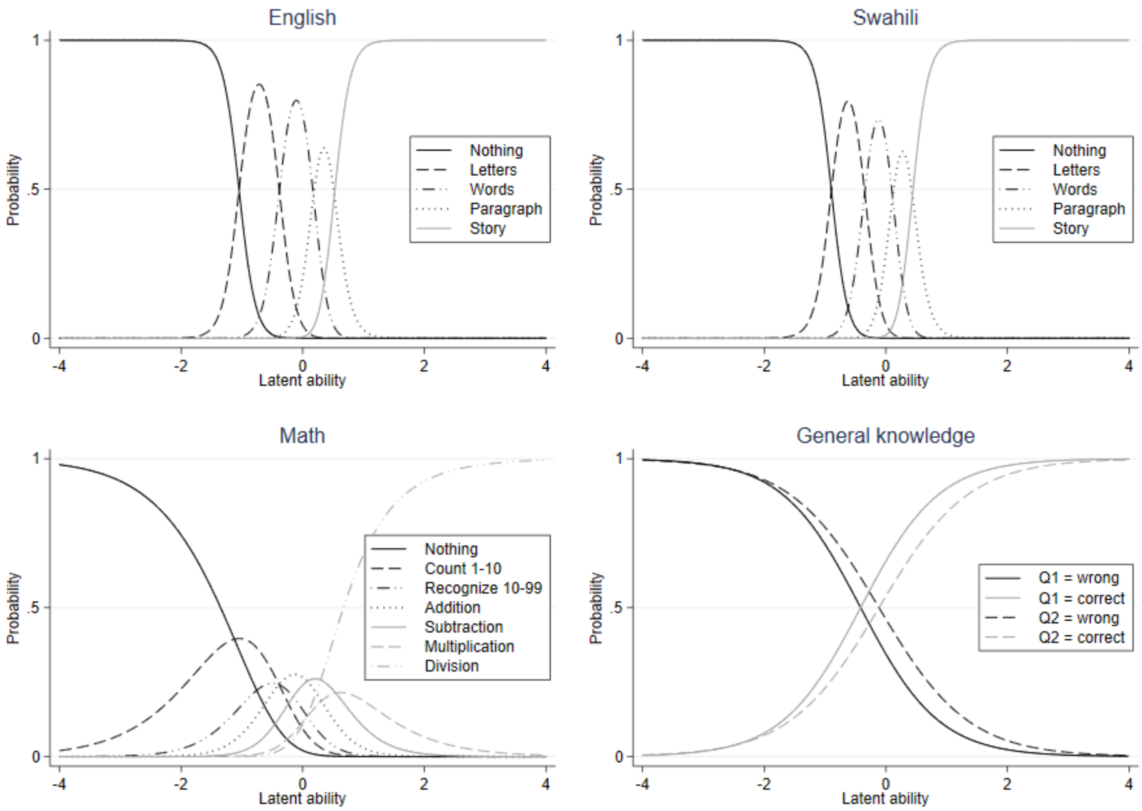
Notes: The y-axis refers to the probability of being able to master whatever category in the different tests. The x-axis refers to the estimated latent ability. The figure is based on children in the age group 6–10 years in 2012.
Source: authors' calculations.

Figure A2: *Category characteristic curves for students in 2013*



Notes: The y-axis refers to the probability of being able to master whatever category in the different tests. The x-axis refers to the estimated latent ability. The figure is based on children in the age group 6–10 years in 2013.
Source: authors' calculations.

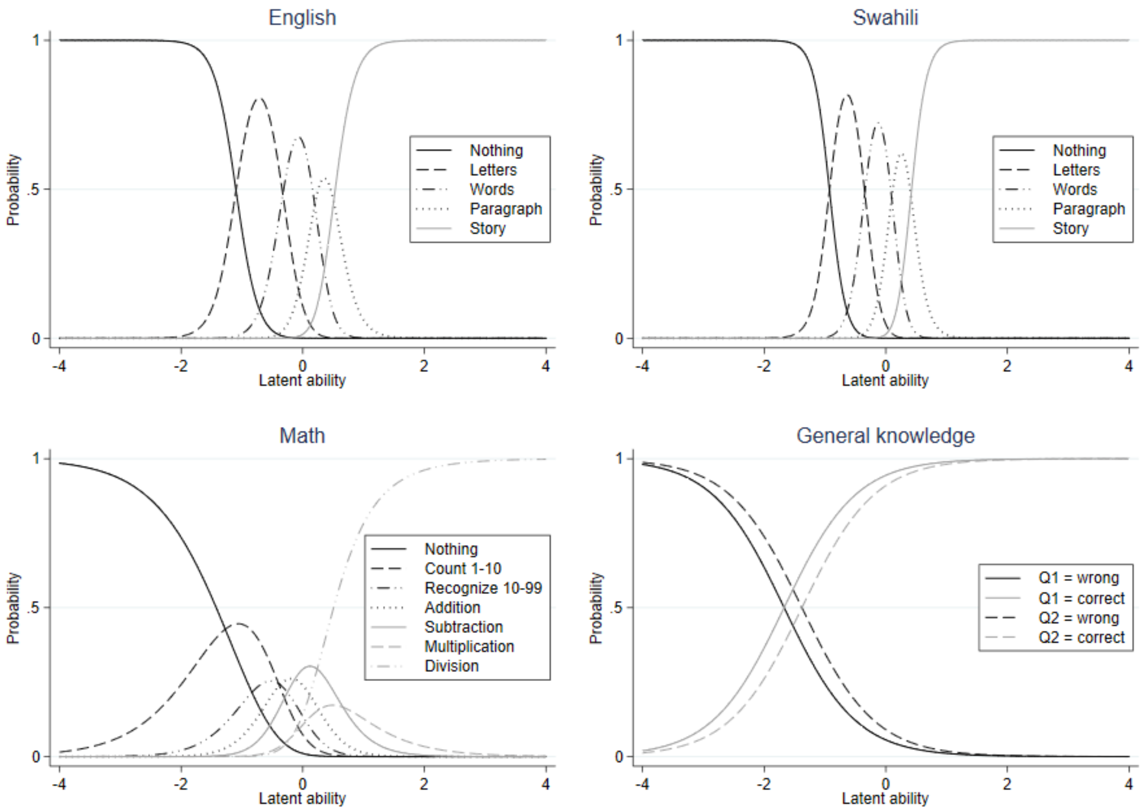
Figure A3: *Category characteristic curves for students in 2014*



Notes: The y-axis refers to the probability of being able to master whatever category in the different tests. The x-axis refers to the estimated latent ability. The figure is based on children in the age group 6–10 years in 2014.

Source: authors' calculations.

Figure A4: *Category characteristic curves for students in 2015*



Notes: The y-axis refers to the probability of being able to master whatever category in the different tests. The x-axis refers to the estimated latent ability. The figure is based on children in the age group 6–10 years in 2015.

Source: authors' calculations.

Table A1: *Assumptions and decisions in regard to the applied methodology*

	Procedure	Approach further explained
(1)	Determine whether the explanatory variables are exogenous, predetermined, or endogenous.	The lagged dependent variable is by definition predetermined under no second-order autocorrelation, age and year dummies are exogenous, and the remaining variables could be affected by current shocks to ability, and hence, they are endogenous.
(2)	Determine the lag structure.	All available lags are used as instruments.
(3)	Decide whether instruments are separated for each time period, or collapsed.	Following Roodman (2009a), instruments are collapsed.
(4)	Determine whether moment conditions from both the difference and levels equations should be used.	To avoid mathematically redundant instruments, time dummies are only included in the levels equation. Splitting the instrument set by each variable and the difference/levels equation, we evaluate the difference-in-Hansen tests of exogeneity. If an instrument subset has a p-value below 0.1, the moment conditions are excluded. We start by the subset with the lowest p-value and re-estimate the model until none of the instrument subsets' joint validity can be rejected. If moment conditions to the difference equation are excluded, we follow Roodman (2009b) and include deeper lags to the levels equation.
(5)	Decide whether to use the one-step or two-step estimator	The two-step estimator is chosen.
(6)	Decide how to calculate standard errors	Standard errors clustered by district-cohort and robust to finite samples are used (Windmeijer 2005).

Source: authors' own.

Table A2: *The effect of private schooling on average latent ability from constrained models*

	OLS	Panel FE	Bias-corrected Panel FE
Ability (lagged)	0.400***	-0.199***	0.001
	(0.020)	(0.045)	(.)
Private	0.385***	0.429***	0.395***
	(0.040)	(0.148)	(0.142)
Never enrolled	-0.997***	-0.808***	-0.861***
	(0.083)	(0.142)	(0.145)
Female	0.031	0.103	0.081
	(0.068)	(0.101)	(0.106)
Age	0.149***	0.428***	0.427***
	(0.008)	(0.021)	(0.021)
Household size	-0.099***	-0.057***	-0.063***
	(0.010)	(0.022)	(0.021)
Assets	-0.038	-0.038	-0.038
	(.)	(.)	(.)
Mothers secondary school	0.395***	0.155	0.187
	(0.046)	(0.116)	(0.116)
N	2,319	2,319	2,319

Notes: The dependent variable is the average standardized latent ability for cohort c in district d at time period t. Each model further includes year dummies. The sample is weighted by the sum of child population weights within the district-cohort-year multiplied by the square root of the number of children within the group. Models are constrained to have the same coefficient estimate for *Assets* as in the GMM model from Table 2. Standard errors clustered at the cohort-district level are presented in parentheses. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Source: authors' calculations.

Table A3: *Robustness analyses*

	No panel gaps	Exclude endogenous variables	District-cohort-gender cells	Difference GMM	Average test scores
Ability (lagged)	0.219***	0.331***	0.160***	0.140	0.288***
	(0.056)	(0.073)	(0.042)	(0.095)	(0.063)
Private	0.965**	1.957***	1.269***	1.867***	0.963***
	(0.403)	(0.626)	(0.399)	(0.666)	(0.353)
Not in school	-1.149***		-0.922***	-0.821**	-1.377***
	(0.443)		(0.247)	(0.328)	(0.334)
Female	-0.247			-0.669	-0.762
	(0.443)			(0.497)	(0.491)
Age	0.231***	0.241***	0.259***	0.229***	0.200***
	(0.031)	(0.038)	(0.017)	(0.036)	(0.025)
Household size	-0.215**		-0.136	-0.079	-0.220**
	(0.097)		(0.090)	(0.102)	(0.102)
Assets	-0.104		0.092	-0.079	-0.079
	(0.103)		(0.100)	(0.158)	(0.094)
Mothers secondary school	0.045		-0.461	-0.265	-0.174
	(0.413)		(0.433)	(0.592)	(0.542)
Instrument count	31	13	27	23	31
Joint Hansen test of over-identifying restrictions (p-value)	0.361	0.485	0.502	0.740	0.569
Test for no second-order autocorrelation (p-value)	0.428	0.322	0.801	0.373	0.574
N	1,355	2,319	4,495	1,262	2,319

Notes: In columns (1)–(4), the dependent variable is the average standardized latent ability for cohort c in district d at time period t. In column (5), the dependent variable is the average standardized test score, and *Ability (lagged)* is also based on this measure. Each model further includes year dummies and a constant. The sample is weighted by the sum of child population weights within the district-cohort-year multiplied by the square root of the number of children within the group. Robust Windmeijer-corrected standard errors, clustered at the district-cohort level, are presented in parentheses. In all columns, the moment conditions for *Mothers secondary school* in the levels equation and the second lag of *Not in school* in the difference equation are excluded, as the p-values for the difference-in-Hansen tests are below 0.1. Age and year dummies are exogenous, the lagged dependent variable is predetermined, and the remaining variables are endogenous. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
Source: authors' calculations.

Chapter 3

The Impacts of Eliminating Secondary School Fees: Evidence from Tanzania

The Impacts of Eliminating Secondary School Fees: Evidence from Tanzania*

Kasper Brandt[1] and Beatrice K. Mkenda[2]

[1]*Department of Economics, University of Copenhagen*

[2]*Department of Economics, University of Dar es Salaam*

Tanzania implemented a fee-free secondary school reform in January 2016. Using variation in district and cohort exposure to the reform, we employ a difference-in-differences strategy to estimate the short-term impacts of the reform. The reform substantially increased enrolment into secondary education. While these enrolment effects were predominantly driven by an increase in public school enrolment, there was also a delayed positive effect on private school enrolment. Districts mostly affected by the reform experienced a significant drop in exam scores relative to less affected districts, which cannot be explained by academic abilities of new students. These findings are in line with a theoretical model on school choice, where some individuals are credit-constrained and the quality of public education is harmed by increased enrolment. (*JEL* I21, I24, I28)

*We wish to thank Youdi Schipper and Sam Jones for providing excellent comments on the paper. Further thanks go to the National Examinations Council of Tanzania for providing public access to a large and useful database on student exam records.

I. Introduction

In 2016, Tanzania implemented a nationwide policy eliminating secondary school fees. Instead, the government pays a compensation fee and an additional capitation grant to cover recurrent expenses except salaries. This is a considerable step towards reaching the Sustainable Development Goal (SDG) indicator 4.1 calling for all children completing free quality primary and secondary education. Similar nationwide reforms are seen in recent years in other countries in Sub-Saharan Africa.[1] Due to the recency of these reforms, the evaluated impacts are limited. At the primary school level, research demonstrates that eliminating fees have a huge impact on enrolment, usually at the cost of worsening quality of education.[2] Whether the same effects take place at the secondary school level is of utmost importance for education finance and allocation of resources. To uphold the same level of quality, schools need more teachers and classrooms. The magnitude of these required investments depends crucially on the local enrolment response.

The existing evidence on the impacts of fee-free secondary education tend to rely on randomized control trials (RCTs) or policies targeting specific subgroups of children (Khandker, Pitt, and Fuwa 2003; Barrera-Osorio et al. 2011; Baird, McIntosh, and Özler 2011; Borkum 2012; Garlick 2013; Hermida 2014; Gajigo 2016; Blimpo, Gajigo, and Pugatch 2019; Duflo, Dupas, and Kremer 2019). These studies find positive enrolment effects, but the magnitude appears to be both context-specific

[1] These countries include Rwanda (2007 and 2012), Uganda (2007), Burundi (2012), Namibia (2016), Kenya (2018), Malawi (2018), Sierra Leone (2018), and Ghana (2019). Other countries have eliminated school fees for targeted groups.

[2] A positive effect on enrolment is found in Ethiopia (Omoeva and Moussa 2018), Kenya (Lucas and Mbiti 2012), Malawi (Al-Samarrai and Zaman 2007), Tanzania (Hoogeveen and Rossi 2013), and Uganda (Deininger 2003; Nishimura, Yamano, and Sasaoka 2008; Grogan 2009). Worsening effects are found on: 1) test scores for students unaffected by the reform in Kenya (Lucas and Mbiti 2012); 2) grade achievement in Tanzania (Hoogeveen and Rossi 2013); 3) student-teacher ratios in Uganda and Ethiopia (Deininger 2003; Omoeva and Moussa 2018); and 4) teacher quality in Malawi (Omoeva and Moussa 2018).

and heterogeneous across treatment groups. In addition, they do not provide a consistent answer on potential spillover effects to other students. Two studies evaluate nationwide policies on eliminating or reducing secondary school fees in Uganda and Kenya, respectively (Masuda and Yamauchi 2018; Brudevold-Newman 2019). Both studies employ a difference-in-differences strategy, exploiting variation in district and cohort exposure to the reforms. Masuda and Yamauchi (2018) find that the elimination of public secondary school fees led to a substantial increase in students taking the final secondary school exam. Moreover, test scores remained unharmed, potentially attributed to the public-private partnership scheme implemented simultaneously to comply with increased demand. Brudevold-Newman (2019) demonstrates that a fee-reduction reform in Kenya increased progression to secondary education, delayed age of first intercourse, first marriage, and first birth, and shifted employment from agriculture to skilled work.

The present paper shows that a fee-free secondary school reform in Tanzania had a substantial impact on progression to secondary school. We use a difference-in-differences strategy to identify the impacts of the fee-free reform. Districts with low pre-reform progression rates from primary to secondary education were more exposed to the reform, and only cohorts who took the the Form Two National Assessment (FTNA) exams after 2014 received any treatment. We construct a measure of treatment intensity based on pre-reform district-level progression rates between primary school and attendance at the FTNA exams. We demonstrate that districts with low pre-reform progression rates did not experience significantly different enrolment trends prior to the reform. The positive enrolment effect for exposed cohorts is robust to controlling for district-specific pre-trends, allowing post-trends to deviate from pre-trends dependent on treatment intensity, and controlling for potential determinants of the pre-reform differences in progression rates. The enrolment effect is predominantly driven by public schools. Two years after the implementation of the reform, however, high-intensity districts also experienced an increase in progression to private schools relative to low-intensity districts, which is in line with a theoretical model assuming public school quality declines following a surge in enrolment.

The paper further demonstrates that the improvements in enrolment came at a cost of learning unattributable to academic abilities of new students. Between 2015 and 2019, FTNA exam scores dropped by an estimated 0.10 to 0.15 standard deviation for students in the district at the 80th percentile, in terms of treatment intensity, relative to students in the district at the 20th percentile. Students induced by the reform to progress performed significantly better in primary school compared to students expected to progress even without the reform. This finding is in line with a theoretical framework, where a large fraction of new students are credit-constrained without fee-free education. High-intensity districts further experienced an increase in the share of students failing relative to low-intensity districts. The estimated increase in the number of students failing due to the reform, however, is substantially lower than the estimated increase in enrolment due to the reform.

The current paper provides a theoretical framework for studying the impacts of eliminating school fees (Section II), describes the Tanzanian education system and the applied data (Section III), describes the strategy for identifying the impacts of the secondary school fee-free reform on enrolment, changes in public-private market structure, and learning (Section IV), presents the results (Section V), discusses the implications of the results (Section VI), and concludes (Section VII).

II. Theoretical framework

The current section provides a theoretical framework for schooling and innate ability of students. The framework leans on Lochner and Monge-Naranjo (2011), but further adds a public-private school distinction and a quality of education dimension. A baseline model assumes individuals always enrol in school when the return is higher than working. Next, credit constraints for some individuals are added to the model. In a third model, quality of public education is allowed to be affected by a public school price reduction. While a public school price reduction leads to higher total enrolment, the separate enrolment effects on the public and private sector depend on the imposed assumptions. Average innate ability of students becomes ambiguous when allowing some individuals to be credit-constrained. The impact on the relative composition of public and private school students also becomes ambiguous when including quality effects from a price reduction.

Basic model

The theoretical framework follows a basic two-period model:

$$U = u(c_0) + \delta u(c_1) \tag{1}$$

where individuals receive utility from consumption based on a concave utility function, $u(\cdot)$. A time discount factor is represented by δ. An individual has three options to choose from: 1) enrol in public school in period 0 and get a medium-wage job in period 1; 2) enrol in private school in period 0 and get a high-wage job in period 1; or 3) work in both periods for a low wage normalized to one. The individual gets a high-wage job after private education due to differences in quality of education. The return to education is *not* additively separable, meaning high-ability individuals are assumed to have a higher return to quality of education.[3] The individual obtains the following

[3] For instance, Glewwe, Kremer, and Moulin (2009) demonstrate that providing free text books in Kenya only had a learning impact on high-ability students.

utility from each option:

$$
\begin{aligned}
U_{pub} &= \delta u(h(a, q_{pub}) - (1+r)P_{pub}) \\
U_{pri} &= \delta u(h(a, q_{pri}) - (1+r)P_{pri}) \\
U_w &= u(1) + \delta u(1)
\end{aligned}
\tag{2}
$$

where $h(a, q)$ is the return to education conditional on student innate ability, a, and quality of education, q. The parameter r refers to the interest rate of borrowing for education, and P is the price of attending public or private school. At specific levels of a, q_{pub}, and P_{pub}, individuals are indifferent between working and attending public education. Likewise, at specific levels of a, q_{pub}, q_{pri}, P_{pub}, and P_{pri}, the individual is indifferent between public and private education. Assuming quality and prices are independent of innate ability, that is students have the same options, low-ability students work in both periods, medium-ability students opt enter public education, and high-ability students opt for private education.

Reducing the price on public education

This subsection analyses the effects on enrolment and innate ability of individuals in school when reducing the price on public education. Three different models are analysed: 1) a baseline model without credit constraints and no quality effects caused by increased enrolment; 2) a model allowing individuals to be credit-constrained; and 3) a model allowing for both credit-constrained individuals and public education quality effects caused by increased enrolment.

A public school price reduction induces both low- and high-ability students to enrol in public school when no one is credit-constrained and there are no public school quality effects from

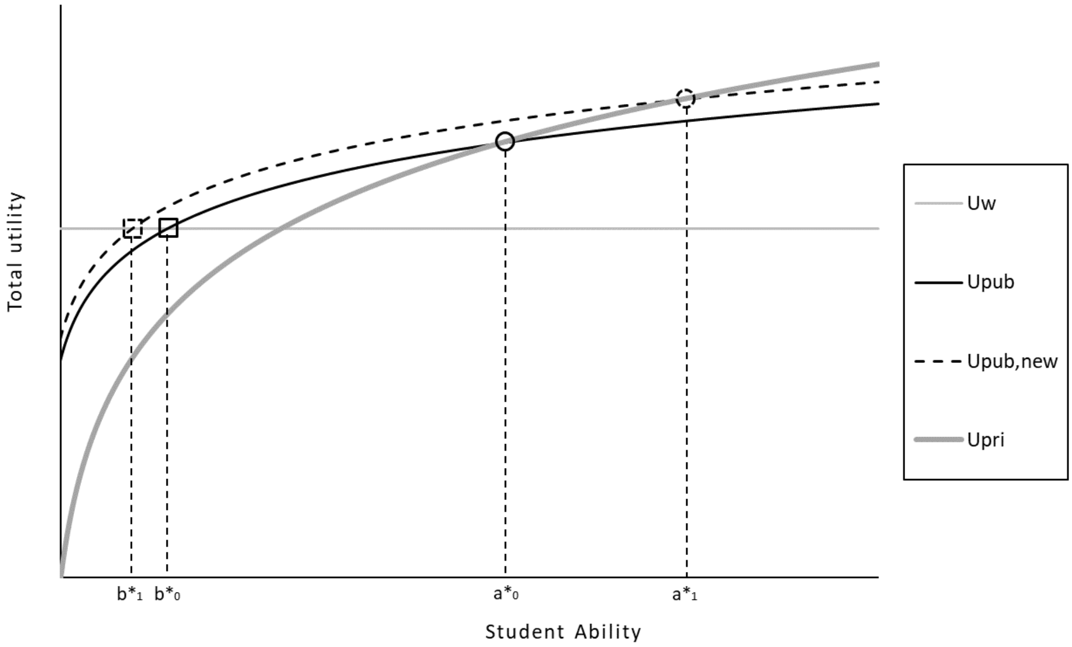


Figure 1: *Baseline model with no credit constraints and no quality effects*

Notes: The model assumes no one is credit-constrained and there are no quality effects caused by higher enrolment in public schools. *Uw* represents utility from working in both time periods. *Upub* and *Upub,new* represent utility from enrolling in public school before and after reducing the price on public education, respectively. *Upri* represents utility from enrolling in private school. The solid square and circle represent the ability cut-off points, where individuals are indifferent between working or public education and public education or private education, respectively, before reducing the price on public education. The dashed square and circle represent equivalent ability cut-off points after reducing the price on public education.

increased enrolment. Figure 1 illustrates what happens to the public education utility curve when reducing the price. The cut-off point for choosing public education instead of working declines from b_0^* to b_1^*. Thus, more low-ability individuals enter public education. At the other end of the ability spectrum, the cut-off point for choosing public education instead of private education increases from a_0^* to a_1^*. Thus, also more high-ability students enter public education. The baseline model predicts that total enrolment increases, public school enrolment increases, private school enrolment decreases, average innate ability for all students decreases, the change in average innate ability for public school students is ambiguous, and average innate ability for private school students increases.

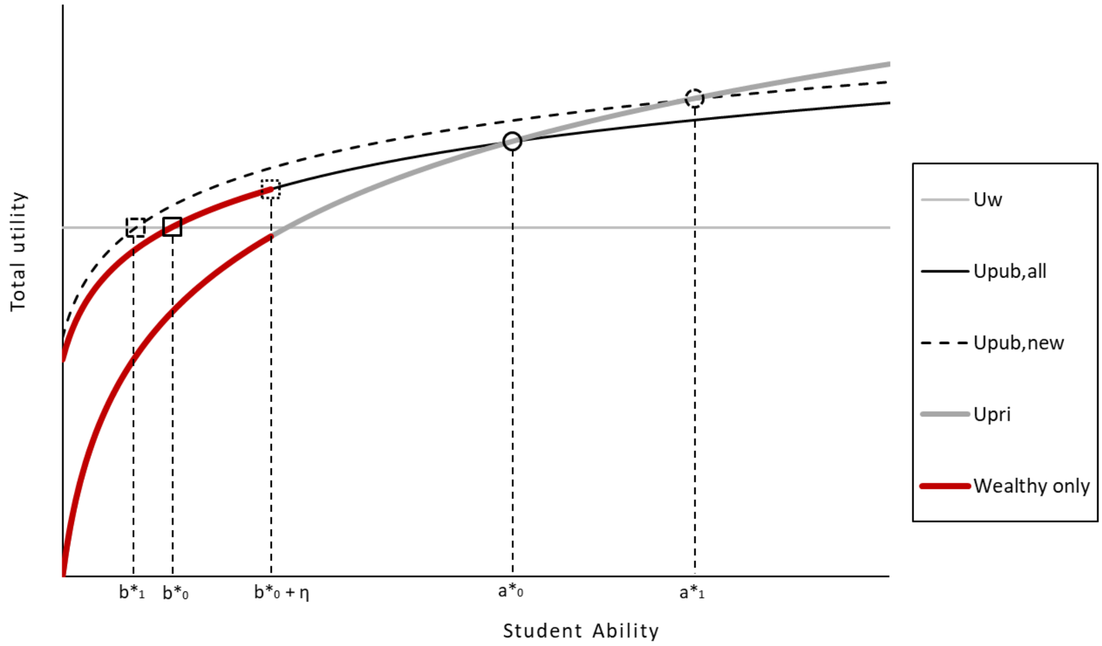


Figure 2: *Credit-constraints for poor and low-ability individuals*

Notes: The model assumes low-ability individuals can be either non-constrained ("wealthy") or credit-constrained ("poor"), whereas high-ability individuals cannot be credit-constrained, and there are no quality effects caused by higher enrolment in public schools. *Uw* represents utility from working in both time periods. *Upub,all* and *Upub,new* represent utility from enrolling in public school before and after a public education price reduction, respectively. *Upri* represents utility from enrolling in private school. *Wealthy only* represents utility from enrolling in public or private school for non-constrained individuals only. The solid square and circle represent the ability cut-off points, where individuals are indifferent between working or public education and public education or private education, respectively, before reducing the price on public education. The dashed square and circle represent equivalent ability cut-off points after reducing the price on public education. The dotted square represents the ability cut-off point between working and public education for credit-constrained individuals prior to a public school price reduction.

When some individuals are credit-constrained, the expected implications of a public education price reduction change. Figure 2 illustrates the implications of reducing the price on public education under the assumption that non-constrained individuals are represented for entire ability spectrum and credit-constrained individuals are low-ability only.[4] Individuals right above a_0^* and non-constrained

[4] The assumption of credit-constrained individuals being low-ability individuals can be justified by a financial market lending money to high-ability individuals only. In Appendix B, the model is analysed under the assumption that individuals are randomly credit-constrained.

individuals between b_0^* and b_1^* change preference to public education. In addition, credit-constrained individuals between $b_0^* + \eta$ and b_1^* enrol in public education, as the new public education utility curve is unlimited for everyone. The effect on average innate ability for individuals in public schools depends on the distribution of credit-constrained and non-constrained individuals.[5] While the effect on average ability in public schools is ambiguous, average ability for individuals in private schools increases as individuals just above a_0^* enrol in public education.

Relaxing the assumption of no public school quality effects following a price reduction further changes the predictions on the composition of public and private school students. Instead of no public school quality effects, we assume quality of public school is harmed due to increased enrolment.[6] If there was no change in enrolment for low-ability individuals when reducing the price on public education, the new ability cut-off point for individuals being indifferent between public and private education would lie in between a_0^* and a_1^* in Figure 1. Hence, quality effects act as a moderating factor on enrolment. As more low-ability and credit-constrained individuals also enrol in public education, the quality of public education might drop enough to make the highest ability public school students switch to private education. Figure 3 illustrates this case.[7] The only certain prediction from this model is that total enrolment increases. The impacts on the relative composition of public and private school students and the average innate ability of individuals in school are ambiguous.

[5] Average innate ability in public schools increases when relatively few individuals have an ability level between b_1^* and b_0^*, there is a large number of credit-constrained individuals just below $b_0^* + \eta$, and there are few individuals in school with ability level above $b_0^* + \eta$. The probability of seeing an increase in average ability is higher when assuming individuals are randomly credit-constrained (see Appendix B)

[6] One could also incorporate quality effects to private education. To keep things simple, however, this is not pursued.

[7] Appendix Figure A1 illustrates a case in which the quality of public education is not severely harmed.

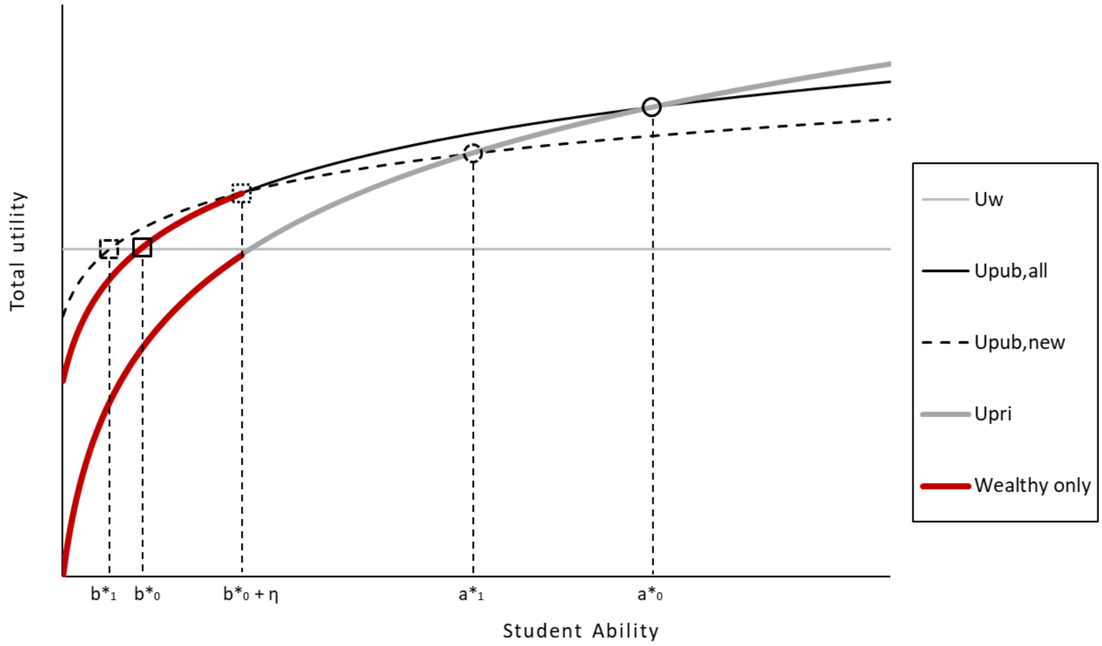


Figure 3: *Credit-constraints and public education quality effects*

Notes: The model assumes low-ability individuals can be either non-constrained ("wealthy") or credit-constrained ("poor"), whereas high-ability individuals cannot be credit-constrained, and there *are* quality effects caused by higher enrolment in public schools. *Uw* represents utility from working in both time periods. *Upub,all* and *Upub,new* represent utility from enrolling in public school before and after a public education price reduction, respectively. *Upri* represents utility from enrolling in private school. *Wealthy only* represents utility from enrolling in public or private school for non-constrained individuals only. The solid square and circle represent the ability cut-off points, where individuals are indifferent between working or public education and public education or private education, respectively, before reducing the price on public education. The dashed square and circle represent equivalent ability cut-off points after reducing the price on public education. The dotted square represents the ability cut-off point between working and public education for credit-constrained individuals prior to a public school price reduction.

III. Context and data

Tanzanian education system

The Tanzanian education system consists of seven years of primary education, followed by two years of lower secondary education, and two years of upper secondary education. After four years of secondary education, students can enrol in vocational or technical educations, or they can proceed with two years of advanced secondary education and university education thereafter. The school year follows the Gregorian calendar and the Primary School Leaving Examination (PSLE) and the Form Two National Assessment (FTNA) exams are taken from medio to ultimo November. The PSLE cannot be retaken and students who failed the PSLE cannot enrol in public secondary education. To gain access to Form 3 in secondary education, students must pass the FTNA. At the PSLE and FTNA, students are tested in five and eight mandatory subjects, respectively. At the FTNA, students may further be tested in elective subjects.

After 14 years of fee-free primary education, the Government of Tanzania announced in February 2015 a new policy plan to make four years of secondary education fee-free. At the time, the fee was 20,000 Tanzanian shillings per student per year.[8] This fee covered costs of operating the school, except staff salaries. During the 2015 election campaign, Dr. John Magufuli (now President John Magufuli) promised to implement the policy plan if elected. After his election as president, the government implemented the policy in November 2015, taking effect from January 2016. The fees previously paid by students are now paid by the government directly to the bank accounts of schools.[9] Other expenses still paid by students include school uniforms, transport, meals, and writing materials. In anticipation of increased enrolment, the government further raised funding for staff. While the key component of the reform was fee-free secondary education, an additional component included voluntary language of instruction.[10] Except for the subject Kiswahili, exams are still conducted in English.

[8] On January 1 2016, 20,000 Tanzanian shillings had a value of USD 9.2, equivalent to approximately 2 percent of median household consumption per adult equivalent household member.

[9] A qualitative study of secondary schools across Tanzania finds that all visited schools indicate they received the monthly payment (HakiElimu 2017). A challenge for schools with increasing enrolment, however, is that the payment is based on enrolment in the previous year.

[10] Previously, the official language of instruction was English. Conversations with school principals and researchers in the field, however, suggest many schools were already teaching in Swahili.

Relative to countries in the region, a high proportion of Tanzanian children persist until the final grade in primary school and an increasing share of students make the transition to secondary education. In 2017, 92 percent of children in a cohort persisted until the final grade in primary school, which is on a par with Kenya and substantially higher than Rwanda, Burundi, and Uganda. (UNESCO 2020). Before 2016, however, the progression rate to secondary education was low in Tanzania relative to countries in the region. Only 56 percent of students finishing primary education progressed to secondary education in 2012 (UNESCO 2020).[11] After the implementation of the secondary school fee-free reform in 2016, the progression rate rose substantially. By 2017, the progression rate had reached 71 percent. Compared to neighbouring countries, Tanzania exceeded the rate in Uganda and came close to the rates in Burundi and Rwanda (UNESCO 2020).

Data sources

The empirical analysis relies on student-level exam records at two different levels of study and aggregation. Table 1 provides an overview of the applied sources of data. First, student-level exam records from the PSLE in each year between 2013 and 2017. While individual exam records from the PSLE are not available in 2011 and 2012, district-level records are available. Second, student-level exam records from the FTNA in each year between 2014 and 2019. Moreover, FTNA enrolment numbers for students in Form 2 in 2013 are used to predict the number of students who took the FTNA exams the same year.[12] This information results in district-level information

[11]The high persistence rate until the end of primary education and the low progression rate to secondary education were arguably legacies of Julius K. Nyerere, the first President of Tanzania, who desired a system of equality and practicality (Nyerere 1967).

[12]District-level dropout rates in Form 2 are calculated using years where both enrolment numbers and number of students who took the FTNA exams are available. Next, enrolment numbers in Form 2 in 2013 are adjusted for the district-level dropout rates to obtain the predicted number of students who took the FTNA exams.

Table 1: *Data availability for individual exam records, district averages, and years of exams*

Level of aggregation	Level of study	Exam years applied
Individual exam records	PSLE	2013–2017
	FTNA	2015–2019
District-level exam takers	PSLE	2011–2017
	FTNA	2013–2019

on progression between the PSLE and FTNA for students who took the FTNA exams in 2013 to 2019. As students spend at least two years in secondary school before taking the FTNA exams, a value-added model can only be used for students who took the FTNA exams in 2015 to 2019. The exams are assessed by a centralized group of examiners.[13]

In order to create a panel at the student level, we match FTNA takers with PSLE takers two years before based on the exact and full name of the student. Students with a name duplicate in either the PSLE or the FTNA are excluded (2.0 percent of PSLE students and 1.5 percent of FTNA students). Of all FTNA students without a name duplicate, 66.9 percent are uniquely identified in the PSLE records two years before. As a placebo test, matching records based on a one year interval between PSLE and FTNA should result in near to zero matches.[14] As expected, only 1.5 percent of FTNA students are matched with a PSLE record one year before. In total, 1,537,231 FTNA students between 2015 and 2019 are uniquely matched with their PSLE record two years before.

Two additional sources of data employed are the President's Office, Regional Administration and Local Government (PO-RALG 2020), and the Visible Infrared Imaging Radiometer Suite (VIIRS)

[13]The National Examinations Council of Tanzania (NECTA) administers both exams. The exam documents are always either locked up or guarded by police and NECTA employees. At the day of the exam, appointed invigilators ensure students follow the exam regulations.

[14]Despite having no name duplicate within one year, a student may have a name duplicate over time. This could lead some of the FTNA students to be matched with a PSLE student the year before.

(Earth Observation Group, NOAA 2020). The former contains information on school ownership, which is used to separate enrolment figures into public and private schools. The latter data source contains night-time radiance (light intensity), which is used as a proxy for economic activity.[15]

Descriptive statistics

Table 2 presents the variables applied and the sample means. As the empirical strategy is split in a district-level analysis and an individual-level analysis, the first seven rows (panel a) are district-level variables used in the analysis of enrolment. The last two variables (panel b) are individual grade point averages (GPAs) at the PSLE and the FTNA used in the analysis of learning. The columns present the sample means for the entire sample, the pre-reform cohort who took the FTNA exams in November 2014, and the most recent cohort who took the FTNA exams in November 2019. Due to changes in administrative boundaries over time and geographical precision of data sources, some districts are merged together to ensure they cover the same geographical area over time.[16]

The first four rows of Table 2 show that district-year cells, on average, have a progression rate of 56 percent between the PSLE and the FTNA, which increases over time.[17] The number of PSLE students is lagged two years in order to compare the same cohort. The numbers of FTNA students

[15] A monthly panel is created for each 750×750 meters cell and cell-specific trends are estimated. Outlier observations more than three standard deviations from the trend are excluded. Next, district-level averages are calculated for each month. As data is available from April 2012 to April 2019, monthly averages in January to March 2012 are based on the district-level yearly growth rate in the subsequent year and light intensities in January to March 2013. Monthly district-level light intensities for May to December 2019 are predicted in a similar manner. Finally, yearly radiance averages are calculated to fit the panel structure of the empirical models.

[16] Appendix Table A1 presents the list of districts used for the analysis and how they have been merged.

[17] The applied measure for progression is not the official measure for progression between primary and secondary education. The official measure is the number of new students enrolling in Form 1 in secondary school relative to the number of students in the final grade of primary school one year before. Hence, the applied measure may understate the official measure due to students dropping out before taking the FTNA.

Table 2: *Descriptive statistics*

	Full sample	Cohort 2014	Cohort 2019
Panel (a): district-year cells			
PSLE students (2 years lag)	6,526	6,658	6,999
FTNA students	3,616	3,295	4,466
Progression rate	0.562	0.496	0.641
Treatment intensity	0.504	0.504	0.504
FTNA fail share	0.102	0.100	0.095
Secondary schools	35.6	34.9	36.9
Night-time radiance	0.304	0.207	0.407
N	910	130	130
Panel (b): individual cells			
PSLE scores	2.26		2.31
FTNA scores	1.37	1.52	1.28
N		See notes	

Notes: Cohort 2014 and Cohort 2019 refer to the cohorts who took the FTNA exams in 2014 and 2019, respectively. The number of observations for *FTNA fail share* in the full sample is 780 as information is not available in 2013. *PSLE students* refers to the number of students who took the PSLE between 2011 and 2017, whereas *FTNA students* refers to the number of students who took the FTNA exams between 2013 and 2019. *Progression rate* refers to the district-level progression rate between the PSLE and the FTNA two years later, and *Treatment intensity* is the progression rate for the cohort who took the FTNA exams in 2014. *FTNA fail share* is the share of FTNA students who failed the exams. *Secondary schools* is the number of schools, and *Night-time radiance* is the average light intensity. *PSLE scores* and *FTNA scores* are average individual scores. In the empirical analysis, they are standardized within each year. Individual PSLE records are only available for the 2015 cohort and onwards. In the full sample, there are 2,745,810 students with an FTNA exam record, of which 1,538,721 have a matched PSLE exam record. In 2014 and 2019, there are 411,091 and 570,730 students with an FTNA exam record, respectively, and 525,157 students have a matched PSLE exam record in 2019.
Source: National Examination Council of Tanzania, Earth Observation Group, NOAA (2020), and authors' calculations.

do not include students who repeat Form 2. We create a measure of treatment intensity based on the pre-reform progression rate for students who took the PSLE and FTNA exams at the end of 2012 and 2014, respectively. We use the share of students who left the school system as the intensity of treatment (1 – progression rate). Figure 4 illustrates the progression rate over time for the four quantiles in terms of treatment intensity. After the announcement of the plan to eliminate fees, progression rates kept declining as was the case prior to the announcement. Between 2015 and 2018, however, the progression rates increased substantially, in particular for high-intensity districts.

While Figure 4 suggests a relative improvement for the districts with the lowest pre-reform progression rates, the overall impact on educational inequality is clearer in Figure 5. Figure 5 illustrates the yearly Gini coefficients in regard to the number of secondary school students in a district relative to the number of primary school students. Right after the announcement of the reform, the coefficient slightly increased. The reason for this is that there is a general drop in the progression rate between 2014 and 2015. While the drop is less significant in terms of percentage points for high-intensity districts, low-intensity districts take up a larger share of secondary school students in 2015 due to the relative drop being marginally larger in high-intensity districts. After the implementation of the reform in 2016, the Gini coefficient drops substantially and it lowers off four years after the implementation (five years after announcement).

Row five, six, and seven of Table 2 present the means for share of FTNA students failing, number of secondary schools, and night-time radiance, respectively. The average share of FTNA students failing has remained constant at around 10 percent. While exam difficulty may have changed, it is an interesting observation that there is no sudden increase after the implementation of the fee-free reform. The number of secondary schools and night-time radiance are potentially affecting both pre-reform levels and post-reform enrolment development. We notice that the average number of

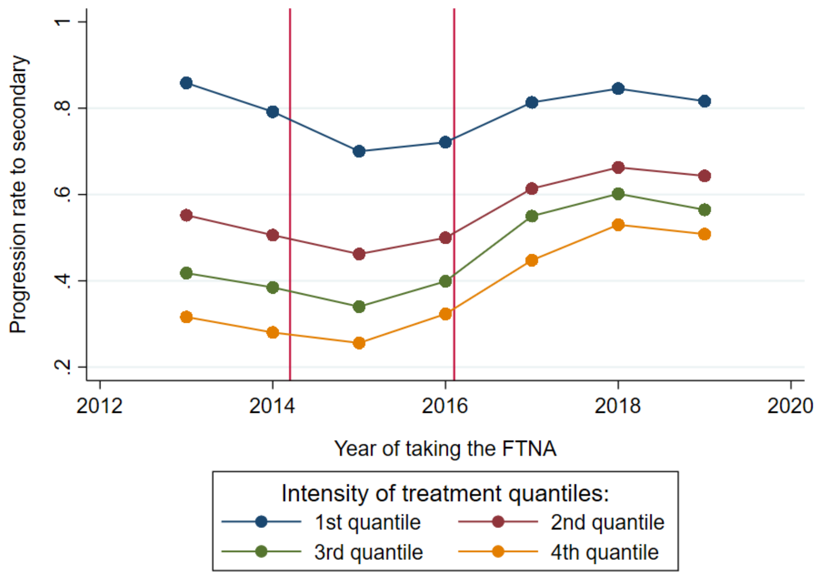


Figure 4: *District-level progression rate over time for different quantiles of treatment intensity*

Notes: The y-axis refers to the number of FTNA students in year t (not repeating Form 2) divided by the number of PSLE students two years before. The first quantile contains the districts least affected by the reform. Red lines represent the time of reform announcement and implementation.
Source: National Examination Council of Tanzania, and authors' calculations.

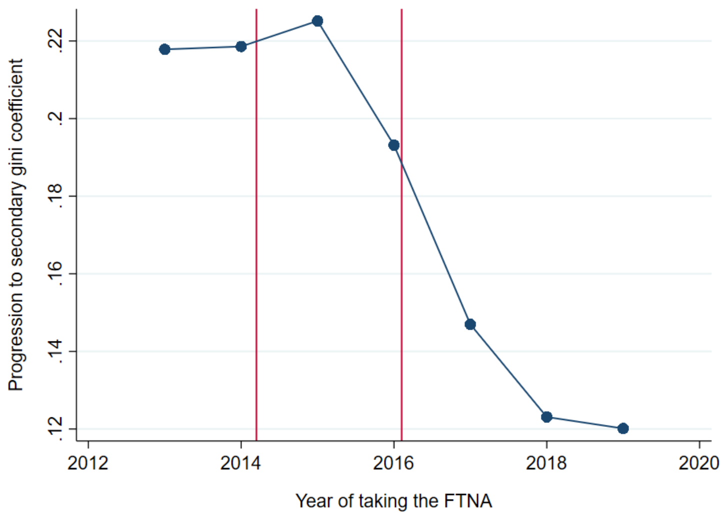


Figure 5: *Gini coefficients for primary to secondary school enrolment*

Notes: The Gini coefficient is the sum of distances between cumulative district percentages of PSLE students and cumulative district percentages of FTNA students divided by the sum of cumulative district percentages of PSLE students. Red lines represent the time of reform announcement and implementation.
Source: National Examination Council of Tanzania, and authors' calculations.

schools per district has increased by two schools between 2014 and 2019, and the average night-time radiance has doubled. The doubling of night-time radiance happens from 2016 to 2017 and appears to be driven by a general measurement change in the raw satellite data. Supporting this hypothesis is that average radiance increased by 0.23 in 2017 independent of lagged radiance level. In the empirical analyses, cohort fixed effects capture such general measurement changes.

The last two rows report the sample means of the individual GPAs at the PSLE and the FTNA. While PSLE scores of FTNA takers are only available for the cohorts who took the FTNA exams in 2015 and onwards, the development is positive between the 2015 cohort and the 2019 cohort. On the other hand, FTNA scores have worsened over the sample period. In the empirical analysis, the exam scores are standardized to have mean zero and standard deviation one within each year.

IV. Identification strategy

The current paper presents two different analyses. The first analysis examines the impact of the fee-free reform on progression from primary to secondary education at the district level. This includes both the overall impact on progression, and divided by public and private education. The second analysis examines learning, where the outcome variables are secondary school exam scores at the individual level and the number of students failing the FTNA exams at the district level.

Impact on enrolment

The identification strategy follows a difference-in-differences approach, relying on district and cohort variation in exposure to the reform.[18] District variation stems from different pre-reform

[18]Choosing a unit of observation that is too small, migration between areas and spillover effects becomes an identification problem. Choosing a unit that is too high, treatment within the unit of observation might vary

progression rates. Districts with an initially high progression rate have less potential to benefit from the reform. Cohort variation stems from the timing of the reform. Students who took the FTNA exams in 2013 and 2014 did not receive any treatment. The 2015 FTNA cohort received an information treatment in the beginning of the second year of secondary education, as the policy plans were announced in February 2015. The 2016 FTNA cohort received the same information in the first year of secondary education and further experienced fee-free secondary education in the second year of secondary education. Lastly, the 2017, 2018, and 2019 FTNA cohorts received full fee-free secondary education and they could also take this into account when they applied for secondary education.

We observe each district-cohort group only once. The empirical specification for evaluating the enrolment effect is:

$$\Delta y_{d,t} = \beta_0 + \sum_{t=2014}^{2019} \beta_{t-2013}(I_d * \delta_t) + \delta_t + \alpha X_{d,t} + \varepsilon_{d,t}, \tag{3}$$

where $y_{d,t}$ is the progression rate from primary to secondary school for district d and FTNA cohort t, δ_t represents cohort fixed effects, I_d is the intensity of treatment for district d, and $X_{d,t}$ includes first differences of the log of secondary schools, contemporary and lagged night-time radiance, and district-specific pre-trends. The beta coefficients measure the impact on enrolment from the treatment intensity variable for the different cohorts. The β_1 measures the impact on enrolment for the 2014 FTNA cohort, which is included to test if high- and low-intensity districts were on a similar path prior to the reform. The rest of the beta coefficients are included to analyse the dynamics of the enrolment effect. The analysis on public-private market structure follows the same empirical

substantially and harm precision (Jepsen 2002). Adhering to related literature, we choose the district as the unit of observation.

strategy, except substituting the outcome variable to be changes in the share of students progressing to public secondary schools and share of students progressing to private secondary schools.

Impact on learning

We change the unit of observation to the individual level when studying the impact of the fee-free reform on exam scores. Standard errors are clustered at the district level, as this is the level of treatment. The analysis contains three different models explaining different dimensions of learning. First, we assume new students that progress due to the reform are similar to the students that would progress without the reform. While any effect of treatment intensity could be due to declining school quality in highly affected districts, it could also be that the composition of students changes, thereby violating the imposed assumption. The first empirical specification for evaluating learning effects is:

$$y_{i,d,t} = \beta_0 + \sum_{t=2015}^{2019} \beta_{t-2014}(I_d * \delta_t) + \delta_t + \gamma_d + \alpha X_{d,t} + \varepsilon_{i,d,t}, \tag{4}$$

where $y_{i,d,t}$ represents FTNA exam scores for student i in district d and cohort t, and γ_d represents district fixed effects. We standardize exam scores to have a mean of zero and a standard deviation of one within each year.

Second, we move to a value-added model identical to Equation 4, but further controlling for PSLE scores. Before estimating the value-added model, we examine whether high-intensity districts experienced a different development in lagged exam scores (PSLE scores) of their secondary school students after the reform compared to low-intensity districts. One can imagine that the reform induced academically weaker students to progress if their return to secondary education was low and the drop in costs made it profitable to progress. On the other hand, some individuals might have been credit-constrained, thereby making progression to secondary education impossible. As emphasized in Section II, the impact of a public school price reduction on innate ability of individuals in school is ambiguous.

Third, the unit of observation reverts to the district level and we examine the share of students failing the FTNA exams. We use the coefficient estimates from this model to predict the number of students failing due to the reform for each district. This estimated number of students failing due to the reform is compared to the estimated number of individuals enrolling due to the reform. In the extreme case where all new students are not learning at all, the estimated number of students failing due to the reform is similar to the estimated number of new students. At the other extreme, where new students learn the same as others and the entrance of new students do not affect others, around 10 percent of new students fail. As difficulty of exams might vary over time, the explanatory variables of interest are year dummies interacted with the measure of treatment intensity. The empirical specification for evaluating failing effects is:

$$y_{d,t} = \beta_0 + \sum_{t=2015}^{2019} \beta_{t-2014}(I_d * \delta_t) + \delta_t + \gamma_d + \alpha X_{d,t} + \varepsilon_{d,t}, \tag{5}$$

where $y_{d,t}$ is the share of students failing the FTNA exams in district d and cohort t.

The difference-in-differences approach assumes that without treatment, post-treatment trends would be the same as pre-treatment trends. We examine the sensitivity of the results to this assumption in a robustness analysis, where post-trends are allowed to deviate from pre-trends dependent on treatment intensity. The issue is further mitigated by controlling for two likely determinants of differences in pre-treatment levels of the outcome variable.

V. RESULTS

Impact on enrolment

Table 3 presents the coefficient estimates from estimating Equation 3 from Section IV. Three different progression rates act as the dependent variable: 1) the overall progression rate to all schools; 2) the progression rate to public schools only; and 3) the progression to private schools only. As the public-private school distinction is only available from 2014 and onwards, columns (2) and (3) do not include a coefficient estimate on the interaction between cohort 2014 and treatment intensity. A full table of all coefficient estimates is reported in Appendix Table A2. Appendix Figures A2 and A3 further illustrate the coefficient estimates associated with the treatment intensity times cohort interaction terms from the first difference model and a district fixed effect model, respectively.

We notice that the impact of treatment intensity depends heavily on the cohort. As expected since the 2013 and 2014 cohorts were not treated, the coefficient estimate associated with treatment intensity for cohort 2014 is insignificant (p-value = 0.24). The 2015 cohort received an 'information treatment', as they were told, with almost one year until the exams, that the government planned to make four years of secondary education fee-free. For this cohort, high-intensity districts experienced an increase in the progression rate relative to low-intensity districts.[19] This result is entirely driven by an increase in the progression rate to public schools. The 2016 cohort received the fee elimination

[19]Instead of interpreting the coefficient as a zero to one treatment, one may reasonably scale down the coefficient to an effect of being in the 80th percentile in terms of treatment intensity relative to being in the 20th percentile. The 80th and 20th percentile districts have treatment intensities of 0.67 and 0.30, respectively. Hence, the 80th percentile district is expected to have a $9.6 \times (0.67 - 0.30) = 3.6$ percentage points higher increase in the progression rate compared to the 20th percentile district between the 2014 and 2015 cohorts.

Table 3: *Yearly impact of the fee-free reform on district-level progression rates*

	Δ Progression rate (all schools)	Δ Progression rate (to public schools)	Δ Progression rate (to private schools)
Treatment intensity × *Cohort*2014	0.048		
	(0.041)		
Treatment intensity × *Cohort*2015	0.096^{***}	0.095^{***}	0.008
	(0.023)	(0.019)	(0.011)
Treatment intensity × *Cohort*2016	0.075^{***}	0.080^{***}	0.004
	(0.024)	(0.019)	(0.011)
Treatment intensity × *Cohort*2017	0.074^{**}	0.059^{**}	0.017^{**}
	(0.029)	(0.026)	(0.008)
Treatment intensity × *Cohort*2018	0.026	0.014	0.032^{**}
	(0.036)	(0.031)	(0.016)
Treatment intensity × *Cohort*2019	0.013	-0.051^{*}	0.070^{***}
	(0.031)	(0.030)	(0.010)
Cohort FE	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes
N	780	650	650
R^2	0.525	0.532	0.266

Notes: Heteroscedasticity-consistent standard errors in parentheses. Cohorts 2014 to 2019 refer to the cohorts who took the FTNA exams in 2014 to 2019, respectively. Additional controls include district-specific pre-trends, first differences of the log of secondary schools, first differences of the log of contemporary night-time radiance, and first differences of the log of one and two years lagged night-time radiance. Appendix Table A2 reports all coefficient estimates. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Source: National Examination Council of Tanzania, Earth Observation Group, NOAA (2020), and authors' calculations.

treatment in the second year. As expected, high-intensity districts increase the overall progression rate even more in 2016 relative to low-intensity districts. Again, this effect is entirely driven by the progression rate to public school.

The 2017, 2018, and 2019 cohorts were all fully exposed to the reform, as students were able to take into account the elimination of fees when they applied for secondary education. The overall progression rate increased more in high-intensity districts than in low-intensity districts between the 2016 and 2017 cohorts. While impacts in the first two years with partial treatment were driven entirely by public schools, high-intensity districts in 2017 also experienced an increase in the progression rate to private schools relative to low-intensity districts. As described in the theoretical framework in Section II, this result could be driven by declining quality of public schools. For the 2018 and 2019 cohorts, high-intensity districts did not experience statistically significant changes in the overall progression rate relative to low-intensity districts. In 2019, however, this null effect masks two diverse effects on progression rates to public and private schools. The progression rate to public schools declined for high-intensity districts relative to low-intensity districts between the 2018 and 2019 cohorts. At the same time, however, high-intensity districts saw a relative increase in the progression rate to private schools.

In economic terms, these results are substantial. Appendix Figure A3 illustrates the coefficient estimates on the interactions between treatment intensity and year of taking the FTNA in a district fixed effect model. The district at the 80th percentile, in terms of treatment intensity, is expected to have a 12 percentage points higher increase in the progression rate compared to the 20th percentile district between 2014 and 2019.[20] As pre-reform progression rates for the 20th and 80th percentile districts are 0.67 and 0.30, respectively, a 12 percentage points narrowing of the gap is a considerable change.

The additional controls demonstrate that an increase in the number of secondary schools is positively correlated with an increase in the overall progression rate. This relationship is entirely

[20]This result is derived by taking the gap in treatment intensity between the 20th and 80th percentile districts $(0.67 - 0.30)$ and multiplying it with the difference between the 2019 and 2014 impact of treatment intensity $(0.38 - 0.06)$

driven by the progression rate to private schools, suggesting the newly established schools are predominantly privately operated. Only changes in the two-years lagged radiance is significantly correlated with changes in the progression rate. This is in line with expectations, as students decide whether to enter secondary education two years before the FTNA. The relationship is negative, suggesting more economic activity at the time of deciding whether to progress to secondary education is negatively correlated with progression. In the theoretical framework, this corresponds to an improvement of the outside option of working in both time periods.

The underlying assumption for the difference-in-differences approach is that post-trends follow pre-trends. We investigate the sensitivity to this assumption by allowing post-trends to deviate from pre-trends dependent on the measure of treatment intensity. Specifically, we define district-specific post-trends as:

$$Post\text{-}trend_{d,t} = Pre\text{-}trend_{d,t} + \theta \times I_d \times (t - 2014), \tag{6}$$

where I_d is the district-specific treatment intensity, and θ is a measure of how much larger the post-trends in high-intensity districts should be relative to low-intensity districts. We let $\theta = 0.1$, which corresponds to a considerable increase in the post-treatment trend for high-intensity districts relative to low-intensity districts. Prior to the reform announcement, pre-trends for districts below and above the median of treatment intensity were, on average, -7.1 and -5.3 percentage points per year, respectively. Setting $\theta = 0.1$, however, post-trends for districts below and above the median of treatment intensity change to -3.7 and 1.4 percentage points per year, respectively. That is, we allow the expected development in progression rates to be larger for high-intensity districts. Appendix Figure A4 compares the baseline coefficient estimates to coefficient estimates where $\theta = 0.1$. The coefficient estimates remain significantly different from zero at the 5 percent significance level when setting $\theta = 0.1$.[21]

Impact on learning

The previous subsection demonstrated a substantial impact of the fee-free secondary school reform on progression rates between primary and secondary education. The present subsection investigates whether this improvement came at a cost of learning. We study both FTNA exam scores at the individual level and the share of students who failed the FTNA at the district level. The latter is used to predict the number of students who failed due to the reform and compare this to the expected number of students induced by the reform to progress to secondary education.

FTNA exam scores

Table 4 presents the results from estimating Equation 4 from Section IV. This is done both with and without controlling for PSLE scores, as including this variable excludes the 2014 FTNA cohort and a fraction of secondary school students where PSLE records could not be matched. Columns (1), (2), and (3) present the full sample model without controlling for PSLE scores, the partial sample model without controlling for PSLE scores, and the partial sample model controlling for PSLE scores, respectively. In order to compare coefficient estimates between models, the base cohort is 2015 for all models. The table further shows the association between treatment intensity and PSLE scores in column (4). A full table of all coefficient estimates is reported in Appendix Table A3.

Columns (1) and (2) demonstrates that high-intensity districts experienced a drop in FTNA

[21] Setting $\theta = 0.37$ makes all interaction terms insignificant at the 5 percent significance level. This level of θ, however, seems highly unrealistic, and we believe setting $\theta = 0$ is most reasonable.

Table 4: *Impact of the fee-free reform on individual exam scores*

	FTNA GPA	FTNA GPA	FTNA GPA	PSLE GPA
Treatment intensity $\times$ *Cohort*2014	0.081			
	(0.079)			
Treatment intensity $\times$ *Cohort*2016	-0.142^{***}	-0.135^{***}	-0.245^{***}	0.191^{***}
	(0.049)	(0.049)	(0.061)	(0.057)
Treatment intensity $\times$ *Cohort*2017	-0.291^{***}	-0.249^{***}	-0.442^{***}	0.335^{***}
	(0.064)	(0.060)	(0.087)	(0.083)
Treatment intensity $\times$ *Cohort*2018	-0.364^{***}	-0.300^{***}	-0.403^{***}	0.179^{*}
	(0.063)	(0.066)	(0.109)	(0.107)
Treatment intensity $\times$ *Cohort*2019	-0.386^{***}	-0.272^{***}	-0.389^{***}	0.203^{*}
	(0.064)	(0.059)	(0.099)	(0.104)
PSLE GPA			0.575^{***}	
			(0.016)	
District and cohort FE	Yes	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes	Yes
N	2,745,748	1,537,231	1,537,231	1,537,231
R^2	0.042	0.048	0.367	0.051

Notes: Standard errors clustered at the district level in parentheses. Cohorts 2014 to 2019 refer to the cohorts who took the FTNA exams in 2014 to 2019, respectively. The base cohort is 2015 in all models. FTNA and PSLE GPAs are standardized within each cohort. Additional controls include log of secondary schools, log of contemporary night-time radiance, and log of one and two years lagged night-time radiance. Appendix Table A3 reports all coefficient estimates. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Source: National Examination Council of Tanzania, Earth Observation Group, NOAA (2020), and authors' calculations.

exam performance after the implementation of the fee-free reform. Treatment intensity did not have a significant correlation with FTNA scores in 2014 relative to 2015. This is in spite of the results in Table 3 showing that the reform had an impact on progression rates already for the 2015 cohort. Between the 2015 and 2016 cohorts, students in high-intensity districts experienced a significant drop in performance relative to their counterparts in low-intensity districts. This effect worsened even more between 2016 and 2018. The magnitudes are slightly smaller when moving to a sample of students where FTNA students are matched with their PSLE records. These students all completed the first two years of secondary education on time and their names were reported the exact same way at the PSLE and the FTNA. This means the types of students in columns (1) and (2) may deviate and create slightly different results.

While columns (1) and (2) shows a negative development in exam scores for students in high-intensity districts, the composition of students may also have changed in regard to academic ability. Including the PSLE GPA as a control variable in column (3), the aggravating impacts of the reform magnify. As seen in column (4), this is caused by secondary school students in high-intensity districts improving their average PSLE performance relative to students in low-intensity districts. These findings are in line with the theoretical framework suggesting credit-constrained individuals could be academically stronger than other students. The magnitude of the results are both statistically and economically significant. The model predicts that students in the district at the 80th percentile in terms of treatment intensity experienced a 0.10 to 0.15 standard deviation drop in FTNA exam scores between 2015 and 2019 relative to students in the district at the 20th percentile.[22] The lower bound is the impact when we do not account for academic ability of students.

In Appendix Table A4, we further distinguish between public and private school students. The negative overall effects on FTNA exam scores were driven entirely by public schools. Between 2015 and 2019, the performance of public school students in high-intensity districts deteriorated progressively compared to public school students in low-intensity districts. Without accounting for PSLE scores, there was no effect of the reform on the FTNA performance of private school

[22]This is calculated by taking the difference in treatment intensity between the 80 percentile and 20th percentile district multiplied by the interaction terms for cohort 2019 in columns (2) and (3) of Table 4.

students. Accounting for PSLE scores, there was a positive and significant effect on FTNA exam scores in 2018 and 2019 for private school students in high-intensity districts. This suggests that the new students who enrolled in private school due to the reform performed slightly worse in primary school relative to other private school students.

One potential confounding factor in the analysis of exam score performance is the “Big Results Now in Education” program. The program introduced the publication of government school rankings both at the primary and secondary school level. Cilliers, Mbiti, and Zeitlin (2020) show that primary schools at the bottom of their within-district ranking were induced by the reform to improve their average PSLE performance. Our findings could be impacted by this, assuming the result holds for secondary schools and for national rankings as well. The direction of the bias depends on whether high-intensity districts have academically weaker or stronger students. If students are academically weaker, the publication of school rankings should improve average exam performance, meaning the negative results in Table 4 are biased upward. The pre-reform correlation between average FTNA performance and treatment intensity is significantly negative (p-value = 0.03), suggesting the results in Table 4 are biased upward given the results from Cilliers, Mbiti, and Zeitlin (2020) hold for secondary schools and national rankings.[23] The authors do, however, suggest that their results are driven by within-district rankings and not national rankings.

Share of students failing

While Table 4 demonstrated that the improvement in progression to secondary education came at a cost of learning, Table 5 investigates whether this negative learning effect was large enough to make more students fail the FTNA exams. Specifically, Table 5 presents the results from estimating

[23] Results are available upon request.

Table 5: *Impact of the fee-free reform on students failing*

	Share failing (all schools)	Share failing (public schools)	Share failing (private schools)
Treatment intensity $\times$ *Cohort*2015	0.117***	0.130***	0.018
	(0.033)	(0.039)	(0.018)
Treatment intensity $\times$ *Cohort*2016	0.151***	0.174***	0.003
	(0.035)	(0.041)	(0.016)
Treatment intensity $\times$ *Cohort*2017	0.182***	0.204***	0.008
	(0.036)	(0.043)	(0.018)
Treatment intensity $\times$ *Cohort*2018	0.150***	0.175***	0.013
	(0.035)	(0.041)	(0.018)
Treatment intensity $\times$ *Cohort*2019	0.141***	0.174***	0.008
	(0.034)	(0.040)	(0.016)
District and cohort FE	Yes	Yes	Yes
Additional controls	Yes	Yes	Yes
N	780	780	718
R^2	0.548	0.526	0.473

Notes: Heteroscedasticity-consistent standard errors in parentheses. Cohorts 2014 to 2019 refer to the cohorts who took the FTNA exams in 2014 to 2019, respectively. The base cohort is 2014 in all models. Additional controls include log of secondary schools, log of contemporary night-time radiance, and log of one and two years lagged night-time radiance. Districts without a private school are excluded in column (3). Appendix Table A5 reports all coefficient estimates. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.
Source: National Examination Council of Tanzania, Earth Observation Group, NOAA (2020), and authors' calculations.

Equation 5 from Section IV. Columns (1), (2), and (3) examine the impact on the share of students failing for all students, public school students only, and private school students only, respectively. A full table of all coefficient estimates is reported in Appendix Table A5. High-intensity districts experienced an increased share of students failing already in 2015 relative to low-intensity districts. This effect worsened until 2017, but declined in 2018 and 2019. In line with the results on exam scores, the negative impact on students failing was entirely driven by public schools.

We use the coefficient estimate associated with treatment intensity in 2019 from the first column of Table 5 to predict the number of students who failed due to the reform.[24] A similar exercise is performed for student enrolment.[25] Figure 6 plots, for each district, the predicted change in students who failed due to the reform (y-axis) against the predicted change in students who progressed due to the reform (x-axis). These are predicted changes between the 2014 and 2019 cohorts. An expected increase in enrolment is associated with an expected increase in students failing. While pre-reform failing rates were around 10 percent, the gradient of the scatter plot in Figure 6 is 0.25. This suggests either a larger proportion of the new students failed relative to students who would progress without the reform, or declining quality of public education leading students who would progress without the reform to fail. On the other hand, the figure also illustrates that the majority of new students arguably learned something during the two years as the scatter plot is considerably below the 45 degree line.

[24]We predict the number of students failing due to the reform by multiplying the coefficient estimate on *Treatment intensity* $\times$ *Cohort2019* with the treatment intensity and number of FTNA students.

[25]While Table 3 uses first differences, the model used for predicting the number of students enrolling in secondary education due to the reform is based on a district fixed effect model.

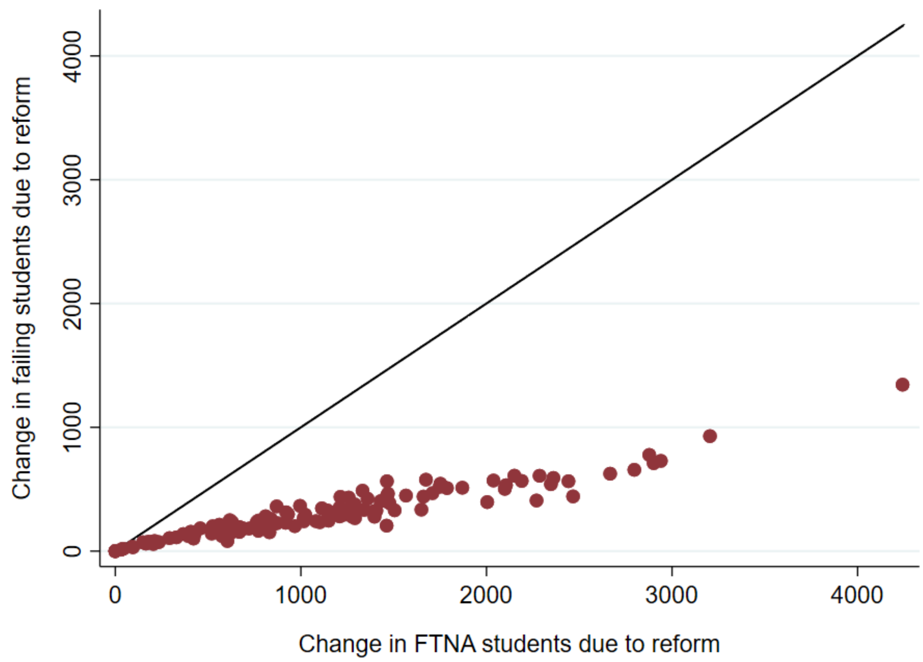


Figure 6: *Predicted changes in enrolment and students failing due to the reform*

Notes: The changes in number of students failing and progressing are based on estimated impacts of the reform between 2014 and 2019. The black line represents the 45 degree line.
Source: National Examination Council of Tanzania, Earth Observation Group, NOAA (2020), and authors' calculations.

VI. Discussion

Secondary school fees remain a substantial barrier to education even when costs are relatively low. Prior to the fee-free secondary school reform, the standard fee for enrolment was 20,000 shillings per year, corresponding to approximately 2 percent of median household consumption per adult. Despite eliminating a relatively small fee, the progression rate to secondary education increases by almost 15 percentage points over a five-year period. During the same period, PSLE scores for secondary school students improve in high-intensity districts relative to low-intensity districts, suggesting academic ability is not driving the decision to progress to secondary education for students at the margin prior to the reform. These developments demonstrate a high price sensitivity for households. Drawing on the theoretical framework, the results are best resembled by the scenario where students are credit-constrained independently of academic ability.

While enrolment surges, the quality of education is at risk. A first sign of worsening quality of public education is the increase in private school enrolment in high-intensity districts relative to low-intensity districts in spite of a wider price gap. Evaluating directly the impact on exam scores, it is clear that high-intensity districts experience a drop in performance relative to low-intensity districts, which is unattributable to students' academic abilities. Two potential reasons for the decline in quality are deterioration of the compensation fee received by the school and lack of teachers and classrooms. The compensation fee is fixed at 20,000 shillings per year, implying that inflation gradually deteriorates the value.[26] Schools further face the challenges of classroom space and hiring teachers. Despite the number of secondary school teachers rising by a larger magnitude than total enrolment, the allocation of new teachers do not favour high-intensity districts. As the enrolment effect varies significantly between districts, in contrast to hiring of new teachers, a mismatch between supply and demand arises.

The results from Tanzania cannot be used for predicting expected outcomes in other countries without caveats. It is important to keep in mind that Tanzania already had a large proportion of children finishing primary school without progressing to secondary education. The mixture of a high primary school completion rate and a low rate of progression to secondary school served as the basis for a large and immediate enrolment response. The negative impacts on learning are related to the enrolment effect. When enrolment immediately surges, it is challenging for schools to adapt with new classrooms and more teachers, and it puts pressure on the authorities to train new teachers

[26]The compensation fee is paid directly to schools, rather than indirectly through local governments, to avoid local capture or reallocation of funds (Reinikka and Svensson 2004).

as quick as possible. The substantial funds allocated to the sector take time to materialize. In the Tanzanian case, new teachers were not allocated disproportionally such that high-intensity districts received relatively more teachers. This could play a significant role in explaining the widening gap in exam score achievement between low-intensity and high-intensity districts.

VII. Conclusion

In January 2016, Tanzania implemented a secondary school reform eliminating public school fees. This led to an immediate and substantial enrolment effect, in particular for districts with a low pre-reform progression rate between primary and secondary education (high-intensity districts). While this effect was driven predominantly by public schools, private schools in high-intensity districts also experienced increased enrolment relative to private schools in low-intensity districts two years after the implementation of the reform. The magnitude of the results suggests that the district at the 80th percentile in terms of treatment intensity experienced a 12 percentage points increase in the overall progression rate between 2014 and 2019 relative to the district at the 20th percentile. The substantial enrolment effect came at a cost of learning. Between 2015 and 2019, FTNA exam scores dropped by an estimated 0.10 to 0.15 standard deviation for students in the district at the 80th percentile, in terms of treatment intensity, relative to students in the district at the 20th percentile. This finding cannot be attributed to the academic abilities of new students. High-intensity district further experienced a significant increase in the share of students failing relative to low-intensity districts.

The Tanzanian case of eliminating secondary school fees provides valuable lessons for other countries planning similar reforms. First, conditional on credit-constraints being a barrier to education, even small price changes can have profound implications on enrolment. Second, a sudden spike in enrolment challenges the education system. Consequently, policy makers should consider a gradual implementation of eliminating fees to make time for adaptation. This adaptation involves building classrooms where needed and training teachers. Importantly, areas are affected differently, meaning allocation of additional resources should be as well.

Looking ahead, the pressure on the public secondary school system is expected to grow in the years to come. The government plans to phase out the PSLE and automatically promote students to lower secondary education in 2021. This policy is projected to increase enrolment in the first four years of secondary education by approximately 370,000 students in 2025 relative to continuing only with fee-free secondary education and no automatic progression (Asim, Chugunov, and Gera 2019). While free secondary education for all children is within reach, the quality component of SDG 4.1 is challenged by the surge in enrolment and geographically disproportional needs for investment.

REFERENCES

Al-Samarrai, Samer, and Hassan Zaman. 2007. "Abolishing School Fees in Malawi: The Impact on Education Access and Equity." *Education Economics* 15 (3): 359–375.

Asim, Salman, Dmitry Chugunov, and Ravinder Gera. 2019. *Fiscal Implications of Free Education.* Technical report 31466. The World Bank.

Baird, Sarah, Craig McIntosh, and Berk Özler. 2011. "Cash or Condition? Evidence from a Cash Transfer Experiment." *The Quarterly Journal of Economics* 126 (4): 1709–1753.

Barrera-Osorio, Felipe, Marianne Bertrand, Leigh L. Linden, and Francisco Perez-Calle. 2011. "Improving the Design of Conditional Transfer Programs: Evidence from a Randomized Education Experiment in Colombia." *American Economic Journal: Applied Economics* 3 (2): 167–195.

Blimpo, Moussa P., Ousman Gajigo, and Todd Pugatch. 2019. "Financial Constraints and Girls' Secondary Education: Evidence from School Fee Elimination in The Gambia." *The World Bank Economic Review* 33 (1): 185–208.

Borkum, Evan. 2012. "Can Eliminating School Fees in Poor Districts Boost Enrollment? Evidence from South Africa." *Economic Development and Cultural Change* 60 (2): 359–398.

Brudevold-Newman, Andrew. 2019. *The Impacts of Lowering the Costs of Secondary Education: Evidence from a Fee Reduction in Kenya.* Mimeo.

Cilliers, Jacobus, Isaac M. Mbiti, and Andrew Zeitlin. 2020. "Can Public Rankings Improve School Performance? Evidence from a Nationwide Reform in Tanzania." *Journal of Human Resources:* 0119–9969R1.

Deininger, Klaus. 2003. "Does Cost of Schooling Affect Enrollment by the Poor? Universal Primary Education in Uganda." *Economics of Education Review* 22 (3): 291–305.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2019. *The Impact of Free Secondary Education: Experimental Evidence from Ghana.* Working Paper.

Earth Observation Group, NOAA. 2020. *Version 1 VIIRS Day/Night Band Nighttime Lights.* https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html. Accessed February 18, 2020.

Gajigo, Ousman. 2016. "Closing the Education Gender Gap: Estimating the Impact of Girls' Scholarship Program in The Gambia." *Education Economics* 24 (2): 167–188.

Garlick, Robert. 2013. *How Price Sensitive Is Primary and Secondary School Enrollment? Evidence from Nationwide Tuition Fee Reforms in South Africa.* Working Paper.

Glewwe, Paul, Michael Kremer, and Sylvie Moulin. 2009. "Many Children Left Behind? Textbooks and Test Scores in Kenya." *American Economic Journal: Applied Economics* 1 (1): 112–135.

Grogan, Louise. 2009. "Universal Primary Education and School Entry in Uganda." *Journal of African Economies* 18 (2): 183–211.

HakiElimu. 2017. *The Impact of the Implementation of Fee-Free Education Policy on Basic Education in Tanzania: A Qualitative Study.* Technical report. Dar-es-Salaam, Tanzania.

Hermida, Priscila. 2014. "Who Benefits from the Elimination of School Enrolment Fees? Evidence from Ecuador." *Desarrollo y Sociedad,* no. 74: 69–132.

Hoogeveen, Johannes, and Mariacristina Rossi. 2013. "Enrolment and Grade Attainment Following the Introduction of Free Primary Education in Tanzania." *Journal of African Economies* 22 (3): 375–393.

Jepsen, Christopher. 2002. "The Role of Aggregation in Estimating the Effects of Private School Competition on Student Achievement." *Journal of Urban Economics* 52 (3): 477–500.

Khandker, Shahidur, Mark Pitt, and Nobuhiko Fuwa. 2003. *Subsidy to Promote Girls' Secondary Education: The Female Stipend Program in Bangladesh.* MPRA PAPER 23688. University Library of Munich.

Lochner, Lance J., and Alexander Monge-Naranjo. 2011. "The Nature of Credit Constraints and Human Capital." *American Economic Review* 101 (6): 2487–2529.

Lucas, Adrienne M., and Isaac M. Mbiti. 2012. "Access, Sorting, and Achievement: The Short-Run Effects of Free Primary Education in Kenya." *American Economic Journal: Applied Economics* 4 (4): 226–253.

Masuda, Kazuya, and Chikako Yamauchi. 2018. *The Effects of Universal Secondary Education Program Accompanying Public-Private Partnership on Students' Access, Sorting and Achievement: Evidence from Uganda.* Working Paper 2018-4. Tokyo, Japan: Center for Economic Institutions, Institute of Economic Research, Hitotsubashi University.

Nishimura, Mikiko, Takashi Yamano, and Yuichi Sasaoka. 2008. “Impacts of the Universal Primary Education Policy on Educational Attainment and Private Costs in Rural Uganda.” *International Journal of Educational Development* 28 (2): 161–175.

Nyerere, Julius K. 1967. “Education for Self-Reliance.” In *Essays on Socialism,* 44–76. Dar es Salaam: Oxford University Press.

Omoeva, Carina, and Wael Moussa. 2018. *The Long-Term Effects of Universal Primary Education: Evidence from Ethiopia, Malawi, and Uganda.* Research Paper 18-02. Washington, D.C: Education Policy and Data Center.

PO-RALG. 2020. *Datasets - Basic Statistics Tanzania.* http://opendata.go.tz/dataset. Accessed January 25, 2020.

Reinikka, Ritva, and Jakob Svensson. 2004. “Local Capture: Evidence from a Central Government Transfer Program in Uganda.” *The Quarterly Journal of Economics* 119 (2): 679–705.

UNESCO. 2020. *Education Statistics.* http://data.uis.unesco.org. Accessed January 25, 2020.

APPENDIX A: ADDITIONAL FIGURES AND TABLES

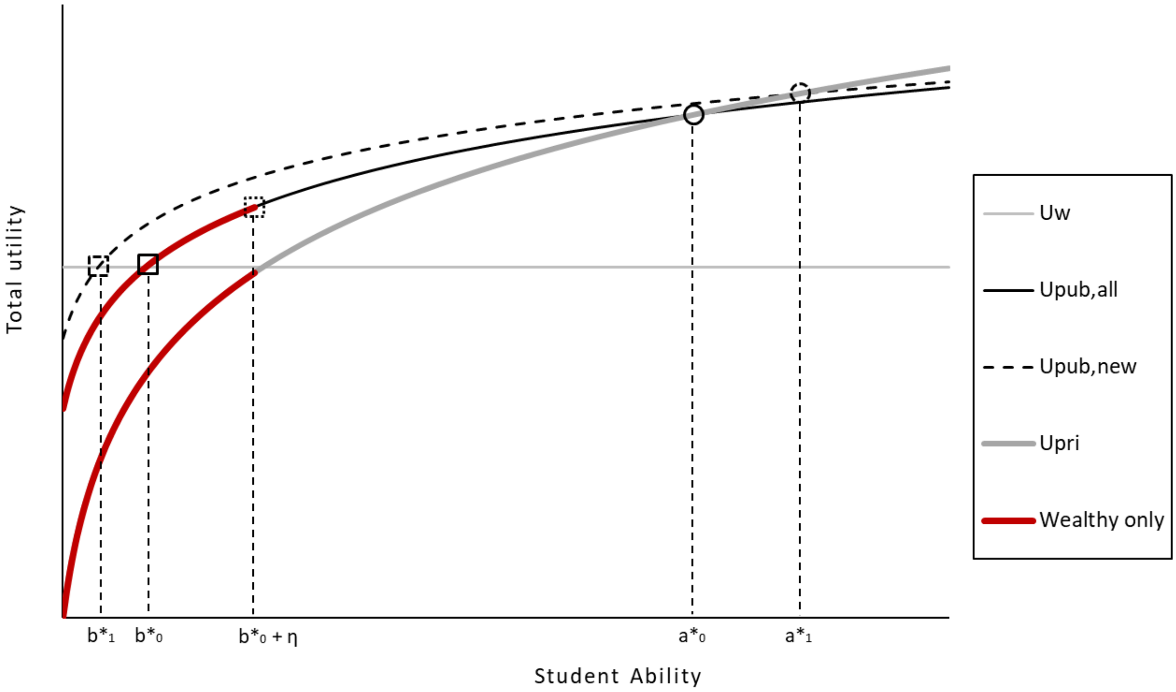


Figure A1: *Credit-constraints and small public education quality effects*

Notes: The model assumes low-ability individuals can be either non-constrained ("wealthy") or credit-constrained ("poor"), whereas high-ability individuals cannot be credit-constrained, and there *are* quality effects caused by higher enrolment in public schools. *Uw* represents utility from working in both time periods. *Upub,all* and *Upub,new* represent utility from enrolling in public school before and after a public education price reduction, respectively. *Upri* represents utility from enrolling in private school. *Wealthy only* represents utility from enrolling in public or private school for non-constrained individuals only. The solid square and circle represent the ability cut-off points, where individuals are indifferent between working or public education and public education or private education, respectively, before reducing the price on public education. The dashed square and circle represent equivalent ability cut-off points after reducing the price on public education. The dotted square represents the ability cut-off point between working and public education for credit-constrained individuals prior to a public school price reduction.

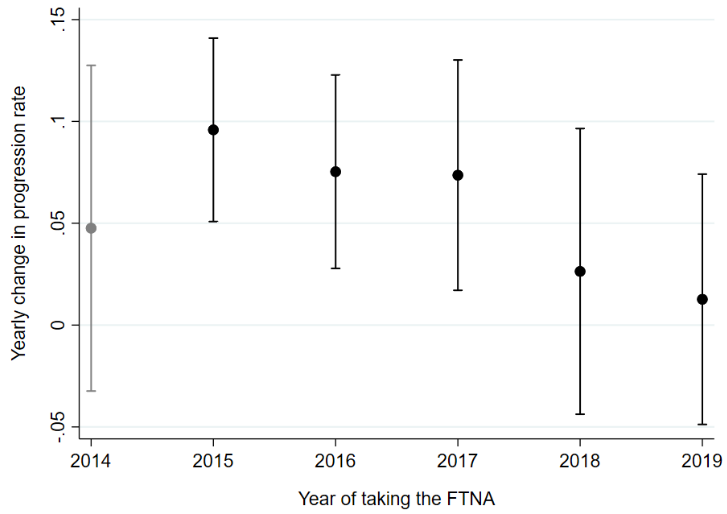


Figure A2: *Yearly impact of the fee-free reform on the overall progression rate*

Notes: The yearly estimates are the interaction terms from Table 3. The confidence intervals are with a significance level of 5 percent. 2014 is an untreated cohort.

Source: National Examination Council of Tanzania, Earth Observation Group, NOAA (2020), and authors' calculations.

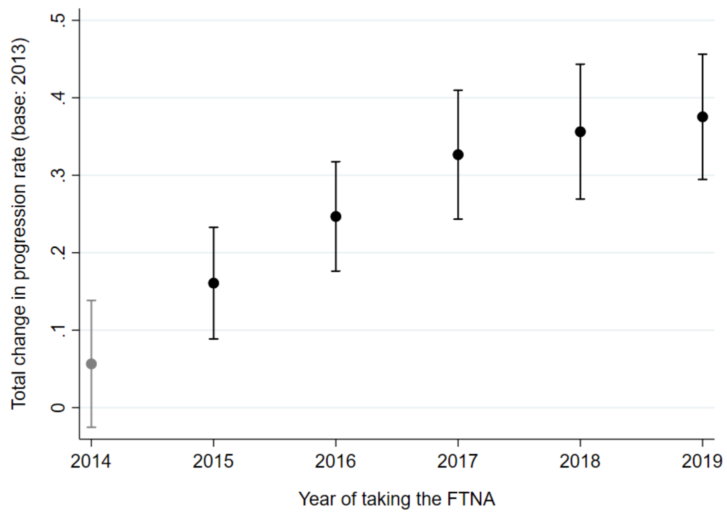


Figure A3: *Total impact of the fee-free reform on the overall progression rate since 2013*

The yearly estimates are the interaction terms from estimating Equation 3 with absolute values instead of first differences and further controlling for district fixed effects. The confidence intervals are with a significance level of 5 percent. 2014 is an untreated cohort.

Source: National Examination Council of Tanzania, Earth Observation Group, NOAA (2020), and authors' calculations.

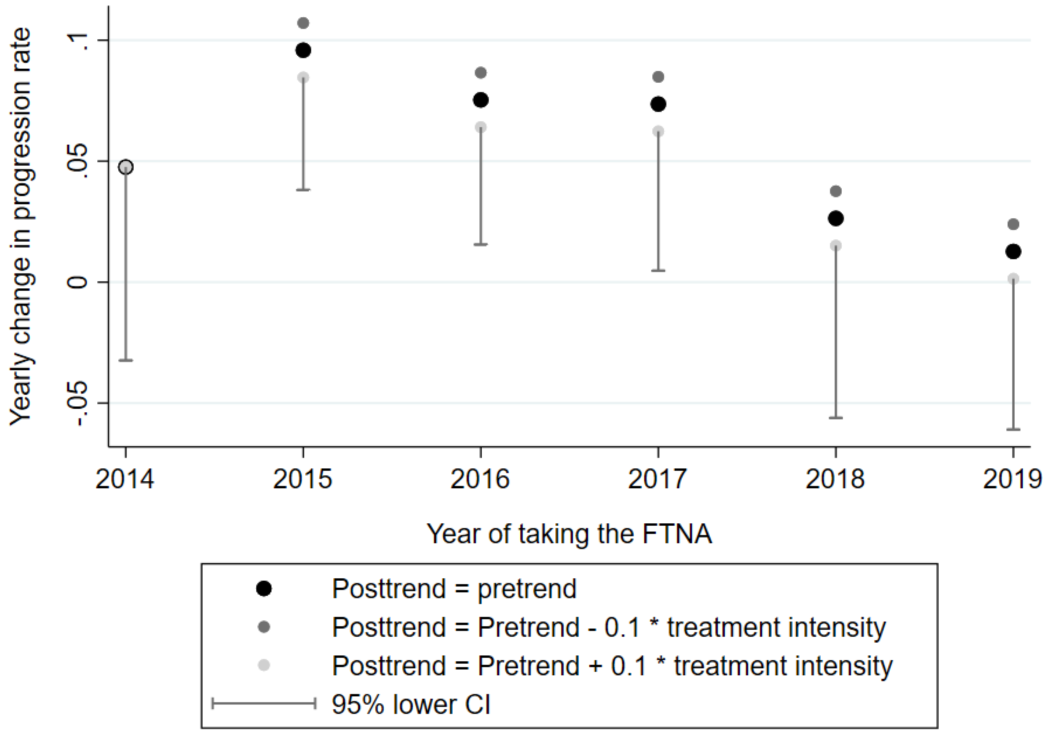


Figure A4: *Yearly impact of the fee-free reform with higher post-trends for high-intensity districts*

Notes: The yearly estimates are the interaction terms from estimating Equation 3, but letting high-intensity districts have a higher post-treatment trend than low-intensity districts. Specifically, the post-trend is defined as $Post\text{-}trend_{d,t} = Pre\text{-}trend_{d,t} + 0.1 \times I_d \times (t - 2014)$. The confidence intervals are with a significance level of 5 percent. 2014 is an untreated cohort.

Source: National Examination Council of Tanzania, Earth Observation Group, NOAA (2020), and authors' calculations.

Table A1: *Districts merged together for the empirical analysis (page 1 of 5)*

District number	District name 1	District name 2	District name 3	District name 4
1	Arusha Rural			
2	Arusha Urban			
3	Babati Urban			
4	Babati Rural			
5	Bagamoyo	Chalinze		
6	Bahi			
7	Bariadi	Itilima		
8	Biharamulo			
9	Bukoba Urban			
10	Bukoba Rural			
11	Bukombe	Mbogwe		
12	Bunda			
13	Chamwino			
14	Chato			
15	Chunya	Songwe		
16	Dodoma			
17	Geita	Nyang'hwale		
18	Hai			
19	Hanang			
20	Handeni Urban	Handeni Rural		
21	Igunga			
22	Ilala			
23	Ileje			
24	Iramba	Mkalama		
25	Iringa Urban			
26	Iringa Rural			
27	Kahama TC	Msalala	Ushetu	
28	Karagwe	Kyerwa		

Districts merged together for the empirical analysis (page 2 of 5)

District number	District name 1	District name 2	District name 3	District name 4
29	Karatu			
30	Kasulu Urban	Kasulu Rural	Buhigwe	
31	Kibaha Rural			
32	Kibaha Urban			
33	Kibondo	Kakonko		
34	Kigoma Urban			
35	Kigoma Rural	Uvinza		
36	Kilindi			
37	Kilolo			
38	Kilombero	Ifakara		
39	Kilosa	Gairo		
40	Kilwa			
41	Kinondoni	Ubungo		
42	Kisarawe			
43	Kishapu			
44	Kiteto			
45	Kondoa	Chemba		
46	Kongwa			
47	Korogwe Rural			
48	Korogwe Urban			
49	Kwimba			
50	Kyela			
51	Lindi Urban			
52	Lindi Rural			
53	Liwale			
54	Longido			
55	Ludewa			
56	Lushoto	Bumbuli		

Districts merged together for the empirical analysis (page 3 of 5)

District number	District name 1	District name 2	District name 3	District name 4
57	Mafia			
58	Magu	Busega		
59	Makete			
60	Manyoni	Itigi		
61	Masasi Urban	Masasi Rural		
62	Maswa			
63	Mbarali			
64	Mbeya Urban			
65	Mbeya Rural			
66	Mbinga	Nyasa		
67	Mbozi	Momba	Tunduma	
68	Mbulu			
69	Meatu			
70	Meru			
71	Missenyi			
72	Misungwi			
73	Mpanda Urban			
74	Mkinga			
75	Mkuranga			
76	Monduli			
77	Morogoro Urban			
78	Morogoro Rural			
79	Moshi Urban			
80	Moshi Rural			
81	Mpanda Rural	Mlele	Nsimbo	Mpimbwe
82	Mpwapwa			
83	Mtwara Urban	Mtwara Rural	Nanyamba	
84	Mufindi	Mafinga		

Districts merged together for the empirical analysis (page 4 of 5)

District number	District name 1	District name 2	District name 3	District name 4
85	Muheza			
86	Muleba			
87	Musoma Urban			
88	Musoma Rural	Butiama		
89	Mvomero			
90	Mwanga			
91	Nyamagana	Ilemela	Mwanza	
92	Nachingwea			
93	Namtumbo			
94	Nanyumbu			
95	Newala			
96	Ngara			
97	Ngorongoro			
98	Njombe Rural	Makambako	Wanging'ombe	
99	Njombe Urban			
100	Nkasi			
101	Nzega			
102	Pangani			
103	Rombo			
104	Rorya			
105	Ruangwa			
106	Rufiji	Kibiti		
107	Rungwe	Busokelo		
108	Same			
109	Sengerema	Buchosa		
110	Serengeti			
111	Shinyanga Urban			
112	Shinyanga Rural			

Districts merged together for the empirical analysis (page 5 of 5)

District number	District name 1	District name 2	District name 3	District name 4
113	Siha			
114	Sikonge			
115	Simanjiro			
116	Singida Urban			
117	Singida Rural	Ikungi		
118	Songea Urban			
119	Songea Rural	Madaba		
120	Sumbawanga Urban			
121	Sumbawanga Rural	Kalambo		
122	Tabora	Uyui		
123	Tandahimba			
124	Tanga			
125	Tarime			
126	Temeke	Kigamboni		
127	Tunduru			
128	Ukerewe			
129	Ulanga	Malinyi		
130	Urambo	Kaliua		

Table A2: *Yearly impact of the fee-free reform on district-level progression rates*

	Δ Progression rate (all schools)	Δ Progression rate (to public schools)	Δ Progression rate (to private schools)
Treatment intensity × *Cohort*2014	0.048		
	(0.041)		
Treatment intensity × *Cohort*2015	0.096***	0.095***	0.008
	(0.023)	(0.019)	(0.011)
Treatment intensity × *Cohort*2016	0.075***	0.080***	0.004
	(0.024)	(0.019)	(0.011)
Treatment intensity × *Cohort*2017	0.074**	0.059**	0.017**
	(0.029)	(0.026)	(0.008)
Treatment intensity × *Cohort*2018	0.026	0.014	0.032**
	(0.036)	(0.031)	(0.016)
Treatment intensity × *Cohort*2019	0.013	−0.051*	0.070***
	(0.031)	(0.030)	(0.010)
Cohort2015	0.006		
	(0.027)		
Cohort2016	0.096***	0.089***	0.005
	(0.026)	(0.013)	(0.010)
Cohort2017	0.193***	0.202***	−0.017**
	(0.028)	(0.016)	(0.008)
Cohort2018	0.115***	0.147***	−0.025**
	(0.028)	(0.020)	(0.011)
Cohort2019	0.101***	0.127***	−0.035***
	(0.030)	(0.018)	(0.009)
Δ*District pre-trend*	0.113***	−0.053*	−0.008
	(0.035)	(0.029)	(0.009)
Δ*Secondary schools (log)*	0.144*	−0.027	0.134***
	(0.075)	(0.091)	(0.042)
Δ*Radiance (contemporary, log)*	−0.056	−0.112	0.065*
	(0.082)	(0.079)	(0.036)
Δ*Radiance (1-year lag, log)*	−0.001	−0.089	−0.001
	(0.104)	(0.100)	(0.040)
Δ*Radiance (2-years lag, log)*	−0.258***	−0.151*	−0.079**
	(0.094)	(0.079)	(0.039)
Constant	−0.085***	−0.091***	−0.004
	(0.024)	(0.012)	(0.007)
N	780	650	650
R^2	0.525	0.532	0.266

Notes: Heteroscedasticity-consistent standard errors in parentheses. Cohorts 2014 to 2019 refer to the cohorts who took the FTNA exams in 2014 to 2019, respectively. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Source: National Examination Council of Tanzania, Earth Observation Group, NOAA (2020), and authors' calculations.

Table A3: *Impact of the fee-free reform on individual exam scores*

	FTNA GPA	FTNA GPA	FTNA GPA	PSLE GPA
Treatment intensity × *Cohort*2014	0.081			
	(0.079)			
Treatment intensity × *Cohort*2016	−0.142***	−0.135***	−0.245***	0.191***
	(0.049)	(0.049)	(0.061)	(0.057)
Treatment intensity × *Cohort*2017	−0.291***	−0.249***	−0.442***	0.335***
	(0.064)	(0.060)	(0.087)	(0.083)
Treatment intensity × *Cohort*2018	−0.364***	−0.300***	−0.403***	0.179*
	(0.063)	(0.066)	(0.109)	(0.107)
Treatment intensity × *Cohort*2019	−0.386***	−0.272***	−0.389***	0.203*
	(0.064)	(0.059)	(0.099)	(0.104)
PSLE GPA			0.575***	
			(0.016)	
Cohort2014	−0.044			
	(0.041)			
Cohort2016	0.049**	0.066**	0.119***	−0.091***
	(0.024)	(0.026)	(0.027)	(0.024)
Cohort2017	0.183***	0.200***	0.288***	−0.152***
	(0.034)	(0.028)	(0.042)	(0.043)
Cohort2018	0.185***	0.139***	0.202***	−0.110*
	(0.044)	(0.036)	(0.054)	(0.059)
Cohort2019	0.240***	0.210***	0.277***	−0.116
	(0.045)	(0.042)	(0.065)	(0.072)
Secondary schools (log)	0.908***	0.159	−0.053	0.370
	(0.236)	(0.205)	(0.363)	(0.372)
Radiance (contemporary, log)	−0.329**	−0.281***	−0.281*	0.000
	(0.140)	(0.104)	(0.149)	(0.152)
Radiance (1-year lag, log)	0.082	0.403**	0.300	0.179
	(0.156)	(0.189)	(0.274)	(0.239)
Radiance (2-years lag, log)	−0.345*	−0.508***	−0.474*	−0.059
	(0.184)	(0.163)	(0.245)	(0.191)
Constant	−3.242***	−0.518	0.301	−1.423
	(0.880)	(0.761)	(1.350)	(1.384)
District FE	Yes	Yes	Yes	Yes
N	2,745,748	1,537,231	1,537,231	1,537,231
R^2	0.042	0.048	0.367	0.051

Notes: Standard errors clustered at the district level in parentheses. Cohorts 2014 to 2019 refer to the cohorts who took the FTNA exams in 2014 to 2019, respectively. The base cohort is 2015 in all models. FTNA and PSLE GPAs are standardized within each cohort. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Source: National Examination Council of Tanzania, Earth Observation Group, NOAA (2020), and authors' calculations.

Table A4: *Impact of the fee-free reform on individual exam scores (public-private distinction)*

	FTNA GPA	FTNA GPA	FTNA GPA	FTNA GPA
Treatment intensity $\times$ *Cohort*2014	0.135		−0.032	
	(0.087)		(0.114)	
Treatment intensity $\times$ *Cohort*2016	−0.153***	−0.229***	0.052	0.037
	(0.044)	(0.057)	(0.064)	(0.074)
Treatment intensity $\times$ *Cohort*2017	−0.320***	−0.427***	0.029	0.014
	(0.062)	(0.082)	(0.063)	(0.078)
Treatment intensity $\times$ *Cohort*2018	−0.417***	−0.430***	0.057	0.157**
	(0.063)	(0.111)	(0.077)	(0.079)
Treatment intensity $\times$ *Cohort*2019	−0.453***	−0.407***	0.015	0.146**
	(0.063)	(0.100)	(0.086)	(0.065)
PSLE GPA		0.496***		0.557***
		(0.017)		(0.015)
Cohort2014	−0.074		0.132***	
	(0.046)		(0.047)	
Cohort2016	0.068***	0.104***	0.003	0.116***
	(0.022)	(0.026)	(0.024)	(0.026)
Cohort2017	0.243***	0.288***	0.137***	0.237***
	(0.031)	(0.038)	(0.038)	(0.035)
Cohort2018	0.277***	0.216***	0.170***	0.079*
	(0.044)	(0.055)	(0.056)	(0.041)
Cohort2019	0.358***	0.272***	0.381***	0.227***
	(0.047)	(0.066)	(0.066)	(0.051)
Secondary schools (log)	0.583***	−0.218	0.872**	0.428
	(0.218)	(0.368)	(0.371)	(0.325)
Radiance (contemporary, log)	−0.404**	−0.314**	0.102	−0.076
	(0.160)	(0.149)	(0.214)	(0.146)
Radiance (1-year lag, log)	0.090	0.373	0.034	0.059
	(0.155)	(0.280)	(0.213)	(0.185)
Radiance (2-years lag, log)	−0.364*	−0.415	−0.381	−0.327
	(0.196)	(0.269)	(0.275)	(0.203)
Constant	−2.212***	0.785	−2.455*	−0.789
	(0.807)	(1.357)	(1.440)	(1.282)
District FE	Yes	Yes	Yes	Yes
Sample (public/private school students)	Public	Public	Private	Private
N	2,304,422	1,388,767	440,985	148,354
R^2	0.0391	0.300	0.109	0.538

Notes: Standard errors clustered at the district level in parentheses. Cohorts 2014 to 2019 refer to the cohorts who took the FTNA exams in 2014 to 2019, respectively. The base cohort is 2015 in all models. FTNA and PSLE GPAs are standardized within each cohort. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Source: National Examination Council of Tanzania, Earth Observation Group, NOAA (2020), and authors' calculations.

Table A5: *Impact of the fee-free reform on students failing*

	Share failing (all schools)	Share failing (public schools)	Share failing (private schools)
Treatment intensity × *Cohort*2015	0.117***	0.130***	0.018
	(0.033)	(0.039)	(0.018)
Treatment intensity × *Cohort*2016	0.151***	0.174***	0.003
	(0.035)	(0.041)	(0.016)
Treatment intensity × *Cohort*2017	0.182***	0.204***	0.008
	(0.036)	(0.043)	(0.018)
Treatment intensity × *Cohort*2018	0.150***	0.175***	0.013
	(0.035)	(0.041)	(0.018)
Treatment intensity × *Cohort*2019	0.141***	0.174***	0.008
	(0.034)	(0.040)	(0.016)
Cohort2015	−0.054***	−0.064***	−0.001
	(0.018)	(0.023)	(0.008)
Cohort2016	−0.073***	−0.088***	0.000
	(0.019)	(0.024)	(0.007)
Cohort2017	−0.093***	−0.115***	−0.010
	(0.021)	(0.026)	(0.008)
Cohort2018	−0.082***	−0.109***	−0.002
	(0.024)	(0.030)	(0.010)
Cohort2019	−0.088***	−0.116***	−0.015
	(0.026)	(0.032)	(0.010)
Secondary schools (log)	−0.217***	−0.212***	−0.006
	(0.063)	(0.069)	(0.036)
Radiance (contemporary, log)	0.132*	0.178**	0.004
	(0.069)	(0.081)	(0.028)
Radiance (1-year lag, log)	−0.039	−0.043	−0.032
	(0.094)	(0.107)	(0.045)
Radiance (2-years lag, log)	0.062	0.045	0.049
	(0.081)	(0.094)	(0.044)
Constant	0.820***	0.817***	0.034
	(0.212)	(0.235)	(0.123)
District FE	Yes	Yes	Yes
N	780	780	718
R^2	0.548	0.526	0.473

Notes: Heteroscedasticity-consistent standard errors in parentheses. Cohorts 2014 to 2019 refer to the cohorts who took the FTNA exams in 2014 to 2019, respectively. The base cohort is 2014 in all models. Districts without a private school are excluded in Column (3). Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Source: National Examination Council of Tanzania, Earth Observation Group, NOAA (2020), and authors' calculations.

APPENDIX B: CREDIT-CONSTRAINED INDIVIDUALS INDEPENDENT OF ABILITY

Section II assumes low-ability poor individuals are credit-constrained, whereas high-ability poor individuals are not. In this robustness section, it is assumed instead that poor individuals are randomly credit-constrained. It is further assumed that there are no effects on quality of public education when reducing the price. Changing the assumption on quality effects have the same implications as in the baseline case where only low-ability poor individuals are credit-constrained.

Total enrolment and enrolment in public education increase following a price reduction on public education. As net return to public education increases, some individuals change preference from private to public education. Consequently, enrolment in private education decreases. While the effect on average ability in public education is still ambiguous, the relative intake of new low- and high-ability individuals changes. That is, more of the new public school students are high-ability individuals compared to the second model in Section II. Some of the new public school students even prefer private education, but only the constraint on public education is alleviated and these high-ability individuals therefore enrol in public education.

Figure B1 is similar to Figure 2 with the exception that poor individuals are randomly credit-constrained. This can be seen as the thick red lines (*Wealthy only*) only represent wealthy individuals. Hence, even poor individuals with a high ability level can be credit-constrained. For instance, at four different ability levels above a_0^*, poor students are credit-constrained. They prefer to attend private education, but they have to work in both periods due to credit constraints. When reducing the price on public education, credit constraints on public education are wiped out. The high-ability students enrol in public schools because the credit constraints on private education remain.

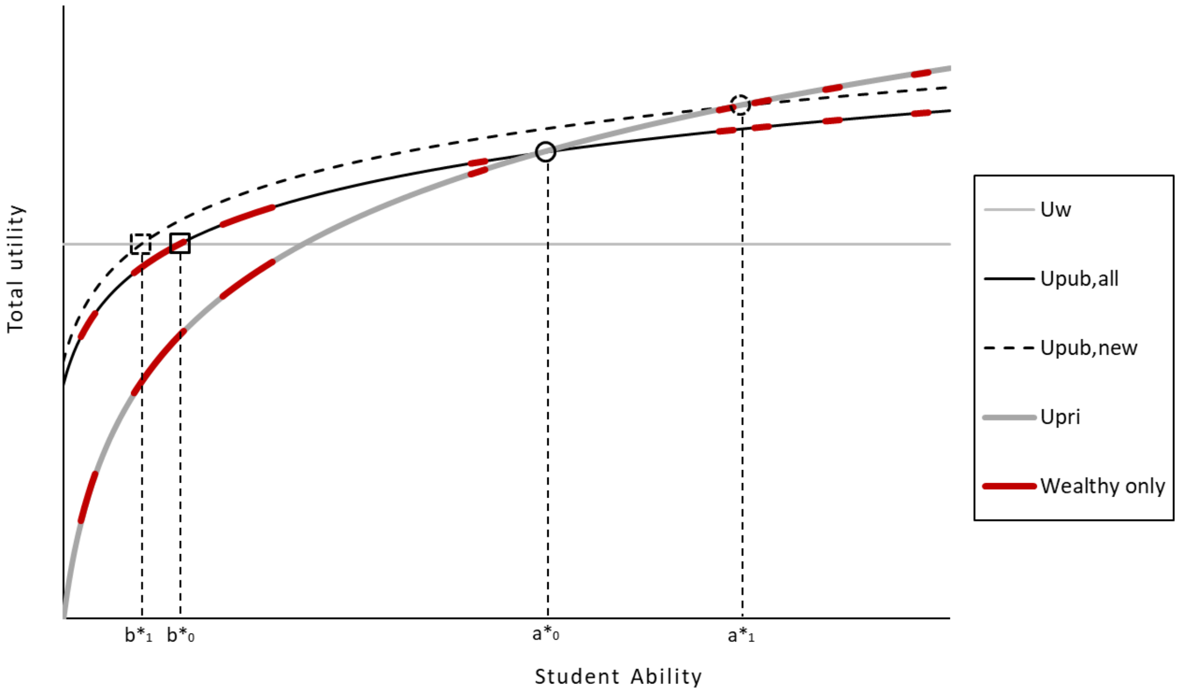


Figure B1: *Credit-constraints randomly distributed among poor individuals*

Notes: The model assumes individuals are randomly credit-constrained and there are no quality effects caused by higher enrolment in public schools. *Uw* represents utility from working in both time periods. *Upub,all* and *Upub,new* represent utility from enrolling in public school before and after a public education price reduction, respectively. *Upri* represents utility from enrolling in private school. *Wealthy only* represents utility from enrolling in public or private school for non-constrained individuals only. The solid square and circle represent the ability cut-off points, where individuals are indifferent between working or public education and public education or private education, respectively, before reducing the price on public education. The dashed square and circle represent equivalent ability cut-off points after reducing the price on public education.

Chapter 4

Predicting Local State Capacity in Sub-Saharan Africa: A Machine Learning Approach

Predicting Local State Capacity in Sub-Saharan Africa: A Machine Learning Approach*

Gustav Agneman[1], Kasper Brandt[1], Christoffer Cappelen[2], and David Sjöberg

[1]*Department of Economics, University of Copenhagen*

[2]*Department of Political Science, University of Copenhagen*

Despite the need to measure state capacity at a sub-national level, most studies still use country-level indicators as rough approximations of the local counterpart. We estimate a measure of state capacity at the 2.5×2.5 *arc-minutes grid cell level (*$\approx$ *5 kilometers) for Sub-Saharan Africa. The measure builds on geocoded survey-based data on local state capacity which we predict and extrapolate using an ensemble of regression trees. We demonstrate the usefulness of measuring state capacity at a disaggregated level by including our local state capacity index as a moderating factor in the relationship between oil wealth and armed conflict. The findings suggest that cells with higher local state capacity face lower risks of conflicts caused by oil price hikes.* (*JEL* H11, H70, Q34)

*We wish to thank Eliana La Ferrara, Chris Blattman, James Robinson, Sir Paul Collier, Austin Wright, David Dreyer Lassen, Andreas Bjerre-Nielsen and Jacob Gerner Hariri for their insightful and generous input.

I. INTRODUCTION

Strong states are thought to discourage uprisings and be more capable of fending off rivals if need be (Fearon and Laitin 2003; Buhaug and Rød 2006; Fjelde and De Soysa 2009; Braithwaite 2010). Yet, empirically identifying the role of state capacity in reducing the risk of conflict has proven complicated, not least because of the challenges in measuring conflict and state capacity at suitable levels of aggregation. Both conflict and state capacity are inherently local phenomena: conflicts have increasingly turned small-scale (Sundberg and Melander 2013), and states, especially in developing countries, rarely pertain equal territorial control within their borders (Herbst 2000; Fearon and Laitin 2003). Consequently, mapping geographic variation in state capacity is imperative to properly account for its impact on conflict.

In this paper, we demonstrate a new approach to measuring local state capacity. First, we construct an index of perceived local state capacity using geocoded survey data from the Afrobarometer (Afrobarometer Data 2004, 2005). The index is based on a factor analysis of several survey items representing the state's territorial reach in terms of law enforcement, information gathering, and public goods provision. Second, we predict the state capacity index using non-parametric machine learning techniques. The model includes as predictors, among others, travel time to the capital, local infrastructural accessibility, population figures, and topographic ruggedness. Third, we extrapolate the model to all 2.5×2.5 arc-minutes grid cells ($\approx 5 \times 5$ kilometers) in Sub-Saharan Africa.[1] We corroborate the validity of our measure of local state capacity by correlating it with contemporary ethnic power relations (Wucherpfennig et al. 2011), pre-colonial centralization (Michalopoulos and

[1] This approach follows Mosser et al. (2019) predicting vaccination coverage across Africa based on survey data and grid cell covariates, and Bergquist et al. (2019) predicting agricultural output for all households in Uganda based on 3,000 survey households and explanatory variables from census data.

Papaioannou 2013) and local vaccination coverage (Mosser et al. 2019).

We further showcase the usefulness of our measure of local state capacity by considering it as a moderating factor in the relationship between oil wealth and conflict. This exercise entails the construction of a panel dataset on oil wealth and oil-related conflicts for Sub-Saharan Africa in 2001–2019.[2] Using this data, we estimate fixed effects models with oil-related conflict as the dependent variable and the interaction between oil price, oil deposits, and local state capacity as the explanatory variable of main interest. The results show that while areas with oil deposits face a higher risk of conflict when the oil price increases, this exogenous shock to oil wealth has no effect on oil-related conflicts in areas with high local state capacity.

Our work contributes to the literature concerned with measuring local state capacity in developing countries, which until now has relied on e.g. historical state institutions (Michalopoulos and Papaioannou 2013; Acemoglu, García-Jimeno, and Robinson 2015; Dell, Lane, and Querubin 2018), census data (Lee and Zhang 2017), sub-national tax collection (Harbers 2015), satellite data (Koren and Sarbahi 2018) and survey data (Luna and Soifer 2017; Fergusson, Molina, and Robinson 2020). While these are innovative approaches, they inevitably entail limitations. Historical data can only inform on the effect of institutions that persisted over time, and is often confined to a specific dimension of the state, such as road provision (Acemoglu, García-Jimeno, and Robinson 2015). The narrow scope is a limitation also when using quality of census data as a proxy for state capacity.[3] Sub-national tax collection data is a useful measure of the extractive capacity of state institutions, but is naturally restricted by incommensurable cross-country differences. Night-time

[2] We define oil conflicts as conflict events in the ACLED data (Raleigh et al. 2010) that include a reference to either oil or petroleum.

[3] Lee and Zhang (2017) argue that deviations from a smooth age distribution within a sub-national area reveal the state is incapable of conducting a reliable census in that area, suggestive of low state capacity.

light emissions and other satellite data, on the other hand, provide for global samples, but can be noisy proxies of local state capacity. While survey data can more accurately capture information on citizens' perceptions of state institutions (Fergusson, Molina, and Robinson 2020), it does not comprehensively cover cross-country territories, thereby providing only a patchy view of local state capacity. Our approach to predict and extrapolate a spatially disaggregated measure of state capacity suggests a novel way forward for studies concerned with measuring state capacity in data scarce territories. By combining the satellite data and survey-based approaches, we overcome some of the limitations inherent in each methodology.

We also contribute to the literature studying the effects of natural resource wealth, in particular oil wealth, on conflict. The point of departure of this literature is the abundant anecdotal evidence of oil-induced conflict in oil-producing developing countries (Ross 2012). Empirical studies have confirmed the link between oil and conflict in specific contexts (Dube and Vargas 2013; Nwokolo 2018), but the external validity of these findings have been contested (Cotet and Tsui 2013; Bazzi and Blattman 2014). A number of studies further suggest that the risk of conflict depends on contextual factors, e.g. the spatial distribution of natural resources (Morelli and Rohner 2015; Lessmann and Steinkraus 2019), redistributive government transfers (Justino 2011; Justino and Martorano 2018; Cordella and Onder 2020), compliance with corporate social responsibility (CSR) (Berman et al. 2017), the relative strength of the government vis-à-vis rebel groups (Buhaug 2010), and whether oil deposits are onshore or offshore (Nordvik 2019). We advance the literature on oil-induced conflict by improving the precision of both the conflict data and the spatial unit of analysis. Moreover, by documenting the moderating role of local state capacity in the oil–conflict relationship, we provide evidence for a contextual factor of first-order importance.

The paper is structured as follows. In Section II, we discuss how previous studies have conceptualized and attempted to measure state capacity. In Section III, we describe the data and strategy used for predicting local state capacity. In Section IV, we validate our predicted measure by linking it to other data sources related to state capacity. In Section V, we describe and present the empirical application of local state capacity by examining its potentially moderating effect on the relationship between oil wealth and risk of conflict. In Section VI, we discuss the limitations of our approach to predict local state capacity. Section VII concludes.

II. State capacity: concept and measurement

A. *State capacity concept*

Over the last decades there has been a renewed focus on the importance of state institutions. Well-functioning states have been argued to promote economic development (Acemoglu and Robinson 2012), democracy (Fukuyama 2005), and other development outcomes. It is perhaps most notable in its absence. At the extreme, weak or failed states suffer from a lack of public goods provision (Rotberg 2003) and an increased likelihood of intra-state violence (Fearon and Laitin 2003; Buhaug and Rød 2006; Fjelde and De Soysa 2009; Braithwaite 2010; Hendrix 2011).

At the center of this literature is the concept of state capacity. State capacity is typically defined as "a government's ability to make and enforce rules, and to deliver services" (Fukuyama 2013, p. 350), or "the ability of the state to effectively implement official goals" (Hanson and Sigman 2013, p. 2). Scholars are sometimes concerned with state capacity in relation to particular functions. Fearon and Laitin (2003), for example, focus on the state's capacity to effectively monitor and access remote areas where rebel groups often form and operate. But state capacity also acts as a more general concept, capturing the state's ability across functions or domains to "get things done".

The concept of state capacity can of course be traced back to Weber's canonical definition of the state as "a human community that (successfully) claims *the monopoly of the legitimate use of physical force* within a given territory" (Weber 1991, p. 78). However, it is perhaps best captured by Michael Mann's concept of the *infrastructural power* of the state: "the capacity of the state to actually penetrate civil society, and to implement logistically political decisions throughout the realm" (Mann 1984, p. 189; see also Soifer 2008; Soifer and vom Hau 2008).

Despite the frequent use of the term state capacity, there is a lack of consensus on the more specific meaning and operationalization of the term, and it is further complicated by the multitude of related and synonymous terms, e.g. state strength, state power, political capacity, or state fragility.[4] The usage ranges from relatively minimalist understandings and measurements, e.g. tax collection and law enforcement (e.g. Thies 2010; Harbers 2015), to more maximalist approaches emphasizing the state's role as provider of a range of public goods and services (e.g. Rotberg 2003). Another development in the literature is that scholars have begun to disentangle the concept, categorizing state capacity into several distinct dimensions such as extractive, coercive, and administrative capacity (e.g. Hendrix 2010; Hanson and Sigman 2013; Berwick and Christia 2018). Extractive capacity refers to the ability to collect revenue, coercive capacity to the enforcement of order, and administrative capacity refers more broadly to the ability to effectively regulate society and deliver public goods and services.

Disentangling the concept of state capacity into multiple dimensions have added to the conceptual development by making it implicit which dimensions, or functions, are of concern. However, it also stands to reason that these dimensions are highly co-dependent. Extractive capacity requires

[4] Some typologies also confuse state capacity with related, but distinct concepts. For instance, Gleditsch and Ruggeri (2010) distinguish between repressive and accommodative forms of state capacity, but this distinction is perhaps related less to the capacity of the state and more to how the state is governed, i.e. the regime type; still important concepts, but also different from the usual conception of state capacity.

coercive and administrative capacity, which is in turn improved by the ability to collect revenue and finance the coercive apparatus. At minimum, a prerequisite, or minimal condition, for state capacity is the ability of the state to project power (Lindvall and Teorell 2016; Berwick and Christia 2018). A high capacity state must, first of all, be able to enforce its rules and collect revenues across the territory it claims to rule (Johnson and Koyama 2017, p. 2).

This leads to another aspect of the state's 'infrastructural power' that is particularly pertinent to the present paper, and which has only recently gained attention, especially in the empirical literature. The state's ability to project power and penetrate society is not necessarily uniformly distributed across its territory. As Herbst (2000) noted "the fundamental problem facing state-builders in Africa ... has been to project authority over inhospitable territories that contain relatively low densities of people" (p. 11). Mann's concept of infrastructural power arguably also implies a spatial dimension of state power: strong states are able to reach and penetrate remote areas, whereas weaker states often lack this ability (see also O'Donnell 1993; Soifer and vom Hau 2008; Fukuyama 2013). This idea can also be found in Boulding's concept of a *loss-of-strength gradient* (Boulding 1962). While initially applied to international conflict, the idea also captures the projection of strength in domestic settings. Basically, a state's strength (or capacity) is largest at its home base and diminishing as one moves away from the centre. The extent of this decline depends on the loss-of-strength gradient capturing the cost of exerting authority across space (see also Buhaug 2010).

B. Measuring state capacity

The spatial heterogeneity of states' projection of power presents researchers with challenges in measurement. Traditionally, the most prominent measure of state capacity is based on indicators of tax revenues—typically some ratio of tax revenues to total GDP. As has long been recognized in the fiscal sociology literature (Schumpeter [1918] 1954; Tilly 1975; Levi 1989), taxation is a central characteristic of the state. The more capable the state is, the more it can extract revenue from its citizens. However, there are several issues with using measures of taxation to proxy state capacity. First, it captures not only extractive capacity but also to some extent bureaucratic and coercive aspects of the state, making it unclear what exactly is proxied by the tax ratio. Second, tax revenues is not only a function of the state's ability to tax, but also of political preferences (Fukuyama 2013).[5] Third, and most importantly for the present purposes, is that tax revenue indicators are typically not readily available across large samples of countries.[6]

Instead of relying on tax revenues or tax ratios to proxy state capacity, scholars have increasingly turned to specific important dimensions of the state to approximate its presence. One innovative approach introduced by Lee and Zhang (2017) aims to capture a minimalist concept of state presence, building on the idea that states are intrinsically in need of information (e.g., in order to tax its citizens, it needs to know its citizens). They use census data at both the national and sub-national level and consider the deviation from a smooth age distribution to develop a measure of state presence. While their approach is conceptually clean and intuitive, it still suffers from a lack of cross-country availability, especially at the sub-national level. Some historical measures of statehood, such as Murdock's political centralization of pre-colonial ethnic groups (Murdock 1959), allow researchers to study "stateness" at a global scale. However, although such measures have been found to positively predict contemporary development through institutional persistence

[5] In addition, some revenue sources are easier to collect than others (Sánchez de la Sierra 2020) (e.g., revenues from trade and natural resources), and total tax revenues are therefore not always reflecting the state's extractive capacity, but the ease with which taxes can be collected.

[6] Harbers (2015) collects data on municipal tax revenues to capture local state capacity in Ecuador, but it is difficult to find comparable data on a larger scale.

(e.g. Gennaioli and Rainer 2007; Michalopoulos and Papaioannou 2013), they lack both precision and generalizability exactly because of the historical dimension.[7]

An emergent literature has propounded a survey-based approach to capture state capacity. Luna and Soifer (2017) propose a survey-based approach to capture the variation of state capacity across space. They focus on specific dimensions, such as territorial reach and taxation, and formulate a series of questions to gauge peoples' experience with the state. For instance, they ask how long it would take for the police to arrive at their home and how often they receive a receipt without requesting one (the latter intended to capture enforcement of VAT taxes). Similarly, Fergusson, Molina, and Robinson (2020) use survey data on tax evasion to measure local state capacity and find significantly positive correlations with measures of clientelism. The survey-based approach has the advantage of enabling specifically tailored questions to the concept of interest, while also capturing sub-national variation. However, conducting surveys across a large number of countries (while still covering the entire territory of each country) is costly and survey-based measures of state capacity are, thus, inherently limited in their spatial reach.

Another approach to measuring local state capacity takes advantage of the increasing availability of geographical information systems (GIS), such as satellite data on both human impacts and geographical factors (e.g. terrain). Koren and Sarbahi (2018), for instance, attempt to overcome the limitations of national-level measures by using night-time light emissions as proxy for sub-national state capacity. However, while available at the global level, light density is a poor single proxy for state capacity. It is typically used to capture local economic development (e.g. Henderson, Storeygard, and Weil 2012; Michalopoulos and Papaioannou 2013). It therefore falls prey to the

[7] Other creative measures such as the presence of post-offices (Acemoglu and Robinson 2012) or the presence of colonial officials and royal roads (Acemoglu, García-Jimeno, and Robinson 2015) have also added to the range of available measures, but these are often confined to specific countries or periods and thus not readily applicable on a larger scale.

same issues as Fearon and Laitin (2003), who use per capita income to capture state capacity at the national level. Income captures not only state capacity, but also other factors potentially linked to the outcome of interest (e.g. in the case of conflict: opportunity costs of rebellion (Collier and Hoeffler 2004) or political preferences (Fukuyama 2013, p. 353)). Still others have used measures of mountainous terrain (Hendrix 2011), road density (Buhaug and Rød 2006), or simply distance to capital (Buhaug 2010). While each measure might correlate with state capacity, they are also at risk of capturing other aspects about local areas.[8] In addition, it is not clear whether one measure is superior to another, or whether these different proxies complement each other in predicting state capacity.

III. A NEW APPROACH TO MEASURING STATE CAPACITY

In this section, we outline the three steps we take to construct a measure of state capacity at the local level. First, we estimate an index of perceived local state capacity using data from Afrobarometer (Afrobarometer Data 2004, 2005). Second, we predict the measure of state capacity using both local and national-level variables. Third, we extrapolate the state capacity index to all 2.5×2.5 arc-minutes grid cells ($\approx 5 \times 5$ kilometers) in Sub-Saharan Africa.

A. *Estimating an index of local state capacity based on survey data*

In order to build a measure of state capacity, we combine several features of the approaches discussed in the previous section. Like e.g. Luna and Soifer (2017) and Fergusson, Molina, and Robinson (2020), we take a survey-based approach using data from the second and third rounds

[8] Distance, for instance, has been argued to be related to conflict for other reasons than state capacity (Campante, Do, and Guimaraes 2019), and local road density does not take into account the geographical and infrastructural constraints at other points on the route from the capital to a particular area.

of the Afrobarometer, comprising 25,103 households.[9] From these survey rounds, we select six questions, outlined in Table 1, that tap into the three core dimensions of state capacity mentioned in the previous section: extractive, coercive, and administrative capacity.

The questions labeled 'Law enforcement' and 'Tax enforcement' ask respondents to indicate on an ordinal scale their belief in the ability of the authorities to enforce the law if a person like themselves committed a serious crime or avoided tax payments. These two clearly relate to coercive and extractive capacity, respectively. We include four additional survey items asking respondents to indicate the ease with which a range of services can be obtained. 'Police ability' again taps into the coercive dimension, whereas the remaining three are meant to capture the administrative capacity of the state. First, we include the ease of obtaining identity documents as an indicator of 'Information capacity'. As several scholars have recognized (e.g. Lee and Zhang 2017; Brambor et al. 2020), a core function of the state is the ability to collect information on its citizens (e.g., in order to tax, enforce law, etc.). Finally, we include two items, 'Service provision' and 'School provision', which tap into the presence of key state institutions and services, and thereby slightly expand our operationalization of local state capacity.

As the index comprises several dimensions of state capacity, we opt for a relatively broad understanding of the state capacity concept. At the conceptual level, the selected survey items link to Mann's (1984) concept of infrastructural power, the ability to penetrate civil society and implement decisions throughout the realm. Without being present in the form of either law enforcement, revenue extraction, information gathering, or provision of public services, the state's ability to penetrate civil society is questionable. Furthermore, in the absence of trust in law and tax

[9] In total, there are 46,778 respondents, but 21,675 respondents did not provide complete answers on the survey items and are therefore dropped.

Table 1: *Selected survey items*

Question	Item	Label
How likely do you think it would be that the authorities could enforce the law if a person like yourself: (Very likely; Likely; Not very likely; Not at all likely)	Committed a serious crime?	**Law enforcement**
	Did not pay a tax on some of the income they earned?	**Tax enforcement**
Based on your experience, how easy or difficult is it to obtain the following services: (Very Easy; Easy; Difficult; Very Difficult)	Help from the police	**Police ability**
	An identity document (such as a birth certificate, driver's license, or passport)	**Information capacity**
	Household services (like piped water, electricity, or phone)	**Service provision**
	A place in primary school for a child	**School provision**

enforcement, the state lacks the ability to implement its rules (decisions).[10]

We next conduct a factor analysis to estimate an index of state capacity based on the six survey items shown in Table 1. Compared to choosing just one proxy for state capacity, deriving a weighted average of six survey items by means of factor analysis mitigates noise. Moreover, instead of arbitrarily deciding how the survey items relate to the index, factor analysis optimally assigns weights (factor loadings) such that an item's contribution to the index is based on its correlation with the latent factor.[11] The resulting factor loadings are displayed in Table 2.

We derive an index to proxy state capacity at the enumeration area (EA) level by taking the average of the individual-level perceived state capacity in each EA. In order to further mitigate

[10] The inclusion of institutions for public goods provision arguably makes this a relatively broad conceptualization. However, what we are ultimately interested in, and what we are capturing below, is not the quality of such institutions *per se*, but rather the presence (or absence) of such institutions at the local level, which—all else equal—indicates a generalized territorial reach of the state.

[11] We use a principal factor analysis after having residualized the input variables on potentially confounding variation due to spatial imbalances in the data on age, household head, gender, and the round of the Afrobarometer. We employ the polychoric package in Stata, which makes possible the use of ordinal-scaled variables in the factor analysis.

Table 2: *Rotated factor loadings from Principal Factor Analysis*

Variable	Factor loading	Variable	Factor loading
Law enforcement	0.237	Information capacity	0.252
Tax enforcement	0.233	Service provision	0.249
Police ability	0.222	School provision	0.195

noise, we restrict the sample to include only EAs with five or more respondents with complete answers. The final sample comprises 2,151 EA–year combinations and 1,938 unique EAs in 17 countries.[12] Figure 1 maps the presence of survey EAs and their estimated state capacity index.

We conduct a series of sanity checks of our survey state capacity measure in Appendix Figure A1. The checks are conducted by splitting the EAs into high and low state capacity using the median value as cutoff, and studying how state capacity relates to variables that convey information about the levels of clientelism, crime, infrastructure, and public trust in institutions. The figure reveals, reassuringly, that EA level state capacity is associated with less violence, crime, and clientelism. Moreover, it is positively correlated with the presence of publicly provided infrastructure, namely post offices and piped water, but not with the presence of religious buildings. Correlating our measure with religious buildings serves as a placebo test, since the provision of religious services is generally not considered a key component of state capacity. Lastly we show that while survey state capacity links positively with trust in courts and in the ruling party, it exhibits no relationship with trust in the political opposition. This last check serves as a placebo test, in the sense that it shows trust is not just generally higher in high state capacity areas. Importantly, except for the presence of post offices, all these relationships remain unchanged when controlling for urban-rural differences,

[12] We exclude Cape Verde as a few key predictors are not available for island nations.

fixed effects at the country level, and clustering at the country level.[13]

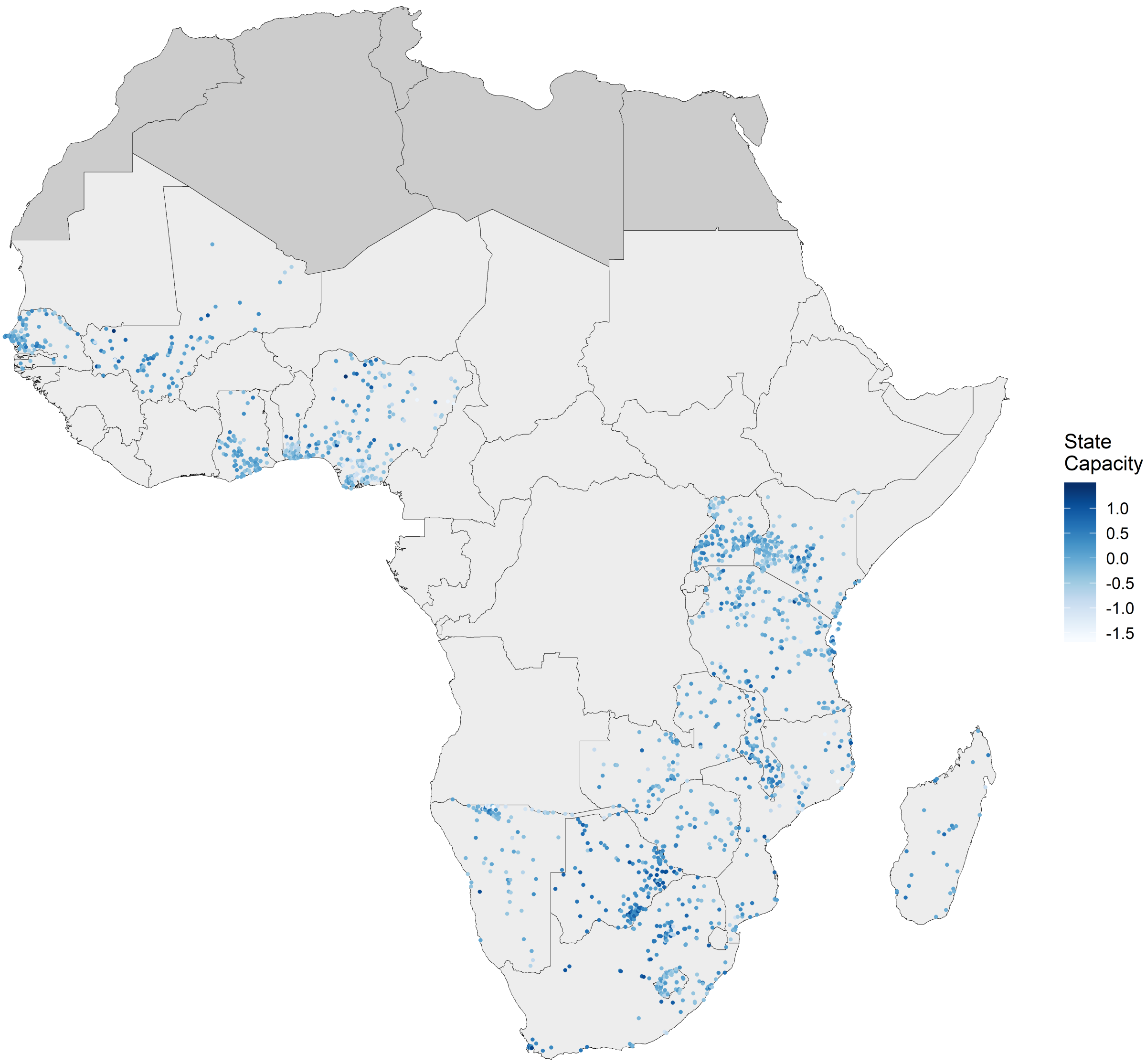


Figure 1: *EA level state capacity*

Notes: The map displays the location and level of state capacity at the EA level.

[13] The regression coefficients and p-values are as follows. Vote buying: regression coefficient $= 0.120$, $N = 1,276$, $p < 0.01$; Feared crime in own home in the past year: regression coefficient $= -0.199$, $N = 2,099$, $p < 0.01$; Family member physically attacked in the past year: regression coefficient $= -0.389$, $N = 2,099$, $p < 0.01$; EA has post office: regression coefficient $= 0.045$, $N = 2,121$, $p = 0.195$; EA has piped water: regression coefficient $= 0.045$, $N = 2,128$, $p < 0.01$; EA has religious building: regression coefficient $= 0.051$, $N = 2,124$, $p = 0.184$; Trust in courts: regression coefficient $= 0.130$, $N = 2,151$, $p < 0.01$; Trust in ruling party: regression coefficient $= 0.149$, $N = 2,151$, $p < 0.01$; Trust in ruling party: regression coefficient $= -0.012$, $N = 2,151$, $p = 0.795$.

B. Predicting the index with machine learning models

Despite Afrobarometer having an exceptional coverage across Africa, for our purposes we need a measure of local state capacity for *all* of Sub-Saharan Africa. The next step is then to develop a procedure for extrapolating values of our measure of the state's territorial reach to areas not covered by the surveys. The fundamental problem is that we do not have a good similar measure of state presence with full continental coverage; however, we do have data on numerous factors likely to either reflect or affect the state's ability to project power across territory. As a starting point, we build on the stylized fact that state power radiates outwards from a political centre (the capital city) and diminishes as one moves away, as Herbst (2000), for instance, has characterized state power in Africa (see also Mann 1984). This idea can also be found in Boulding's concept of a *loss-of-strength gradient* (Boulding 1962). Researchers have attempted to capture this idea by using geodesic distance from the capital to each grid cell or local road density (e.g. Buhaug and Rød 2006; Buhaug 2010). While these are natural choices to measure the *loss-of-strength gradient*, other local factors are arguably related to state capacity, as well as the absolute strength from which power radiates.

We rely on national-level indicators to capture the baseline capacity of the political centre. Then, to capture the sub-national territorial reach of the state, we model the *loss-of-strength gradient* as depending on a set of local factors. Some make it easier to project power, others make it harder. We distinguish between two dimensions: the demand for state presence and the cost of reaching territory. On the demand side, we posit that the state is more willing to invest in state reach (or power) in populous and wealthy areas (where for instance the revenue potential is higher). On the cost side, several factors affect the price of increasing the reach of the state. We focus on geographical, topological, and infrastructural constraints on the state's ability to reach territory. The farther away an area is from the capital city (the political centre), the higher the price for reaching that area. The price is further increased if the state has to cross inaccessible terrain such as mountains and dense forests. Conversely, prior investments in infrastructure such as roads can mitigate the cost of future projection by making it easier to cross territory that is hard to reach.

Another way of understanding this approach is that we develop a composite index of the state's ability to reach territory using data on the demand for state presence and the cost of projecting power into the hinterland. However, instead of relying on a naive additive (and linear) index construction, we use the survey-based measure to estimate the relative (non-parametric) weights of each input to the prediction model.

Specifically, we use a two-step prediction procedure, where we separately predict between- and within-country variation in our measure of state capacity. First, we use a simple linear model to predict between-country variation based on two relevant Worldwide Governance Indicators (WGIs) capturing government effectiveness and regulatory quality (Kaufmann and Kray 2019). Next, we predict state capacity at the EA level, using predictors available at the local 2.5×2.5 arc-minutes level ($\approx 5 \times 5$ kilometers). We demean the outcome variable and the predictors by country to ensure we capture within-country variation.[14] The predictors include, among others, travel time to the capital and urban center, night-time light emissions, population density, and forest cover.[15] While these predictors arguably link with demand for state capacity and cost of reaching the territory through different channels, the causal structures remain unknown and unexplored.[16] Finally, the

[14] We use both absolute and standardized deviations from the country means for the predictors, as it is not given a priori whether the relationship to within-country variation in state capacity is better reflected by absolute or relative measures of the predictors.

[15] Appendix Table A1 presents the full list of predictors and a short description of each.

[16] The causal structures between state capacity and the explanatory variables are trivial for prediction purposes.

between- and within-country predictions are added up.

For the prediction of within-country state capacity variation, we employ an ensemble model consisting a bagging algorithm, several random forest algorithms, and several gradient boosting algorithms.[17] Employing the 'SuperLearner' package in R, each of the underlying algorithms is run on ten folds and the performance of an algorithm is evaluated based on the accuracy of the predicted outcome variable in the holdout sample. The higher the performance of an algorithm, the higher the weight in the ensemble model. The model further accounts for algorithms producing similar predictions by penalizing the weights given to these algorithms. The main advantage of the ensemble model is that predictions are as good as the predictions from any input algorithm (van der Laan and Dudoit 2003; van der Laan, Polley, and Hubbard 2007; van der Laan and Rose 2011). An additional advantage is that the method hedges from cherry picking a specific model and therefore from researcher biases.

The additive separability of between- and within-country state capacity eliminates a problem of overfitting and improves precision for countries with very low state capacity. Including national-level predictors in the ensemble model largely corresponds to including country fixed effects, which would prevent us from extrapolating to non-surveyed countries. Furthermore, countries outside the common support of the survey sample, in terms of baseline country state capacity, would be given the same baseline state capacity as the lowest (or highest) sample country. The disadvantage of imposing additive separability is that we do not allow the association between local predictors and local state capacity to depend on the baseline capacity of the state.

While both national-level government effectiveness and regulatory quality are significantly

[17] Appendix B briefly introduces each of the three non-parametric algorithms. For a more elaborate discussion on regression trees, see James et al. (2013).

correlated with our measure of state capacity,[18] one cannot interpret the effect and magnitude of each explanatory variable in the ensemble model. Appendix Figure A2, however, illustrates the 'importance' of each predictor in one of the algorithms embedded in the ensemble model.[19] In particular, population density and night-time light emissions, followed by travel time to a city and the capital, are found to be important predictors in a random forest algorithm with two predictors in each tree.

The ensemble model employed to explain within-country variation in state capacity gives weight to one random forest algorithm with two predictors in each tree (74 percent) and one gradient boosting model with interaction depth of two and a shrinkage parameter of 0.005 (26 percent).[20] In an out-of-sample testing exercise, Appendix Figure A3 illustrates that the predicted measure of state capacity strongly correlates with the survey measure even within countries (*p-value* < 0.0001). As expected, however, the survey measure exhibits large variation, and the prediction model picks up only a small fraction of this variation. EA observations are based on a median of seven households, which arguably introduces noise. The relatively low R^2 for the holdout sample, when considering within-country variation, may reflect the difficulties in predicting perceived local state capacity using only satellite data as inputs. While this poses a limitation on our prediction strategy, we find it preferable to rely on a data-driven approach to decide on the relative weights of the predictors rather than choosing a single measure or an arbitrary combination. In the following subsection, we further show that the bivariate relationships between predicted state capacity and its predictors exhibit the expected trends.

[18] The t-values from the simple linear regression are 4.5 and 2.9, respectively. Results are available upon request.

[19] With random forest, one may evaluate the importance of each explanatory variable. This is done by separately permuting each variable and deriving the mean squared error (MSE). This MSE is compared to the MSE from the baseline non-permuted model. The higher the increase in MSE from permuting a variable, the more informative it is.

[20] For each algorithm embedded in the ensemble model, we restrict end nodes to have at least 25 observations.

C. Extrapolation to non-surveyed cells

Having calibrated the prediction model, we extrapolate to all of Sub-Saharan Africa. We first extrapolate a simple linear model based on two WGIs to derive a country-level baseline measure of state capacity. Next, we extrapolate the country-demeaned local state capacity measure based on the ensemble model described in the previous subsection. We add up these two measures to derive the final measure of local state capacity at the 2.5×2.5 arc-minutes grid cell level ($\approx 5 \times 5$ kilometers), yielding 1,166,249 predicted state capacity cells in year 2000.[21] Figure 2 maps the predicted measure of local state capacity, where dark blue represents higher predicted state capacity.

The map in Figure 2 reveals a few patterns worth highlighting. First, local state capacity is highly clustered at the national level. Among the countries with low state capacity, we find Liberia, Somalia, Zimbabwe (relative to its neighbors), and The Democratic Republic of the Congo (DRC). In 2000, Liberia had just seen the end to the Second Liberian civil war (Sawyer 2005), while Somalia had been plagued by the Somali Civil War since 1991 (Elmi and Barise 2006). Zimbabwe was experiencing economic and political turmoil following the "Fast Track Land Reform" (Richardson 2007) and DRC was in the middle of the Second Congo War (Hesselbein 2007). On the other end, some of the countries exhibiting higher state capacity are Botswana[22] and Ghana, who are known for their political stability relative to the rest of the region (Naudé 2013). As a mediocre country in the map, Mali might be surprising to some, given its current political instability. In the beginning of the 2000s, however, Mali experienced a stable and peaceful period, depicted by the description of Mali by the American development agency (USAID) as "one of the most enlightened democracies

[21] In the empirical application in Section V, we aggregate the measure to the 0.5×0.5 degree grid cell level in order to link the data to relevant variables. To mitigate the risk of endogeneity bias, we use predetermined state capacity. One can, however, easily extend the prediction exercise to cover more recent years.

[22] See Acemoglu, Johnson, and Robinson (2002) for an in-depth analysis of the political history of Botswana.

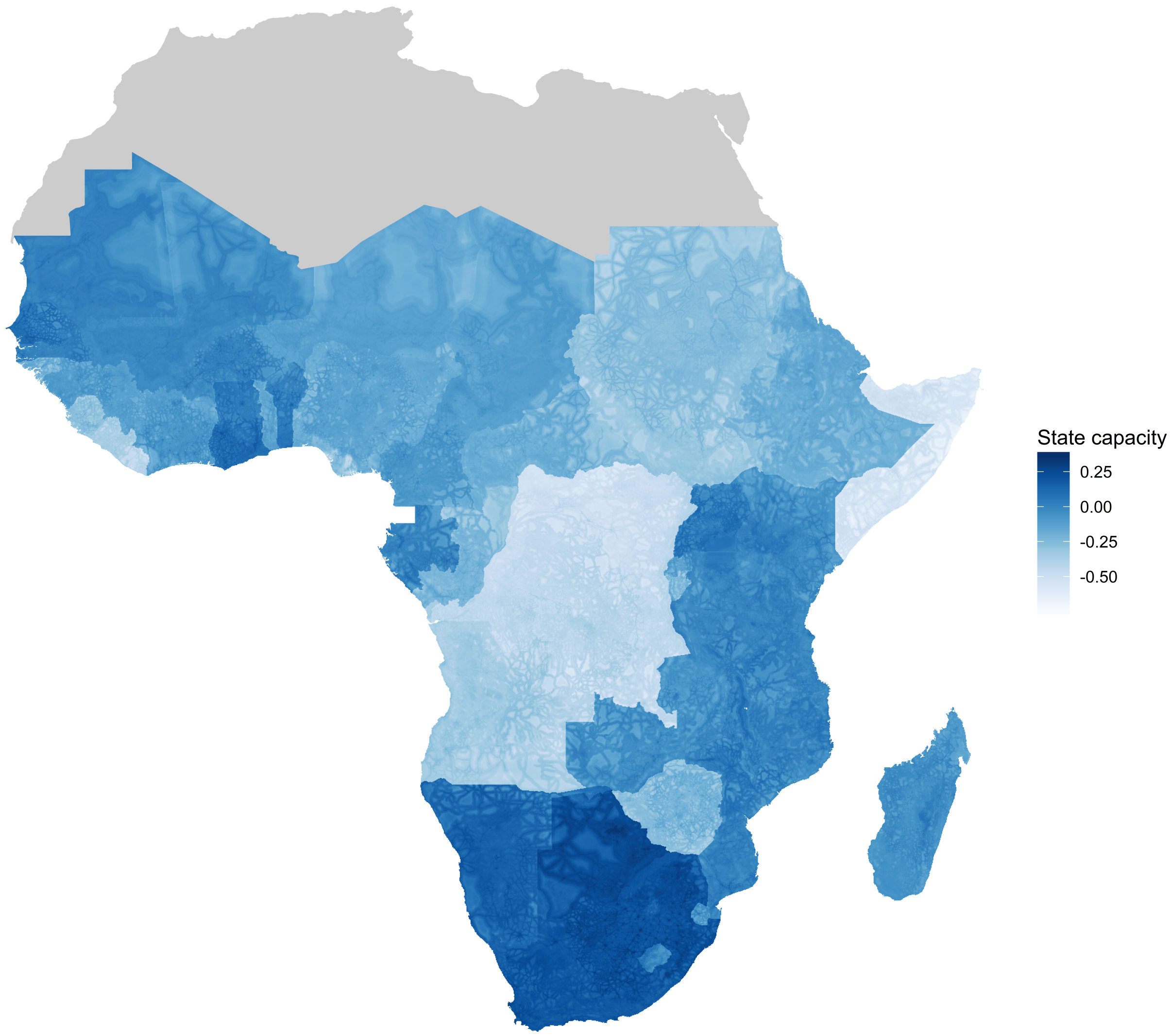


Figure 2: *Predicted measure of local state capacity in year 2000*

Notes: The map illustrates the predicted measure of state capacity in year 2000 at the 2.5×2.5 arc-minutes grid cell level ($\approx 5 \times 5$ kilometers). Grey cells represent excluded countries from North Africa. Equatorial Guinea is excluded due to missing *travel time to the capital*, as the capital is located on an island in the Gulf of Guinea, meaning we cannot predict within-country state capacity.

in all of Africa" (Koenig 2016, p. 114).

Second, while predicted state capacity is clearly clustered at the national level, the map also displays within-country variation. As machine learning predictions trade off interpretability for accuracy, we cannot simply state the effect and magnitude of each explanatory variable. We do, however, notice that there are within-country networks and clusters of either low or high predicted state capacity. These networks and clusters represent topology, infrastructural accessibility, and travel times to the border or economic and political centres. Furthermore, and in particular for high state capacity countries, there are several dark-colored 'dots'. These represent capitals and larger cities, where the presence of the state and its serviceability ought to be highest. Appendix Figure A4 further examines the bivariate within-country relationships between the predicted measure of state capacity and selected explanatory variables. We notice, in line with expectations, that predicted state capacity is negatively correlated with travel time to the capital, travel time per kilometer to the capital, and forest cover. Conversely, predicted state capacity is positively correlated with night-time light emissions and population density. While the bivariate relationship between predicted state capacity and elevation above sea level is fairly constant for cells one standard deviation below and above the country elevation average, cells far below and above the average are predicted to have higher state capacity.[23]

IV. Validating the measure of local state capacity

A. *Ethnic Power Relations*

Ethnic favoritism characterizes the politics of many African countries (Dickens 2018). This is apparent for instance in Kenya, where the ruling ethnic group has been shown to disproportionately allocate road investments to their home regions (Burgess et al. 2015). If this is true for the continent at large, we should observe that territories inhabited by politically powerful groups enjoy higher levels of state capacity. We test this proposition by linking our measure of state capacity with the

[23] One potential limitation of the prediction strategy is uncommon support between survey and extrapolation data. In Section VI, we further discuss how certain outlier areas can be given inaccurate predictions.

GeoEPR dataset (Vogt et al. 2015). The GeoEPR contains geocoded data on the political power of all African ethnic groups that are linked with a specific territory.[24] We code ethnic groups categorized as "Dominating", "Monopoly" or "Senior Partner" in the year 2000 as politically powerful. The data includes 662 ethnic groups, of which 151 are coded as politically powerful. We next spatially merge the EPR-polygons with our state capacity measure at the 2.5×2.5 arc-minutes grid cell level ($\approx 5 \times 5$ kilometers). Finally, we correlate these measures to investigate the relationship between ethnic political power and state capacity.

Figure 3 depicts within-country correlations between political power and our measure of local state capacity in binscatter plots. Political power is positively associated with our measure of local state capacity with a Pearson correlation coefficient of 0.055 ($N = 1,150,206$, $p < 0.01$). Importantly, the relationship between political power and predicted state capacity is positively correlated also when considering only countries that *were not* surveyed by the Afrobarometer (Pearson correlation coefficient = 0.03, $N = 644,440$, $p < 0.01$). This suggests that the predicted measure of local state capacity carries relevance also in countries that did not contribute to the data used in the prediction model. Country fixed effects regression estimates with standard errors clustered at the ethnic group level, however, indicate that the correlations are relatively weak with p-values of 0.12 and 0.34 for the full sample and restricted sample, respectively. This drop in significance is driven by clustering over large spatial areas of ethnic polygons, thereby substantially limiting statistical power.

[24] Political power is a categorical variable taking one of the following labels: "Dominating", "Monopoly", "Senior Partner", "Junior Partner", "Irrelevant", "Discriminated", "Powerless", "Self-excluded", and "State Collapse".

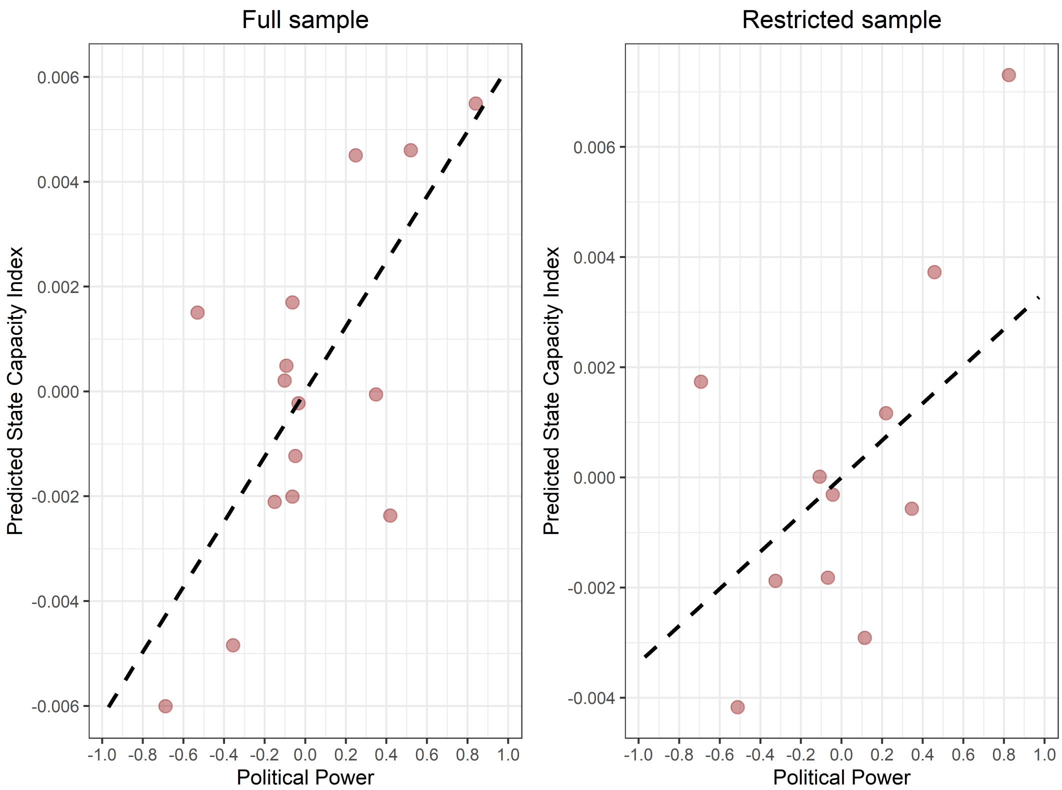


Figure 3: *Correlation between local state capacity and ethnic political power*

Notes: The figure displays bin-scattered relationships between ethnic political power and, respectively, the full sample of state capacity (left) and the restricted sample with only non-surveyed countries (right). The variables are demeaned at the country level in order to depict within-country variation.

B. Pre-colonial centralization

According to numerous scholars, state institutions tend to persist over time (see e.g. Michalopoulos and Papaioannou 2013; Acemoglu, García-Jimeno, and Robinson 2015). Given that this proposition holds true in our data, we should observe a positive correlation between our measure of state capacity and the degree of political centralization in pre-colonial times. Pre-colonial centralization[25] at the society level was coded in the Ethnographic Atlas by Murdock (1959) according to the following categories: “No levels (no political authority beyond community)”, “One level (e.g., petty

[25] The variable is more precisely defined as “Jurisdictional Hierarchy Beyond Local Community”, and contains information on centralization for 843 pre-colonial societies in Africa.

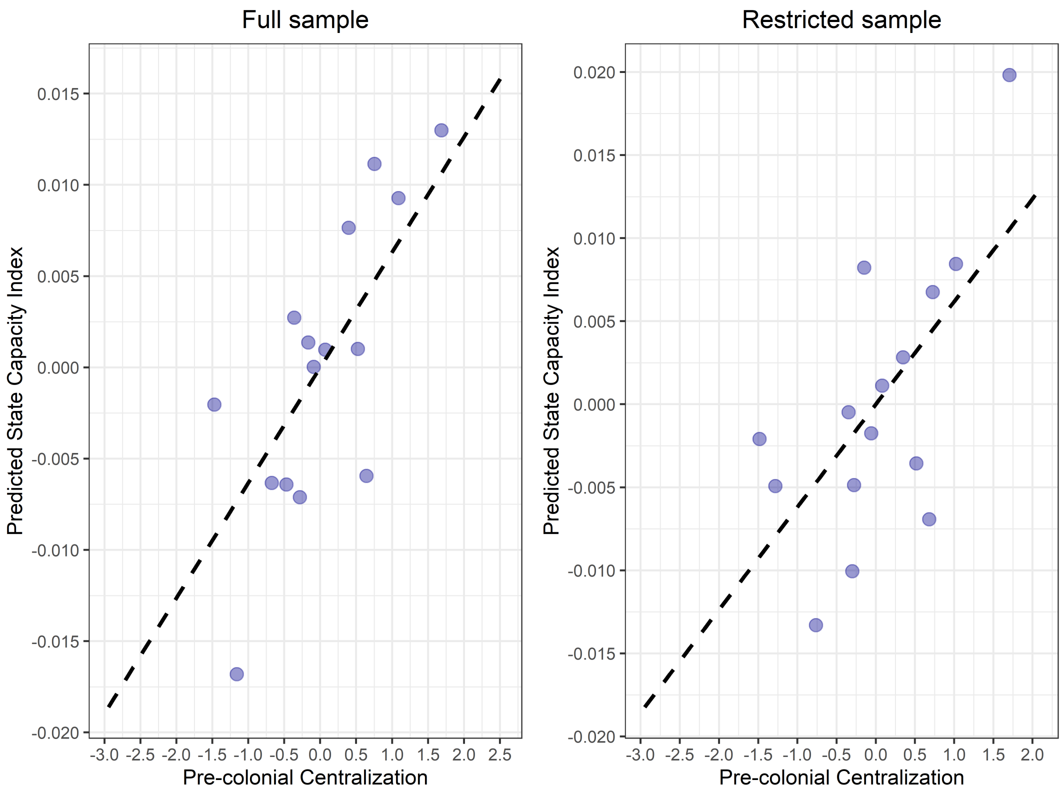


Figure 4: *Correlation between local state capacity and pre-colonial centralization*

Notes: The figure displays bin-scattered relationships between pre-colonial centralization and, respectively, the full sample of state capacity (left) and the restricted sample with only non-surveyed countries (right). The variables are demeaned at the country level in order to depict within-country variation.

chiefdoms)", "Two levels (e.g., larger chiefdoms)", "Three levels (e.g., states)", and "Four levels (e.g., large states)". We create an index of pre-colonial centralization[26] at the 2.5×2.5 arc-minutes grid cell level, by overlaying our measure of local state capacity with geocoded societies from the Ethnographic Atlas (Michalopoulos and Papaioannou 2013).

Figure 4 shows how our measure of state capacity correlates with pre-colonial centralization within countries. The correlation is positive and significant for the full sample (Pearson correlation coefficient = 0.120, $N = 722,945$, $p < 0.01$) and when focusing solely on countries not covered by the Afrobarometer (Pearson correlation coefficient = 0.123, $N = 370,784$, $p < 0.01$). The

[26]The index is used as a continuous variable that ranges from 1 (no political authority beyond community) to 5 (large states).

strong association between predicted state capacity and pre-colonial centralization is confirmed in country fixed effects regressions with standard errors clustered at the pre-colonial society level. The regression coefficient estimates are statistically significant both for the full sample including all cells with a pre-colonial society ($p < 0.01$) and for the restricted sample including only cells in non-surveyed countries ($p < 0.05$). This finding further corroborates the spatial distribution of our state capacity measure.

C. Local vaccination coverage

For the final validation exercise, we compare our predicted values of state capacity to data of local vaccination coverage. Vaccinations are generally considered one of the most cost-effective methods of improving public health. However, vaccine-preventable diseases remain a major cause of mortality (Mosser et al. 2019), especially in low-income countries. Part of the reason is that vaccinations require capable state institutions to administer, record, and inform about vaccinations. Arguably, the presence of the state is a minimal condition for effective vaccination programs. As such, vaccination coverage is a useful proxy for the state's administrative capacity. This has also been recognized by Soifer (2012) who uses vaccination rates to capture the administrative capacity of Latin American states.

Data on vaccination rates are typically only available at the national (sometimes also the administrative unit) level, which limits the usefulness as proxy for local variations in state capacity. However, Mosser et al. (2019) recently published data on local diphteria-pertussis-tetanus (DPT) vaccine coverage of children aged 12–23 months in Africa. The DPT vaccine is included in the WHO's Expanded Programme on Immunisation (EPI), a standardized vaccination programme designed to increase childhood vaccinations throughout the world, and DPT coverage is widely used to measure the performance of routine vaccine delivery systems. Similar to our approach, they use a Bayesian geostatistical model based on survey data and a suite of spatial covariates to estimate annual (2000–2016) local DPT vaccine coverage at a high spatial resolution (5×5 km). We use estimates for completion of three doses of the DPT vaccine (DPT3) in the year 2000 to proxy for local state reach.[27] The spatial resolution matches our predicted measure of state capacity and are therefore readily comparable.[28] In line with the above approach, we further demean the values using the country mean to isolate the within-country variation.

Figure 5 shows the binned scatterplot of DPT3 vaccination coverage and our predicted measure of local state capacity. The left panel includes the full sample of Sub-Saharan Africa, whereas the right panel includes only countries not included in the Afrobarometer survey data. As both figures illustrate, predicted state capacity correlates positively with local DPT3 coverage, further corroborating our measure of local state capacity. The within-country Pearson correlation for the full sample is 0.20 ($N = 923,848$, $p < 0.01$) and for the non-survey sample it is 0.24 ($N = 526,297$, $p < 0.01$). Country fixed effects regression estimates with standard errors clustered at the country level yield significant coefficient estimates both for the full sample of cells ($p < 0.01$) and for the restricted sample including only cells in non-surveyed countries ($p < 0.01$). Thus, the results lend further credibility to our predicted measure of local state capacity.

[27] Completion of the initial DPT vaccination routine requires three doses (DPT3), but not all children complete the vaccine. DPT3 coverage is therefore more demanding in terms of state capacity than the initial dose (DPT1).

[28] The raster grids do not share the same origin, i.e. they do not align perfectly. The DPT3 raster is therefore resampled using the `Raster` package in R, which uses bilinear interpolation to match values to the shifted grid cells.

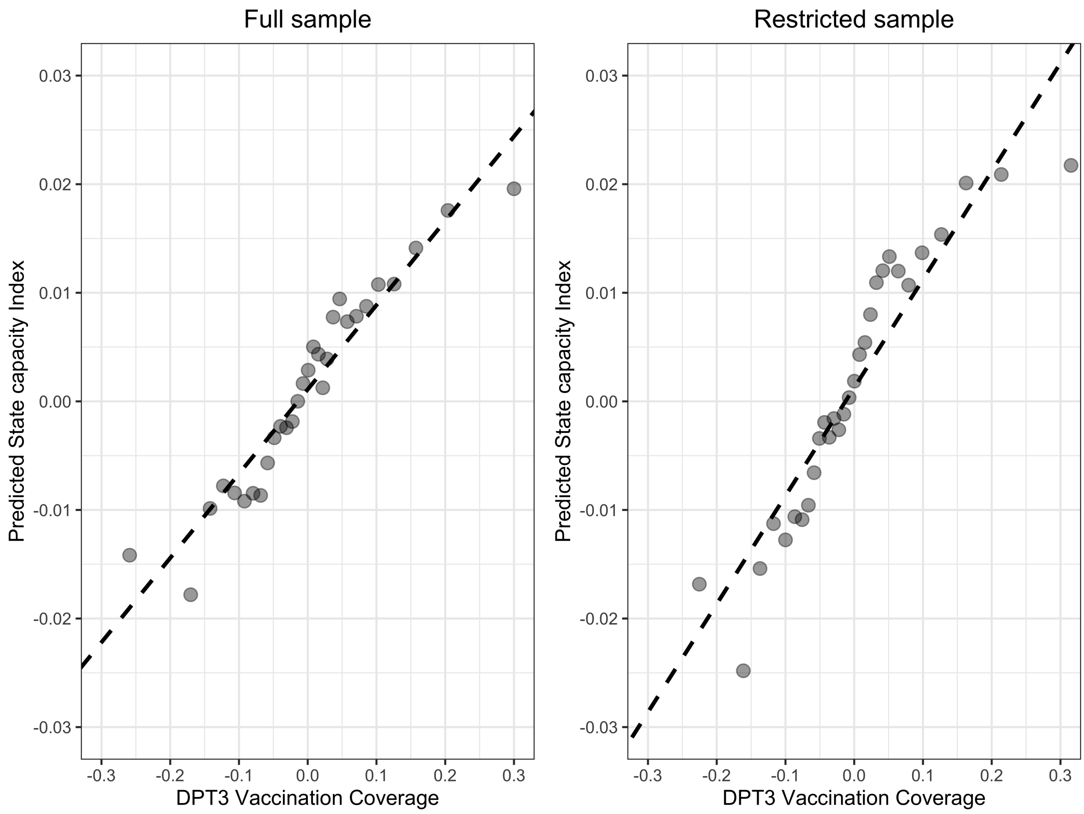


Figure 5: *Correlation between local DPT3 vaccination coverage and predicted state capacity*

Notes: The figure displays bin-scattered relationships between local DPT3 vaccination coverage and predicted state capacity. The left panel shows the scatterplot for the full sample of Sub-Saharan Africa. The right panel includes only Sub-Saharan African countries not included in the Afrobarometer survey. The variables are demeaned at the country level in order to depict within-country variation.

V. State capacity and oil-induced conflict across space and time

A. Background

We now show an application of our local state capacity measure by investigating its potentially moderating role in the oil wealth–conflict relationship. Oil, just like other extractive resources, has been found to increase the risk of conflict in many contexts (Dube and Vargas 2013; Berman et al. 2017; Nwokolo 2018; Nordvik 2019). One mechanism that has been proposed to explain this link is the so-called greed channel, which posits that the concentrated and seizable value of oil allows armed groups to engage in both extortion and resource grab of oil assets. A second mechanism, the grievances channel, emphasizes the perceived injustices of the unequal distribution of oil rents, which may lead to unrest in the form of protests, riots or sabotage directed towards the oil industry. Lastly, oil rents can impact the risk of conflict indirectly, through what Bazzi and Blattman (2014) label the state capacity effect. The direction of the state capacity effect is a priori ambiguous. On the one hand, oil wealth may strengthen states' ability to repress conflict, by generating revenues that can be used either to buy off potential insurgents or to invest in repressive capacity (which deter uprisings) (Ross 2012). On the other hand, oil rents have frequently been linked with corruption and extortion, which may weaken states' strength both by deteriorating capacity and and by eroding public confidence in state institutions.

While most studies find that increased oil wealth leads to higher risk of conflict, a few studies suggest that oil wealth has no impact and that it can even decrease the risk of conflict. Bazzi and Blattman (2014) use cross-country panel data to estimate the association between changes in commodity prices and the onset, intensity, and ending of armed conflicts. The authors find no effect of current oil prices and a negative effect of lagged oil prices on conflict incidence. In a panel fixed effects model with countries as the unit of observation, Cotet and Tsui (2013) find no significant effect of oil discovery on conflict onset, conditional on oil exploration.

The somewhat inconsistent findings in the oil–conflict literature is partly due to a failure to measure relevant conflict events at relevant geographical units. By focusing on all types of conflicts, and at coarse levels of spatial aggregation (often at the country level), many previous empirical studies present estimates that suffer from attenuation bias. Another reason why the estimated effect of oil wealth varies between studies may simply reflect that the effect of oil wealth on conflict depends on the spatio-temporal context. Contextual factors such as country institutions

(Buhaug 2010; Berman et al. 2017), companies Corporate Social Responsibility (CSR) commitments (Berman et al. 2017), the spatial distribution of resources (Morelli and Rohner 2015; Lessmann and Steinkraus 2019) and redistributive government transfers (Justino 2011; Justino and Martorano 2018; Cordella and Onder 2020) have been found to moderate the risk of conflict.

In this study, we investigate the role of local state capacity as a moderating factor in the oil–conflict relationship. We hypothesize that local state capacity decreases the risk of oil-induced conflict by reducing the incentives of confrontation. Higher levels of state capacity implies higher extractive[29], coercive[30], and administrative capacity.[31] These dimensions of state capacity, although clearly distinct from each other, jointly determine states' ability to avoid and, if need be, repress conflict. The extractive capacity determines states' ability to invest in coercive and administrative capacity. In turn, coercive capacity can disincentivize conflict as it decreases the likelihood of capturing oil rents (Bazzi and Blattman 2014). Administrative capacity, on the other hand, determines the scope for accommodative measures which may reduce the rationale for grievances (Hendrix 2010).

A rich literature on the role of state capacity to suppress conflict supports this underlying hypothesis. Especially for intra-state violence, the lack of state capacity to deter and fend off rebel insurgencies has often been emphasized as a key explanation for civil war onset in developing countries (e.g. Fearon and Laitin 2003; DeRouen and Sobek 2004; Fearon 2005; Buhaug and Rød 2006; Lacina 2006; Fjelde and De Soysa 2009; Braithwaite 2010; Gleditsch and Ruggeri 2010; Hendrix 2011). For example, Fearon and Laitin (2003) argue that insurgencies are more likely to

[29] Our survey state capacity index contains this dimension in the survey item Tax enforcement.

[30] Captured by the survey items Law enforcement and Police ability in the survey state capacity measure.

[31] This dimension is represented in our survey state capacity measure by the survey items Information capacity, Service provision and School provision.

form in areas where the state lacks the capacity to effectively monitor. However, due to the lack of readily available measures of local state capacity, empirical studies have generally relied on state-level variables of state capacity, despite acknowledging that the local level is more appropriate (e.g., Berman et al. 2017). Moreover, due to the circular nature of the relationship between state capacity and armed conflict, establishing the role of state capacity in mitigating the risk of conflict has proven challenging.[32]

B. Empirical approach

In order to investigate how state capacity conditions the risk of conflict, we create a panel dataset on state capacity, oil wealth, and oil-induced conflicts for Sub-Saharan Africa in 2001-2019. First, we aggregate our predicted state capacity index in year 2000 to 0.5×0.5 degree grid cells ($\approx 55 \times 55$ kilometers). We link this with data on oil deposits in year 2000, and yearly data on oil prices and oil-induced conflicts in 2001-2019. The interaction between oil deposits in year 2000 and the oil price constitutes our exogenous spatio-temporal measure of oil wealth. Higher oil prices means that oil deposits become relatively more valuable, which could activate the greed or grievances channel and trigger conflict. The panel structure allows us to estimate fixed effects models to explore how changes in oil wealth impact the risk of conflict in oil cells conditional on local state capacity, holding constant potentially confounding variation at both the cell and the year level.

C. Data

Our measure of local state capacity in the empirical analysis is the measure we construct and present in Section III. To ease the interpretation of the results in the empirical analyses, we standardize the

[32] As emphasized by Thies (2010), there is also a causal link from civil war to state capacity.

state capacity measure and add the minimum value such that the lowest possible value is 0 and a unit equals one standard deviation.

The data on geocoded petroleum deposits is obtained from PRIO-GRID data 2.0 (Lujala, Rod, and Thieme 2007). Since conflict, and even the risk of conflict, could affect hydrocarbon exploration and production decisions, endogeneity is a first-order concern that we need to circumvent. In order to do so, we define cells as oil cells if they contained onshore oil resources in year 2000.[33] Next, we create a time- and space-varying measure of oil wealth by interacting the spatial indicator of oil deposits with the log of the average yearly international oil price.[34]

To measure oil-related conflict events at the grid cell level, we obtain geocoded conflict data from the Armed Conflict Location and Event dataset (ACLED) (Raleigh et al. 2010). ACLED provides information on the location and exact date as well as the nature of conflict events. The data builds on information compiled by a range of stakeholders, including news agencies, researchers, and humanitarian organizations. While the ACLED data could contain measurement error due to inaccurate reporting, and the level of misreporting likely could differ between high and low state capacity regions, we alleviate concerns of reporting inconsistencies biasing our results by including both year and cell fixed effects.

We exploit the detailed information provided by ACLED to create measures of conflict that are closely linked with our hypothesis. First, we classify oil-related conflict events based on qualitative information on the recorded events. Specifically, we label an event ‘oil-related’ if it contains the words “oil” or “petroleum”.[35] By applying a narrow definition of oil conflict, we are able to estimate

[33] We combine the variables *petroleum_s* and *petroleum_y* in year 2000 from PRIO-GRID 2.0 into one indicator, since the combination of both constructs a measure on “known petroleum deposits in year 2000”, which is the most suitable definition of an oil cell for our purposes.

[34] The oil price is measured as an equally weighed average spot price of Brent, Dubai and West Texas Intermediate crude oil prices. The data is obtained from www.indexmundi.com.

[35] We further label two special cases that include the strings “right-to-oil” and “gas-and-oil” as oil-related, and we label

with better precision the role of state capacity in mitigating the risk of oil-induced conflict. Next, we define two subsets of oil conflicts to allow for suggestive tests of the mechanisms linking oil wealth to conflict, and to study the moderating role of state capacity in the respective channels. First, we focus on violent oil-related conflict events,[36] which potentially relate to contests over oil rents, in order to test for the greed channel. Second, by using as our dependent variable oil events classified by ACLED as either riots or protests, we investigate civilian actions related to oil grievances.

The ACLED dataset for Sub-Saharan Africa from January 1st 2001 to December 31st 2019 includes 767 geocoded oil-related conflict events. Out of these, 404 are coded as "violent events" and 313 are labeled "demonstrations".[37] The mere fact that there *are* oil-related conflict events provides an indication of oil wealth impacting the risk of conflict. In Figure 6, we map oil deposits, state capacity, and conflict events at the cell level. The spatial location of the majority of oil-related conflict events are in, or in close proximity to, oil cells, which provides a simple sanity check of our coding of oil conflicts.

D. Empirical methodology

In order to abstract from time- and space-confounding variation, we include both year and cell fixed effects in all regressions, such that we essentially estimate difference-in-differences models. The key identifying assumption is that oil price shocks are orthogonal to conflict events in our sample.[38] Since oil production in Sub-Saharan Africa constituted only 7.3 percent of world output in 2008, and 5.6 percent in 2018 (Dudley 2019), we deem this to be a reasonable assumption. Our empirical

conflict events with the word combinations "palm oil", "cooking oil" and "oil mill" as non oil-related.

[36] This subset of conflict events encompasses battles, violence against civilians, and explosions/remote violence.

[37] The remaining 60 oil-related events encompass non-violent actions included in the ACLED data.

[38] If the international oil price responds to conflict events in the oil-producing cells, simultaneity bias arise.

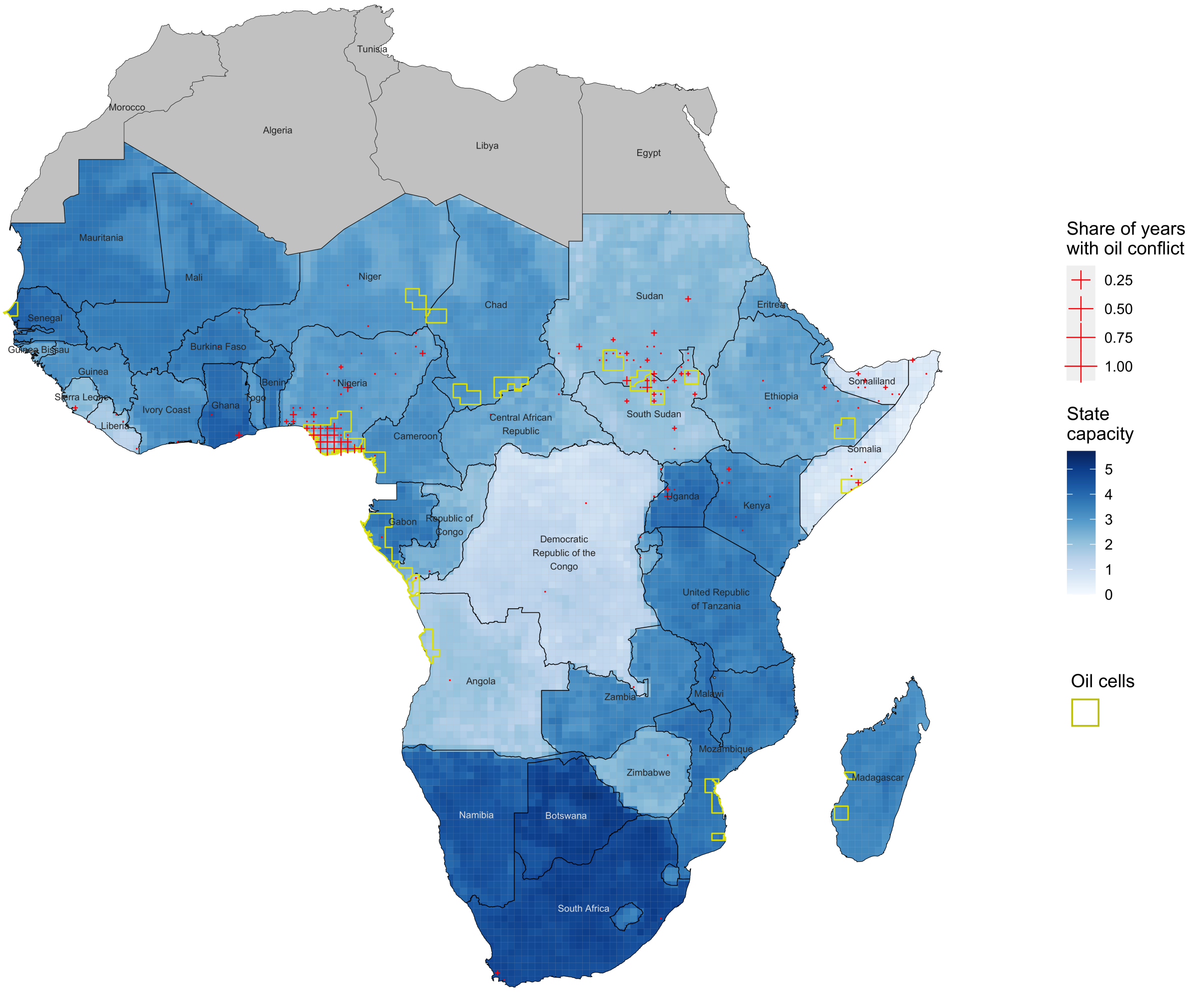


Figure 6: *State capacity, oil cells, and oil conflict events in Sub-Saharan Africa 2001-2019*

Notes: The map plots oil deposits and oil conflict data onto our grid cell state capacity measure. The state capacity index is standardized and re-scaled to have a minimum of zero for the empirical analysis.

estimations all build on variants of the following model:

$$\begin{aligned} Oil\_Conflict_{it} =& \beta_0 + \beta_1 Oilprice_t \times Oilcell_i + \beta_2 Oilprice_t \times SC_i + \\ & \beta_3 Oilprice_t \times Oilcell_i \times SC_i + \gamma_i + \delta_t + \varepsilon_{it}, \end{aligned} \tag{1}$$

where the outcome variable represents whether there was an oil-related conflict event in cell i in year t; *Oilprice* is the international oil price; *Oilcell* indicates whether the cell had known oil deposits in year 2000; *SC* is the predicted measure of state capacity; γ_i and δ_t are cell and year fixed effects, respectively; and ε is the error term. Year fixed effects capture the potential effect of the *Oilprice* in cells without known oil deposits, whereas the interaction with *Oilcell* captures the additional effect for cells with known oil deposits.[39] The triple interaction term captures whether the effect from the interaction between *Oilprice* and *Oilcell* is conditional on state capacity.

In an extension of the baseline model, we consider the contagious nature of conflict across space and time. In oil cells, the likelihood of oil conflict is 5.2 percent if there was no oil conflict in the past year, but 17.1 percent if there was. Similarly, while the risk of oil conflict in oil cells is 5.3 percent in a year if none of the neighboring cells experience oil conflict that year, the risk of conflict is 8.4 percent if at least one neighbor recorded an oil conflict. We account for the auto-correlation and spatial spillovers by including temporal and spatial lags of the outcome variable, respectively.[40] To recognize that standard errors may correlate across space and time, we cluster them at the country level. This specification allows standard errors to correlate both across large territories and over the entire sample period, an approach that generally increases the estimated standard errors relative to clustering of standard errors at more limited spatial or temporal levels.[41]

We further consider a continuous measure of "oil deposits" by instead of using a dummy indicator for oil cell, we use the inverse (square root) distance to an oil cell (labelled *oil proximity*). Lastly, we present a placebo test by substituting current with lead oil prices. While auto-correlation in prices may cause the lead prices to pick up an effect associated with current prices, the precision

[39] While the coding of the outcome variable makes it unlikely for cells without oil deposits to be affected by a change in the oil price, three scenarios make it feasible. First, oil is found in Sub-Saharan Africa also *after* the year 2000. Second, the purchasing power of consumers is affected by oil price fluctuations. Third, conflict in oil cells is contagious and may spillover into neighboring cells.

[40] This approach introduces dynamic panel bias, meaning the coefficient estimate on the lagged dependent variable is biased downward. Despite having 19 time periods and the bias being an issue mostly in shorter time panels, one should keep this in mind when interpreting the coefficient estimate.

[41] Indeed, the estimated standard errors are significantly smaller if we cluster instead at the cell or at the country–year level (results not reported).

ought to be smaller, and hence we expect significance to drop.

E. Results

We now turn to the empirical link between oil wealth and conflict, and the role of local state capacity as a moderating factor in this relationship. First, we show in Figure 7 how the risk of oil-related conflict events differ for high and low state capacity cells. The risks of oil-induced conflict, violent oil events, and oil demonstrations, are substantially higher in oil cells with low state capacity compared to cells with high state capacity. The difference is significant and of a magnitude of 10. The effect remains significant when including country fixed effects, showing that our measure of local state capacity matters also for the spatial distribution of oil conflicts *within* countries (*p-value* < 0.05). While these results indicate a moderating role of state capacity, as low state capacity oil cells appear much more prone to oil-related conflicts compared to high state capacity oil cells, the differences in levels are not necessarily causal. In order to test if oil wealth induces oil-related conflict, and if this effect is conditional on the level of state capacity, we estimate panel regressions with our indicator variables for oil conflict as the outcome variables, and oil price, oil deposits, and state capacity as the explanatory variables. The results are displayed in Table 3.

Table 3 shows the impact of international oil prices on oil-related conflict in cells with oil deposits relative to cells without oil deposits. In Panel A columns 1 and 2, we investigate how increased oil prices differentially impact the risk of conflict in oil cells, separately for low and high levels of state capacity. The results depict an interesting pattern. While the risk of oil-related conflict increases with oil prices in low state capacity cells, the impact in high state capacity oil cells is negative. In column 3 we show the effect for the full sample, where state capacity is added as a

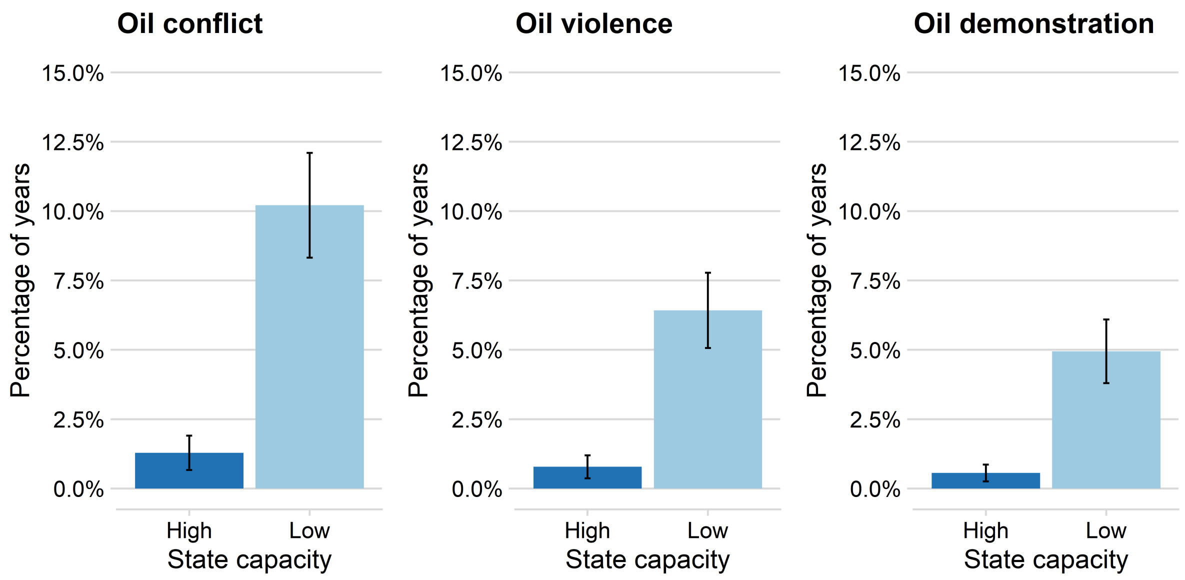


Figure 7: *The frequency of oil-related conflicts and battles in oil cells by level of state capacity*

Notes: The figure plots the share of years that oil cells experience oil-related conflicts (left), oil-related violent events (middle), and oil-related demonstrations (right). In all three plots, oil cells are divided into high and low state capacity based on the median predicted state capacity in year 2000.

continuous variable in a triple interaction term with oil cell and oil prices. The results corroborate the insights from columns 1 and 2. If oil prices increase by 25 percent year to year, the likelihood of oil-related conflict increases by 0.5 percentage points ($0.053 \times log(1.25)$) for oil cells with the lowest level of state capacity (where state capacity $= 0$). This is a relatively large increase, since the share of years with oil-related conflict in oil cells is 5.9 percent. This effect is muted in cells with higher levels of state capacity. For a cell with state capacity at the 75th percentile (3.6 standard deviations higher than the minimum state capacity), the increased risk of oil-related conflict in oil cells due to a 25 percent year to year price hike is estimated at 0.1 percentage points.[42]

Next, we investigate whether the effect of oil price shocks differ for violent oil events and oil-related demonstrations. In columns 4-6, we display the regression models with a dummy indicator for oil-related violent events, e.g. battles and remote violence, as the dependent variable. The effects

[42] $0.053 \times log(1.25) + (-0.00045 \times log(1.25)) \times 3.6 + (-0.011 \times log(1.25)) \times 3.6$.

Table 3: *The conditional effect of oil price fluctuations on the likelihood of oil conflict 2001–2019*

	Oil event (1)–(3)			Oil violence (4)–(6)			Oil demonstration (7)–(9)		
	Low SC (1)	High SC (2)	Interaction (3)	Low SC (4)	High SC (5)	Interaction (6)	Low SC (7)	High SC (8)	Interaction (9)
Panel A:									
Oil price × Oil cell	0.039**	-0.0014*	0.053**	0.025**	0.00003	0.038*	0.014*	-0.0014*	0.016*
	(0.017)	(0.00076)	(0.024)	(0.011)	(0.00015)	(0.020)	(0.0068)	(0.00068)	(0.0080)
Oil price × State capacity			-0.00045			-0.00039*			-0.00005
			(0.00031)			(0.00020)			(0.00013)
Oil price × State capacity × Oil cell			-0.011*			-0.0085*			-0.0028
			(0.0058)			(0.0050)			(0.0022)
R-squared	.317	.128	.292	.251	.065	.244	.234	.130	.211
Observations	80,142	79,591	159,733	80,142	79,591	159,733	80,142	79,591	159,733
Panel B:									
Oil price × Oil cell	0.035**	-0.0013*	0.048**	0.022**	0.00005	0.035*	0.011**	-0.0014*	0.012**
	(0.015)	(0.00075)	(0.022)	(0.0095)	(0.00021)	(0.018)	(0.0051)	(0.00067)	(0.0059)
Oil price × State capacity			-0.00036			-0.00030*			-0.00004
			(0.00024)			(0.00015)			(0.00010)
Oil price × State capacity × Oil cell			-0.0099*			-0.0079*			-0.0019
			(0.0053)			(0.0046)			(0.0019)
Dependent variable temporal lag	0.078***	-0.033	0.062***	0.052**	-0.070***	0.047**	0.034	-0.036	0.019
	(0.014)	(0.028)	(0.015)	(0.022)	(0.0086)	(0.020)	(0.029)	(0.037)	(0.035)
Dependent variable spatial lag	0.028***	0.0022	0.022***	0.027***	0.0058	0.025***	0.054***	-0.0019	0.036***
	(0.0065)	(0.0062)	(0.0058)	(0.0078)	(0.0091)	(0.0074)	(0.011)	(0.0087)	(0.012)
R-squared	.322	.129	.295	.253	.069	.246	.235	.131	.211
Observations	80,142	79,591	159,733	80,142	79,591	159,733	80,142	79,591	159,733
Cell & year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable mean	.0044	.00067	.0025	.0027	.00011	.0014	.0019	.00053	.0012
Dependent variable std. dev.	.066	.026	.050	.052	.011	.038	.043	.023	.035

Notes: The dependent variables in columns (1)-(3), (4)-(6), and (7)-(9) are cell-year indicator variables for 'oil-related conflict event', 'violent oil-related conflict event' and 'oil-related demonstration', respectively. Columns (1), (4), and (7) restrict the sample to cells below the median state capacity, whereas columns (2), (5), and (8) restrict the sample to cells above median state capacity. Columns (3), (6), and (9) display results from models including the continuous measure of local state capacity in a triple interaction with oil cell and the oil price. *Oil price* is the log of the average oil price for each year, *Oil cell* is an indicator variable on known oil deposits in year 2000, and *State capacity* refers to our measure of local state capacity in year 2000. Standard errors clustered at the country level are in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

mimic the ones obtained when using the broad measure of oil conflicts: oil price hikes increase the risk of violent oil events in low state capacity cells, but the risk is significantly moderated for higher levels of state capacity. In fact, for oil cells with higher than median state capacity, there is virtually no effect of oil price shocks on the likelihood of experiencing a violent oil event. A similar pattern is apparent also for oil demonstrations. Increased oil prices positively affect the risk of oil-related demonstrations in low state capacity oil cells, whereas in high state capacity cells the effect is reversed. Although the triple interaction is insignificant for this subset of oil-related conflict events, the gist from the panel models is that oil price shocks increase the risk of oil-related conflicts, but only in oil cells with low local state capacity in the year 2000. The results support both greed and grievances as underlying mechanisms linking oil price shocks to conflict, and suggest a role of local state capacity in counteracting both channels.

In Panel B we explore the robustness of the baseline findings to accounting for both time and spatial spillovers. As we saw in Subsection D, oil conflicts tend to persist over time, and the risk of oil conflict is associated with oil conflicts in the adjacent area.[43] As expected, the year and space lags are positively correlated with the dependent variable in columns 1, 4, and 7, where we consider only low state capacity cells. Strikingly, this is not the case for high state capacity cells (columns 2, 5 and 8), which indicates that local state capacity may not only matter for the occurrence of oil conflicts, it seems to also mitigate the risk of conflict persisting as well as spreading across territories. Importantly, when we include the year and spatial lags, the baseline results remain qualitatively unchanged. Even though this exercise arguably entails "over-controlling", the stability of the coefficient estimates is reassuring.

[43] Since conflict in neighboring cells may be jointly determined, this positive correlation does not necessarily imply a causal link.

We now consider a few alterations to our preferred specification presented above. First, we present results in Appendix Table A2 from using lead and lagged instead of current oil prices in the regression models. Whereas the lagged oil price conveys information that may matter for present conflict, for instance if greed and grievance effects build up over time, the only way that lead prices could affect conflict is through its correlation with present oil prices. We therefore anticipate that using the lagged oil prices should produce results that align with our baseline specifications, while employing lead oil prices should result in substantially weakened effects. As expected, Appendix Table A2 shows no significant effect of the triple interaction term when lead prices are used. Lagged oil prices, on the other hand, display similar effects as current oil prices on the risk of oil conflict in oil cells.

Next, we substitute our measure of oil deposits from a dummy indicator to the inverse (square root) distance to oil cells. By operationalizing oil deposits as a value that dissipates with distance to oil, we capture potential spillovers in oil wealth, e.g. due to transport networks. This exercise, shown in Appendix Table A3, does not impact the results substantially, but improves the precision of the estimates somewhat. Lastly, we consider an even more restrictive model than the baseline, by including country–year fixed effects, instead of only year fixed effects. Even this alteration to our baseline model does not impact the estimates appreciably, further supporting the insights on state capacity as a moderating factor.

VI. Discussion

A number of limitations, both with regards to the construction of our local state capacity measure, and of the empirical application of the measure as a moderating factor in the oil–conflict relationship, are worth highlighting. First, the within-country prediction model suffers from uncommon support between the extrapolation areas and the areas with survey data. Since the machine learning algorithm divides areas into "groups" of predicted state capacity based on the range of input variables, it will inevitably lump together areas that are not entirely similar. This is of particular concern for areas in the extrapolation data that are not represented in the survey data. For instance, extreme altitude areas, which are not surveyed by the Afrobarometer, but still feature in the data that we extrapolate to, are placed in the same groups as the highest-altitude areas in the survey data. While this arguably should lower precision in some extrapolated areas, we show in Section IV that our measure of local state capacity links with the variables we use for validation also in countries not covered by the Afrobarometer.

Second, the prediction strategy relies on the assumption of additive separability between national and sub-national predictors. That is, the strategy does not account for potential interactions between predictors at the local and national levels. While this is arguably a strong assumption, it is also necessary, as including national-level predictors in the machine learning model essentially corresponds to including country fixed effects, which increases the risk of overfitting. Separating our prediction model in two steps allows us to overcome this issue, and to instead capture universal associations in the within-country prediction model.

Third, our measure of local state capacity is to a large extent driven by differences in country-level predictors.[44] While the relative weights of within- and between-country variations in our state capacity measure are optimized through our prediction models, we cannot rule out potential biases due to omitted variables for either the national or sub-national predictors. We can, however, show that the within-country variation that we do capture conveys important information. In all validation

[44] Country differences account for 93 percent of the variation in our measure of local state capacity.

checks in Section IV, we show that our measure of local state capacity matters for *within* country variation. Moreover, if we in the oil-conflict panel regressions use instead the country mean of our local state capacity measure, the conditioning effect of state capacity is somewhat smaller in magnitude and insignificantly estimated.[45]

Fourth, our prediction model incorporates determinants that may moderate the risk of oil-induced conflict for reasons other than local state capacity. For instance, we are not able to abstract from the potentially confounding covariance between local state capacity and local income levels, and as a consequence we cannot rule out alternative explanations for the results presented in Section V.

Fifth, as a time-invariant measure of state capacity based on data from the year 2000, our measure does not account for the changes in state capacity that occurred during the study period. While constructing a panel dataset of local state capacity is feasible, the usefulness of such a measure, at least for the present purposes, is not obvious. The improved precision of local state capacity would come at the price of endogenous time-variation, an issue we mitigate by using a pre-determined variable of state capacity.

Sixth, measuring only oil-induced conflicts entails both methodological advantages and disadvantages. The increased precision enables us to reduce noise in the estimates, but also reduces the number of data points that convey information.[46] Furthermore, since oil conflicts by construction occur mainly in oil regions, the effect of oil prices on oil conflicts may be biased due to an underlying positive time trend in both the oil price and oil conflicts. This, however, does not impact the main finding, namely the moderating role of local state capacity. We further corroborate this finding by

[45] Oil conflict: regression coefficient$= -0.0090$, $N = 159,733$, $p = 0.18$; Violent oil conflict: regression coefficient$= -0.0047$, $N = 159,733$, $p = 0.33$; Oil demonstration: regression coefficient$= -0.0035$, $N = 159,733$, $p = 0.35$.

[46] Each year, 12.5 percent of cells in Sub-Saharan Africa experienced some type of conflict event, but only 0.25 percent experienced oil-related conflict. However, in oil cells, oil-related conflicts constitute more than 4 percent of all conflict events.

means of panel regressions with cell fixed effects where we exclude cells that do not have any oil deposits. These regressions confirm the moderating role of local state capacity.[47]

VII. Conclusion

The conceptualization of state capacity is ever contested and to date there exists no universal definition of the term. While scholars disagree about what exactly state capacity encompasses, there is no controversy regarding the centrality of state capacity in shaping economic, social, and political life. Variability in states' ability to project power across territories, especially for developing countries, highlight the need to measure state capacity at a local level. But due to the data scarce nature of these contexts, doing so has proven challenging.

In this paper, we have provided a novel methodology to measure state capacity at a spatially disaggregated level, and corroborated the relevance of the measure in several exercises. In order to operationalize state capacity, we considered three important dimensions: the state's extractive, coercive, and administrative capacity. We trained a tree-based prediction model on geocoded survey data on state performance, using publicly available satellite data to measure within-country variation in state capacity. We extrapolated the resulting measure of state capacity to all 2.5×2.5 arc-minutes grid cells in Sub-Saharan Africa.

In order to validate the spatial distribution of our novel measure of local state capacity, we correlated it with variables that have been argued to convey information about state capacity in previous work, namely political power of ethnic groups, pre-colonial centralization, and vaccination coverage. Using only *within* country variation, we showed with this exercise that our measure

[47] Oil conflict: Oil prices $\times$ State capacity interaction coefficient$= -0.011$, $N = 3,686$, $p < 0.1$; Violent oil conflict: Oil prices $\times$ State capacity interaction coefficient$= -0.009$, $N = 3,686$, $p < 0.1$; Oil demonstration: Oil prices $\times$ State capacity interaction coefficient$= -0.003$, $N = 3,686$, $p = 0.219$.

of local state capacity linked positively with these factors that, for different reasons, indicate the power of the state at the sub-national level. Finally, we employed the measure of state capacity in fixed effects panel models on the relationship between oil wealth and conflict, and documented a suggestive moderating effect of local state capacity.

While our approach entails certain disadvantages discussed in Section VI, this paper provides a first account of a data-driven solution using machine learning to measure local state capacity. This methodology allows us to overcome the data scarcity inherent to this literature. The disaggregated data on state capacity will be made publicly available in order to promote empirical research on the determinants and consequences of state capacity. Future research employing similar methodologies would benefit from using more data in the training model and identifying additional relevant predictors. Moreover, the flexible nature of the framework presented in this paper allows for various conceptualizations of state capacity including disentangling the concept into its constitutive dimensions. Researchers interested in specific dimensions of state capacity – e.g., extractive, coercive, or administrative capacity – could easily adapt the approach to their specific needs by simply augmenting the index used for training the prediction model. Both the data and methodology thus add to a burgeoning literature on the role of the state, and provide a new approach to the study and measurement of the state at the level at which most citizens actually experience and interact with the state.

REFERENCES

Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *American Economic Review* 105 (8): 2364–2409.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2002. *An African Success Story: Botswana.* SSRN SCHOLARLY PAPER ID 304100. Rochester, NY: Social Science Research Network.

Acemoglu, Daron, and James Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity and Poverty.* New York: Crown.

Afrobarometer Data. 2004, 2005. [Benin, Botswana, Ghana, Kenya, Lesotho, Madagascar, Malawi, Mali, Mozambique, Namibia, Nigeria, Senegal, South Africa, Tanzania, Uganda, Zambia, Zimbabwe], [Rounds 2 and 3], [Years 2004 and 2005], Available at http://www.afrobarometer.org.

Balk, D. L., U. Deichmann, G. Yetman, F. Pozzi, S. I. Hay, and A. Nelson. 2006. "Determining Global Population Distribution: Methods, Applications and Data." In *Advances in Parasitology,* edited by Simon I. Hay, Alastair Graham, and David J. Rogers, 62:119–156. Global Mapping of Infectious Diseases: Methods, Examples and Emerging Applications. Academic Press.

Bazzi, Samuel, and Christopher Blattman. 2014. "Economic Shocks and Conflict: Evidence from Commodity Prices." *American Economic Journal: Macroeconomics* 6 (4): 1–38.

Bergquist, Lauren F., Benjamin Faber, Thibault Fally, Matthias Hoelzlein, Edward Miguel, and Andres Rodriguez-Clare. 2019. *Scaling Agricultural Policy Interventions: Theory and Evidence from Uganda.* Working Paper.

Berman, Nicolas, Mathieu Couttenier, Dominic Rohner, and Mathias Thoenig. 2017. "This Mine Is Mine! How Minerals Fuel Conflicts in Africa." *American Economic Review* 107 (6): 1564–1610.

Berwick, Elissa, and Fotini Christia. 2018. "State Capacity Redux: Integrating Classical and Experimental Contributions to an Enduring Debate." *Annual Review of Political Science* 21 (1): 71–91.

Boulding, Kenneth E. 1962. *Conflict and Defense: A General Theory.* New York: Harper & Row.

Braithwaite, Alex. 2010. "Resisting Infection: How State Capacity Conditions Conflict Contagion." *Journal of Peace Research* 47 (3): 311–319.

Brambor, Thomas, Agustín Goenaga, Johannes Lindvall, and Jan Teorell. 2020. "The Lay of the Land: Information Capacity and the Modern State:" *Comparative Political Studies* 53 (2): 175–213.

Buhaug, Halvard. 2010. "Dude, Where's My Conflict?: LSG, Relative Strength, and the Location of Civil War." *Conflict Management and Peace Science* 27 (2): 107–128.

Buhaug, Halvard, and Jan Ketil Rød. 2006. "Local Determinants of African Civil Wars, 1970-2001." *Political Geography* 25 (3): 315–335.

Burgess, Robin, Remi Jedwab, Edward Miguel, Ameet Morjaria, and Gerard Padró i Miquel. 2015. "The Value of Democracy: Evidence from Road Building in Kenya." *American Economic Review* 105 (6): 1817–1851.

Campante, Filipe R., Quoc-Anh Do, and Bernardo Guimaraes. 2019. "Capital Cities, Conflict, and Misgovernance." *American Economic Journal: Applied Economics* 11 (3): 298–337.

Center for International Earth Science Information Network, International Food Policy Research Institute, The World Bank, and Centro Internacional de Agricultura Tropical. 2011. *Global Rural-Urban Mapping Project, Version 1 (GRUMPv1): Population Count Grid.*

Collier, Paul, and Anke Hoeffler. 2004. "Greed and Grievance in Civil War." *Oxford Economic Papers* 56 (4): 563–595.

Cordella, Tito, and Harun Onder. 2020. "Sharing Oil Rents and Political Violence." *European Journal of Political Economy:* In Press.

Cotet, Anca M., and Kevin K. Tsui. 2013. "Oil and Conflict: What Does the Cross Country Evidence Really Show?" *American Economic Journal: Macroeconomics* 5 (1): 49–80.

Danielson, Jeffrey, and Dean Gesch. 2011. *Global Multi-Resolution Terrain Elevation Data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011–1073.*

Dell, Melissa, Nathan Lane, and Pablo Querubin. 2018. "The Historical State, Local Collective Action, and Economic Development in Vietnam." *Econometrica* 86 (6): 2083–2121.

DeRouen, Karl R., and David Sobek. 2004. “The Dynamics of Civil War Duration and Outcome:” *Journal of Peace Research* 41 (3): 303–320.

Dickens, Andrew. 2018. “Ethnolinguistic Favoritism in African Politics.” *American Economic Journal: Applied Economics* 10 (3): 370–402.

Dube, Oeindrila, and Juan F. Vargas. 2013. “Commodity Price Shocks and Civil Conflict: Evidence from Colombia.” *The Review of Economic Studies* 80 (4): 1384–1421.

Dudley, Bob. 2019. *BP Statistical Review of World Energy, 68th Ed.* Statistical Review. London, UK: BP.

Earth Observation Group, NOAA. 2019. *Version 4 DMSP-OLS Nighttime Lights Time Series.* Dataset. National Oceanic and Atmospheric Administration.

Elmi, Afyare Abdi, and Dr Abdullahi Barise. 2006. “The Somali Conflict: Root Causes, Obstacles, and Peace-Building Strategies.” *African Security Review* 15 (1): 32–54.

Fearon, James D. 2005. “Primary Commodity Exports and Civil War.” *Journal of Conflict Resolution* 49 (4): 483–507.

Fearon, James D., and David D. Laitin. 2003. “Ethnicity, Insurgency, and Civil War.” *American Political Science Review* 97 (1): 75–90.

Fergusson, Leopoldo, Carlos A Molina, and James A Robinson. 2020. *The Weak State Trap.* Working Paper 26848. National Bureau of Economic Research.

Fjelde, Hanne, and Indra De Soysa. 2009. “Coercion, Co-Optation, or Cooperation?: State Capacity and the Risk of Civil War, 1961-2004.” *Conflict Management and Peace Science* 26 (1): 5–25.

Fukuyama, Francis. 2005. “Building Democracy After Conflict: "Stateness" First.” *Journal of Democracy* 16 (1): 84–88.

———. 2013. “What Is Governance?” *Governance* 26 (3): 347–368.

Gennaioli, Nicola, and Ilia Rainer. 2007. “The Modern Impact of Precolonial Centralization in Africa.” *Journal of Economic Growth* 12 (3): 185–234.

Gleditsch, Kristian S., and Andrea Ruggeri. 2010. “Political Opportunity Structures, Democracy, and Civil War.” *Journal of Peace Research* 47 (3): 299–310.

Hanson, Jonathan K., and Rachel Sigman. 2013. *Leviathan’s Latent Dimensions: Measuring State Capacity for Comparative Political Research.* SSRN SCHOLARLY PAPER ID 1899933. Rochester, NY: Social Science Research Network.

Harbers, Imke. 2015. “Taxation and the Unequal Reach of the State: Mapping State Capacity in Ecuador.” *Governance* 28 (3): 373–391.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. 2012. “Measuring Economic Growth from Outer Space.” *American Economic Review* 102 (2): 994–1028.

Hendrix, Cullen S. 2011. “Head for the Hills? Rough Terrain, State Capacity, and Civil War Onset.” *Civil Wars* 13 (4): 345–370.

———. 2010. “Measuring State Capacity: Theoretical and Empirical Implications for the Study of Civil Conflict:” *Journal of Peace Research* 47 (3): 273–285.

Herbst, Jeffrey. 2000. *States and Power in Africa: Comparative Lessons in Authority andControl.* Princeton Studies in International History and Politics. New Jersey: Princeton University Press.

Hesselbein, Gabi. 2007. *The Rise and Decline of the Congolese State: An Analytical Narrative on State-Making.* Working Paper 21. Crisis States Research Centre.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning: With Applications in R.* New York: Springer.

Johnson, Noel D., and Mark Koyama. 2017. “States and Economic Growth: Capacity and Constraints.” *Explorations in Economic History* 64:1–20.

Justino, Patricia. 2011. “Carrot or Stick? Redistributive Transfers versus Policing in Contexts of Civil Unrest.” *IDS Working Papers* 2011 (382): 1–29.

Justino, Patricia, and Bruno Martorano. 2018. “Welfare Spending and Political Conflict in Latin America, 1970-2010.” *World Development* 107:98–110.

Kaufmann, Daniel, and Aart Kray. 2019. *Worldwide Governance Indicators.* Dataset. New York: World Bank.

Klein Goldewijk, Kees, Arthur Beusen, Jonathan Doelman, and Elke Stehfest. 2017. “Anthropogenic Land Use Estimates for the Holocene – HYDE 3.2.” *Earth System Science Data* 9 (2): 927–953.

Koenig, Nicole. 2016. *EU Security Policy and Crisis Management: A Quest for Coherence.* Routledge.

Koren, Ore, and Anoop K. Sarbahi. 2018. "State Capacity, Insurgency, and Civil War: A Disaggregated Analysis." *International Studies Quarterly* 62 (2): 274–288.

Lacina, Bethany. 2006. "Explaining the Severity of Civil Wars." *Journal of Conflict Resolution* 50 (2): 276–289.

Lee, Melissa M., and Nan Zhang. 2017. "Legibility and the Informational Foundations of State Capacity." *The Journal of Politics* 79 (1): 118–132.

Lessmann, Christian, and Arne Steinkraus. 2019. "The Geography of Natural Resources, Ethnic Inequality and Civil Conflicts." *European Journal of Political Economy* 59:33–51.

Levi, Margaret. 1989. *Of Rule and Revenue.* Berkeley: University of California Press.

Lindvall, J., and J. Teorell. 2016. *State Capacity as Power: A Conceptual Framework.* STANCE Working Paper 1. Lund, Sweden: Lund University.

Lujala, Päivi, Jan Ketil Rod, and Nadja Thieme. 2007. "Fighting over Oil: Introducing a New Dataset:" *Conflict Management and Peace Science* 24 (3): 239–256.

Luna, Juan Pablo, and Hillel David Soifer. 2017. "Capturing Sub-National Variation in State Capacity: A Survey-Based Approach." *American Behavioral Scientist* 61 (8): 887–907.

Mann, Michael. 1984. “The Autonomous Power of the State: Its Origins, Mechanisms and Results.” *European Journal of Sociology / Archives Européennes de Sociologie* 25, no. 2 (November): 185–213.

Michalopoulos, Stelios, and Elias Papaioannou. 2013. “Pre-Colonial Ethnic Institutions and Contemporary African Development.” *Econometrica* 81 (1): 113–152.

Morelli, Massimo, and Dominic Rohner. 2015. “Resource Concentration and Civil Wars.” *Journal of Development Economics* 117:32–47.

Mosser, Jonathan F., William Gagne-Maynard, Puja C. Rao, Aaron Osgood-Zimmerman, Nancy Fullman, Nicholas Graetz, Roy Burstein, et al. 2019. “Mapping Diphtheria-Pertussis-Tetanus Vaccine Coverage in Africa, 2000–2016: A Spatial and Temporal Modelling Study.” *The Lancet* 393 (10183): 1843–1855.

Murdock, George Peter. 1959. *Africa: Its Peoples and Their Culture History.* New York: McGraw Hill Text.

Naudé, Wim. 2013. *Development Progress in Sub-Saharan Africa: Lessons from Botswana, Ghana, Mauritius, and South Africa.* Oxford University Press.

Nelson, A. D. 2008. *Travel Time to Major Cities: A Global Map of Accessibility.* Dataset. Luxembourg: Office for Official Publications of the European Communities.

Nordvik, Frode Martin. 2019. “Does Oil Promote or Prevent Coups? The Answer Is Yes.” *The Economic Journal* 129 (619): 1425–1456.

Nwokolo, Arinze. 2018. *Oil Price Shocks and Civil Conflict: Evidence from Nigeria.* Working Paper 274. Sussex, UK: The Institute of Development Studies, University of Sussex.

O'Donnell, Guillermo. 1993. "On the State, Democratization and Some Conceptual Problems: A Latin American View with Glances at Some Postcommunist Countries." *World Development,* SPECIAL ISSUE, 21, no. 8 (August): 1355–1369.

Raleigh, Clionadh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature." *Journal of Peace Research* 47 (5): 651–660.

Richardson, Craig J. 2007. "How Much Did Droughts Matter? Linking Rainfall and GDP Growth in Zimbabwe." *African Affairs* 106 (424): 463–478.

Ross, Michael L. 2012. *The Oil Curse: How Petroleum Wealth Shapes the Development of Nations.* Princeton University Press.

Rotberg, Robert I., ed. 2003. *When States Fail: Causes and Consequences.* Princeton, N.J: Princeton University Press.

Sánchez de la Sierra, Raúl. 2020. "On the Origins of the State: Stationary Bandits and Taxation in Eastern Congo." *Journal of Political Economy* 128 (1): 32–74.

Sawyer, Amos. 2005. *Beyond Plunder: Toward Democratic Governance in Liberia.* Boulder: Lynne Rienner Publishers.

Schumpeter, Joseph A. [1918] 1954. "The Crisis of the Tax State." In *International Economic Papers,* edited by A. T. Peacock. Numer 4. London: Translated by Wolfgang F. Stolper and Richard A. Musgrave.

Shaver, Andrew, David B. Carter, and Tsering Wangyal Shawa. 2019. "Terrain Ruggedness and Land Cover: Improved Data for Most Research Designs." *Conflict Management and Peace Science* 36 (2): 191–218.

Soifer, Hillel. 2012. "Measuring State Capacity in Contemporary Latin America." *Revista de ciencia política (Santiago)* 32 (3): 585–598.

———. 2008. "State Infrastructural Power: Approaches to Conceptualization and Measurement." *Studies in Comparative International Development* 43 (3): 231.

Soifer, Hillel, and Matthias vom Hau. 2008. "Unpacking the Strength of the State: The Utility of State Infrastructural Power." *Studies in Comparative International Development* 43 (3): 219.

Sundberg, Ralph, and Erik Melander. 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research* 50 (4): 523–532.

Thies, Cameron G. 2010. "Of Rulers, Rebels, and Revenue: State Capacity, Civil War Onset, and Primary Commodities:" *Journal of Peace Research* 47 (3): 321–332.

Tilly, Charles, ed. 1975. *The Formation of National States in Western Europe.* Princeton, NY: Princeton University Press.

van der Laan, Mark, and Sandrine Dudoit. 2003. *Unified Cross-Validation Methodology For Selection Among Estimators and a General Cross-Validated Adaptive Epsilon-Net Estimator: Finite Sample Oracle Inequalities and Examples.* U.C. Berkeley Division of Biostatistics Working Paper Series Paper 130.

van der Laan, Mark J., Eric C. Polley, and Alan E. Hubbard. 2007. "Super Learner." *Statistical Applications in Genetics and Molecular Biology* 6 (1).

van der Laan, Mark J., and Sherri Rose. 2011. *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer Series in Statistics. New York: Springer-Verlag.

Vogt, Manuel, Nils-Christian Bormann, Seraina Rüegger, Lars-Erik Cederman, Philipp Hunziker, and Luc Girardin. 2015. "Integrating Data on Ethnicity, Geography, and Conflict: The Ethnic Power Relations Data Set Family." *Journal of Conflict Resolution* 59 (7): 1327–1342.

Weber, Max. 1991. "Politics as a Vocation." In *From Max Weber: Essays in Sociology,* edited by H. H. Gerth and C. W. Mills, 77–128. London: Routledge.

Wucherpfennig, Julian, Nils B. Weidmann, Luc Girardin, Lars-Erik Cederman, and Andreas Wimmer. 2011. "Politically Relevant Ethnic Groups across Space and Time: Introducing the GeoEPR Dataset." *Conflict Management and Peace Science* 28 (5): 423–437.

Appendix A: Additional figures and tables

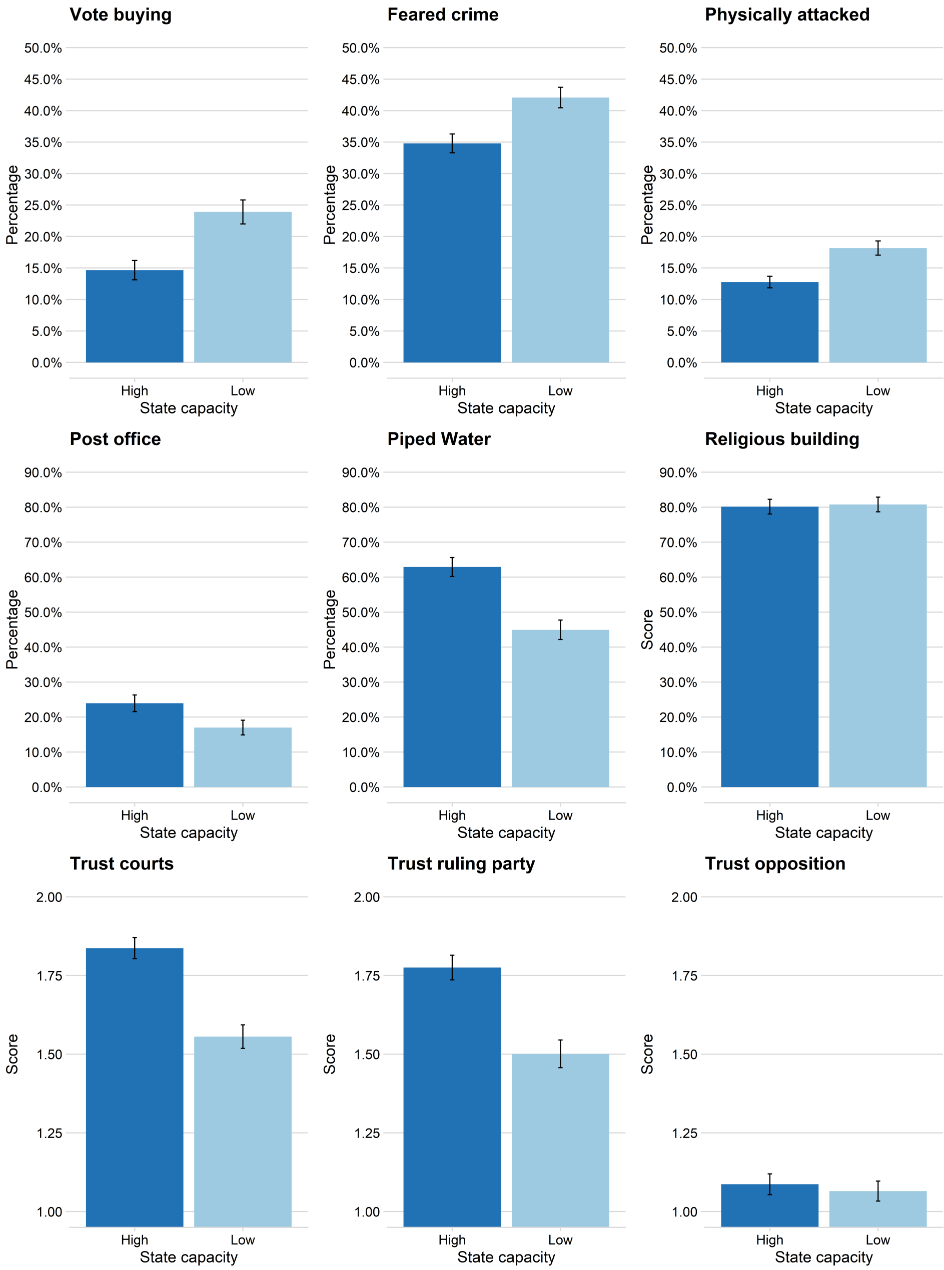


Figure A1: *Survey data sanity check*

Notes: This figure provides sanity checks of the survey data. The first row shows the share of EA-level vote buying, fear of crime, and actual physical attacks for high (above median) and low (below median) state capacity EAs, respectively. The second row features bar graphs on physical infrastructure reported by the enumerators in the visited areas. The last row depicts levels of trust in the courts, ruling party, and political opposition.
Source: Afrobarometer and authors' calculations.

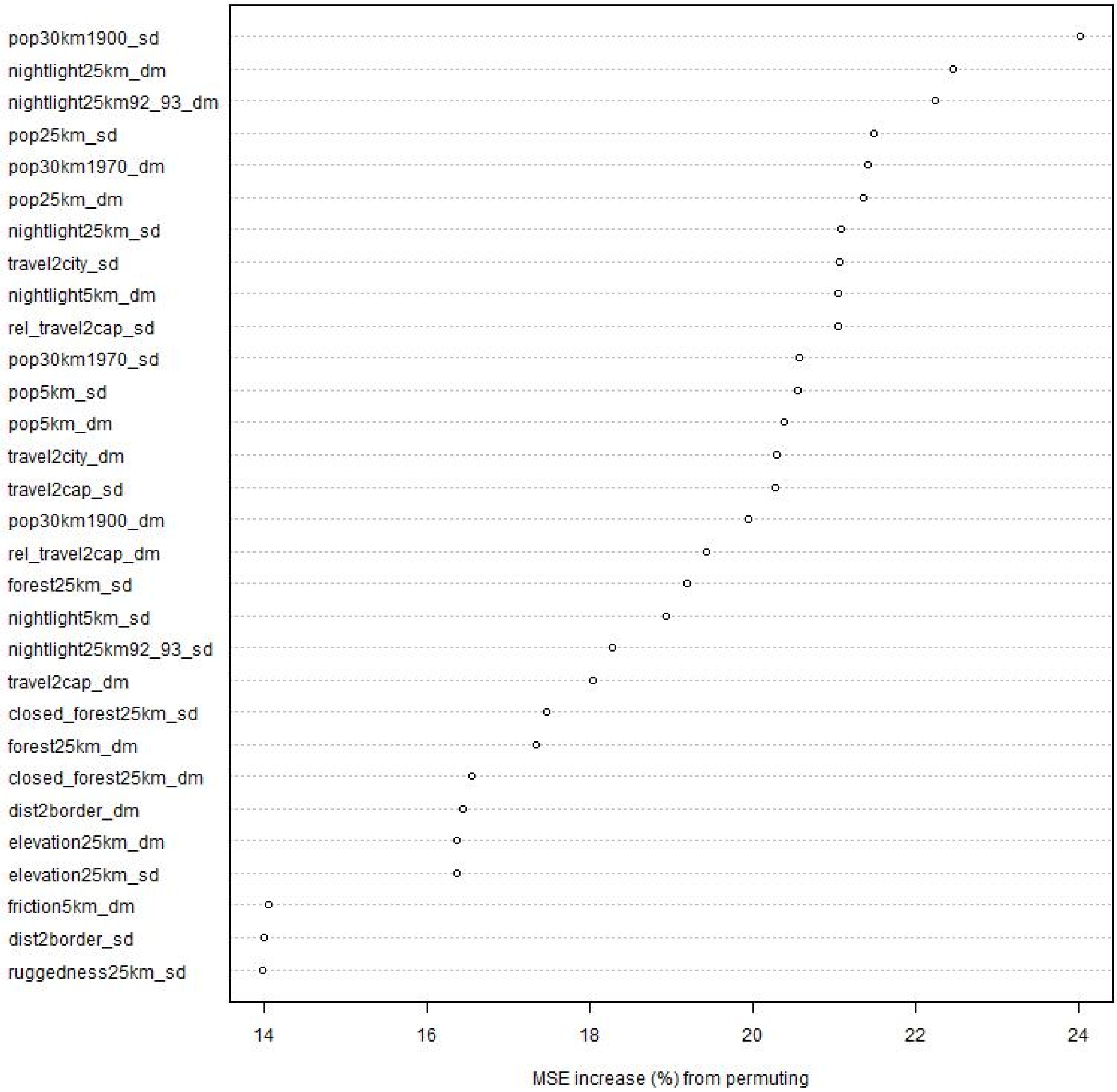
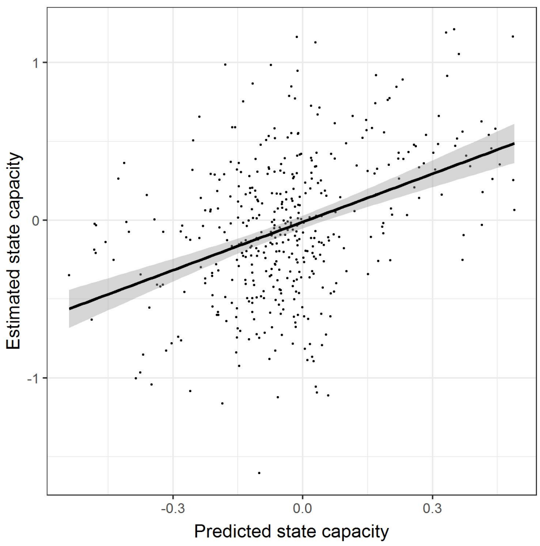
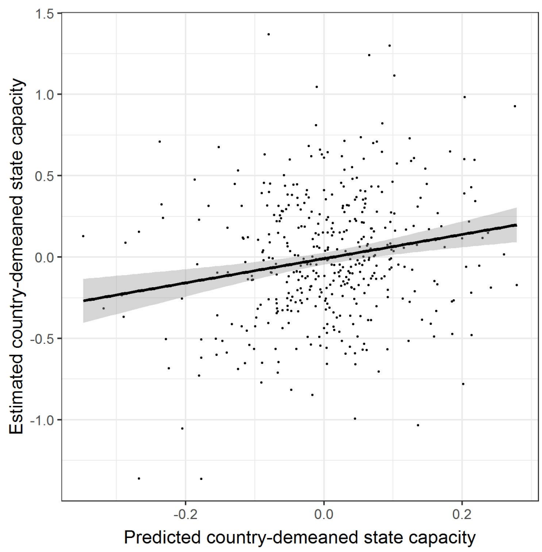


Figure A2: *Variable importance in Random Forest model with two predictors in each tree*

Notes: The x-axis refers to the mean increase in MSE from permuting a specific variable. The higher the value, the more important is the variable in predicting the outcome variable in the random forest algorithm. Suffix "_sd" refers to variables that have been within-country standardized, whereas suffix "_dm" refers to variables that have been country-demeaned.
Source: Authors' illustration.



(a) *Country baseline + country-demeaned predictions*



(b) *Country-demeaned predictions only*

Figure A3: *Relationship between predicted and survey state capacity for a hold-out test sample*

Notes: The left graph shows the relationship between the estimated state capacity index (based on factor analysis) and the predicted state capacity index (based on between- and within-country prediction models). The right graph shows the relationship between the country-demeaned state capacity index and the predicted country-demeaned state capacity index. The best linear fits in the left and right graphs have highly statistically significant (*p-values* < 0.0001) coefficient estimates of 1.02 and 0.74, respectively, and the shaded areas represent 95 percent confidence intervals. The R^2 in the left and right graphs are 0.158 and 0.037, respectively. For the illustrative purpose of this graph, the test sample is not used to fit the ensemble model. In the ensemble model used for extrapolation, all EAs are used.
Source: Authors' illustration.

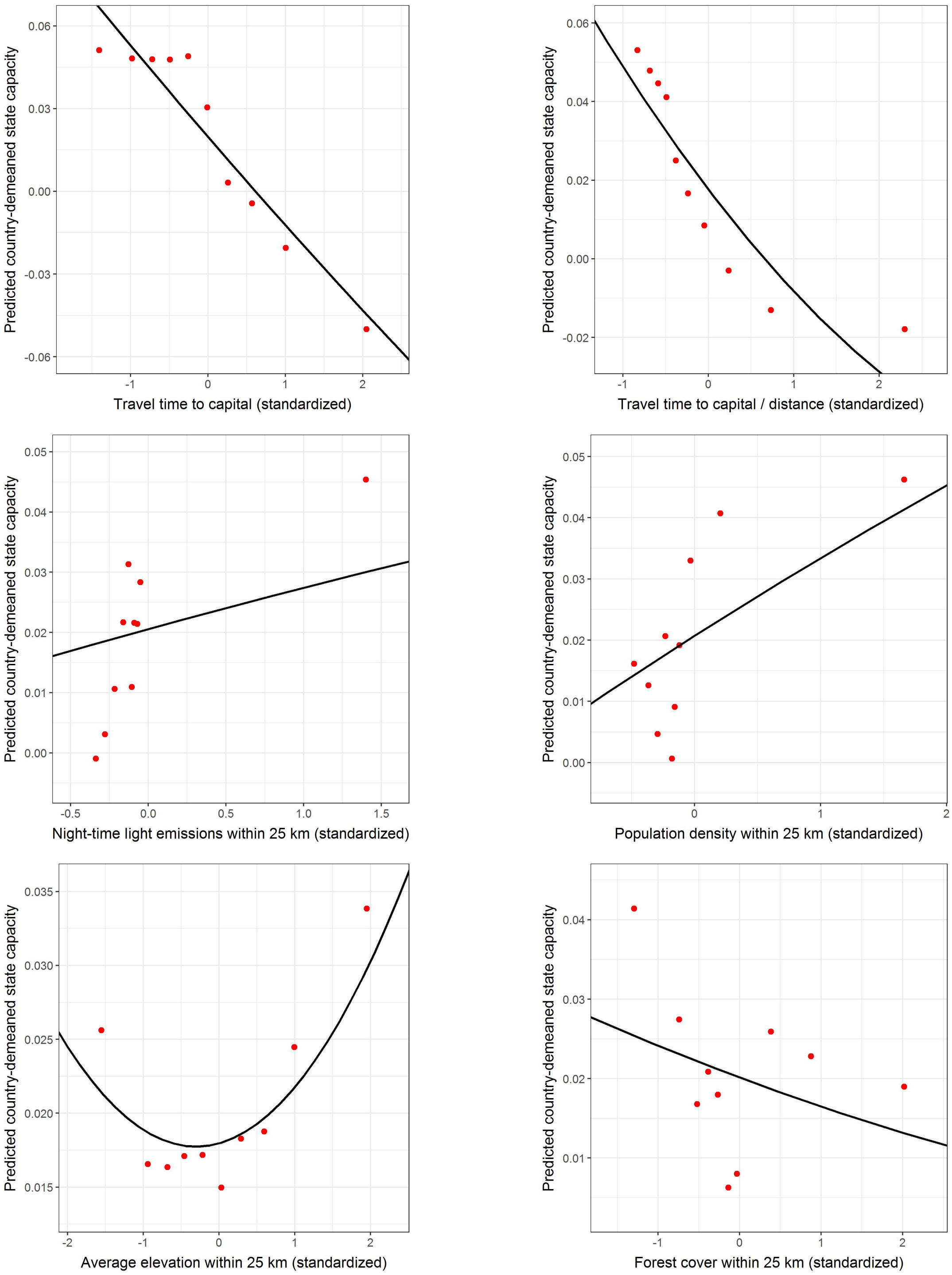


Figure A4: *Bivariate relationships between predicted state capacity and predictors*

Notes: The plots show the average predicted country-demeaned state capacity and average of the predictor for the different deciles of the predictor. The illustrated predictors are standardized within each country. The black lines represent best quadratic fits to underlying non-binned data.
Source: Authors' illustration.

Table A1: *Country-demeaned and standardized variables used for predicting local state capacity*

Predictor name	Description
travel2cap	Travel time to country capital in year 2000.
rel_travel2cap	Travel time to country capital divided by geodesic distance to capital . (travel time per kilometer)
travel2city	Travel time to nearest urban center in year 2000.
dist2border	Geodesic distance to the country border.
friction5km	Infrastructural accessibility within a 5 kilometers radius in year 2000 (easiness of crossing a 1×1 kilometer cell).
nightlight25km92_93	Average night-time light emissions within a 25 kilometers radius in 1992 . and 1993
nightlight5km	Average night-time light emissions within a 5 kilometers radius in 2000.
nightlight25km	Average night-time light emissions within a 25 kilometers radius in 2000.
pop30km1900	Population density within a 30 kilometers radius in 1900.
pop30km1970	Population density within a 30 kilometers radius in 1970.
pop5km	Population density within a 5 kilometers radius in 2000.
pop25km	Population density within a 25 kilometers radius in 2000.
elevation25km	Average elevation above sea leavel in meters within a 25 kilometers radius.
ruggedness25km	Average topographic ruggedness index (TRI) within a 25 kilometers radius. For each 2.5×2.5 arc-minutes grid cell ($\approx 5 \times 5$ kilometers), we take the mean of the absolute differences between a cell's elevation and its eight surrounding neighbors. Next, we take the average TRI for cells within a 25 kilometers radius.
forest25km	Share covered by forest within a 25 kilometers radius.
closed_forest25km	Share covered by closed forest within a 25 kilometers radius.

Notes: All predictors are included both as country-demeaned variables and as standardized within-country variables. Travel time and friction variables are from Nelson (2008), night-time light emission variables are from Earth Observation Group, NOAA (2019), population data from year 2000 is from Center for International Earth Science Information Network et al. (2011) and Balk et al. (2006), historic population data is from Klein Goldewijk et al. (2017), elevation and ruggedness are based on Danielson and Gesch (2011), and forest cover is from Shaver, Carter, and Shawa (2019).

Table A2: *The conditional effect of oil price fluctuations on the likelihood of oil conflict; Lag and lead prices*

	Oil event (1)–(3)			Oil violence (4)–(6)			Oil demonstration (7)–(9)		
	Low SC (1)	High SC (2)	Interaction (3)	Low SC (4)	High SC (5)	Interaction (6)	Low SC (7)	High SC (8)	Interaction (9)
Panel A:									
Lagged oil price × Oil cell	0.049**	0.00031	0.070**	0.026**	0.0016	0.039**	0.023*	-0.0012**	0.031*
	(0.022)	(0.0019)	(0.031)	(0.0097)	(0.0017)	(0.018)	(0.014)	(0.00055)	(0.017)
Lagged oil price × State capacity			-0.00058			-0.00048*			-0.00011
			(0.00039)			(0.00024)			(0.00016)
Lagged oil price × State capacity × Oil cell			-0.015*			-0.0085**			-0.0063*
			(0.0074)			(0.0046)			(0.0035)
R-squared	.319	.128	.294	.252	.065	.244	.235	.130	.212
Observations	80,142	79,591	159,733	80,142	79,591	159,733	80,142	79,591	159,733
Dependent variable mean	.0044	.00067	.0025	.0027	.00011	.0014	.0019	.00053	.0012
Dependent variable std. dev.	.066	.026	.050	.052	.011	.038	.043	.023	.035
Panel B:									
Lead oil price × Oil cell	0.025	0.0013	0.029	0.026	0.0010	0.036	0.0029	0.00029	-0.00087
	(0.017)	(0.0011)	(0.025)	(0.020)	(0.0012)	(0.033)	(0.0028)	(0.00043)	(0.0077)
Lead oil price × State capacity			-0.00037*			-0.00025**			-0.00009
			(0.00021)			(0.00011)			(0.00011)
Lead oil price × State capacity × Oil cell			-0.0047			-0.0072			0.0011
			(0.0059)			(0.0076)			(0.0024)
R-squared	.318	.131	.294	.258	.068	.251	.226	.136	.206
Observations	75,924	75,402	151,326	75,924	75,402	151,326	75,924	75,402	151,326
Dependent variable mean	.0044	.00066	.0039	.0028	.00012	.0015	.0018	.00053	.0012
Dependent variable std. dev.	.066	.026	.062	.053	.011	.038	.042	.023	.034

Notes: The dependent variables in columns (1)-(3), (4)-(6), and (7)-(9) are cell-year indicator variables for 'oil-related conflict event', 'violent oil-related conflict event' and 'oil-related demonstration', respectively. Columns (1), (4), and (7) restrict the sample to cells below the median state capacity, whereas columns (2), (5), and (8) restrict the sample to cells above median state capacity. Columns (3), (6), and (9) display results from models including the continuous measure of local state capacity in a triple interaction with oil cell and the oil price. *Lagged oil price* refers to the *Oil price* in $t-1$, whereas *lead oil price* refers to the *Oil price* in $t+1$. *Oil cell* is an indicator variable on known oil deposits in year 2000, and *State capacity* refers to our measure of local state capacity in year 2000. Standard errors clustered at the country level are in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A3: *The conditional effect of oil price fluctuations on the likelihood of oil conflict: Oil proximity*

	Oil event (1)–(3)			Oil violence (4)–(6)			Oil demonstration (7)–(9)		
	Low SC (1)	High SC (2)	Interaction (3)	Low SC (4)	High SC (5)	Interaction (6)	Low SC (7)	High SC (8)	Interaction (9)
Panel A:									
Oil price × Oil proximity	0.062**	-0.0015	0.096**	0.040**	0.00018	0.068**	0.021**	-0.0016	0.028**
	(0.025)	(0.0013)	(0.039)	(0.016)	(0.00027)	(0.033)	(0.010)	(0.0011)	(0.013)
Oil price × State capacity			0.00056			0.00036			0.00023
			(0.00040)			(0.00036)			(0.00015)
Oil price × State capacity × Oil proximity			-0.022**			-0.016*			-0.0057
			(0.0097)			(0.0086)			(0.0034)
R-squared	.318	.128	.293	.251	.0646	.244	.234	.130	.211
Observations	80,142	79,591	159,733	80,142	79,591	159,733	80,142	79,591	159,733
Panel B:									
Oil price × Oil proximity	0.055**	-0.0014	0.087**	0.035**	0.00019	0.062**	0.017**	-0.0016	0.022**
	(0.022)	(0.0013)	(0.036)	(0.014)	(0.00027)	(0.029)	(0.0076)	(0.0011)	(0.0093)
Oil price × State capacity			0.00056			0.00038			0.00017
			(0.00036)			(0.00033)			(0.00012)
Oil price × State capacity × Oil proximity			-0.019**			-0.015*			-0.0041
			(0.0089)			(0.0078)			(0.0030)
Dependent variable temporal lag	0.078***	-0.033	0.061***	0.052**	-0.070***	0.047**	0.034	-0.036	0.019
	(0.014)	(0.028)	(0.015)	(0.022)	(0.0086)	(0.020)	(0.029)	(0.037)	(0.035)
Dependent variable spatial lag	0.028***	0.0022	0.022***	0.027***	0.0058	0.025***	0.054***	-0.0019	0.036***
	(0.0066)	(0.0062)	(0.0058)	(0.0078)	(0.0091)	(0.0073)	(0.011)	(0.0087)	(0.012)
R-squared	.324	.129	.297	.256	.0693	.248	.245	.132	.217
Observations	80,142	79,591	159,733	80,142	79,591	159,733	80,142	79,591	159,733
Cell & year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable mean	.0044	.00067	.0025	.0027	.00011	.0014	.0019	.00053	.0012
Dependent variable std. dev.	.066	.026	.050	.052	.011	.038	.043	.023	.035

Notes: The dependent variables in columns (1)-(3), (4)-(6), and (7)-(9) are cell-year indicator variables for 'oil-related conflict event', 'violent oil-related conflict event' and 'oil-related demonstration', respectively. Columns (1), (4), and (7) restrict the sample to cells below the median state capacity, whereas columns (2), (5), and (8) restrict the sample to cells above median state capacity. Columns (3), (6), and (9) display results from models including the continuous measure of local state capacity in a triple interaction with oil proximity and the oil price. *Oil price* is the log of the average oil price for each year, *Oil proximity* is a continuous variable capturing the inverse of the square root of distance to known oil deposits in year 2000, and *State capacity* refers to our measure of local state capacity in year 2000. Standard errors clustered at the country level are in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table A4: *The conditional effect of oil price fluctuations on the likelihood of oil conflict: Country-year fixed effects*

	Oil event (1)–(3)			Oil violence (4)–(6)			Oil demonstration (7)–(9)		
	Low SC (1)	High SC (2)	Interaction (3)	Low SC (4)	High SC (5)	Interaction (6)	Low SC (7)	High SC (8)	Interaction (9)
Panel A:									
Oil price × Oil cell	0.038**	-0.00071	0.056**	0.024**	0.00013	0.042*	0.013**	-0.00083	0.015*
	(0.016)	(0.00064)	(0.023)	(0.010)	(0.00015)	(0.021)	(0.0061)	(0.00057)	(0.0075)
Oil price × State capacity			-0.0013			-0.0010			-0.00022
			(0.0014)			(0.00096)			(0.00075)
Oil price × State capacity × Oil cell			-0.012**			-0.0100*			-0.0027
			(0.0057)			(0.0055)			(0.0023)
R-squared	.326	.139	.301	.260	.078	.253	.247	.143	.223
Observations	80,028	79,534	159,676	80,028	79,534	159,676	80,028	79,534	159,676
Panel B:									
Oil price × Oil cell	0.034**	-0.00059	0.052**	0.022**	0.00021	0.039**	0.011**	-0.00081	0.012*
	(0.015)	(0.00073)	(0.022)	(0.0094)	(0.00025)	(0.019)	(0.0050)	(0.00060)	(0.0062)
Oil price × State capacity			-0.0011			-0.00089			-0.00022
			(0.0013)			(0.00090)			(0.00067)
Oil price × State capacity × Oil cell			-0.011**			-0.0094*			-0.0021
			(0.0054)			(0.0051)			(0.0022)
Dependent variable temporal lag	0.075***	-0.032	0.059***	0.052**	-0.072***	0.046**	0.027	-0.037	0.013
	(0.015)	(0.028)	(0.015)	(0.022)	(0.0079)	(0.021)	(0.029)	(0.036)	(0.033)
Dependent variable spatial lag	0.019**	-0.0074	0.014**	0.020***	-0.00026	0.018***	0.043***	-0.013	0.026**
	(0.0072)	(0.0074)	(0.0065)	(0.0068)	(0.0098)	(0.0065)	(0.011)	(0.011)	(0.012)
R-squared	.331	.140	.304	.264	.083	.256	.254	.145	.226
Observations	80,028	79,534	159,676	80,028	79,534	159,676	80,028	79,534	159,676
Cell & country × year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable mean	.0044	.00067	.0025	.0027	.00011	.0014	.0019	.00053	.0012
Dependent variable std. dev.	.066	.026	.050	.052	.011	.038	.043	.023	.035

Notes: The dependent variables in columns (1)-(3), (4)-(6), and (7)-(9) are cell-year indicator variables for ‘oil-related conflict event’, ‘violent oil-related conflict event’ and ‘oil-related demonstration’, respectively. Columns (1), (4), and (7) restrict the sample to cells below the median state capacity, whereas columns (2), (5), and (8) restrict the sample to cells above median state capacity. Columns (3), (6), and (9) display results from models including the continuous measure of local state capacity in a triple interaction with oil cell and the oil price. *Oil price* is the log of the average oil price for each year, *Oil cell* is an indicator variable on known oil deposits in year 2000, and *State capacity* refers to our measure of local state capacity in year 2000. Standard errors clustered at the country level are in parentheses. Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Appendix B: Tree-based prediction methods

Bagging uses bootstrapping to generate B random data sets from the training sample. For each random data set, we use the predictor that minimizes the residual sum of squares (RSS) by cutting the sample into two subgroups and taking the mean of the outcome variable. This procedure continues until the algorithm wants to divide a subgroup such that at least one of the proposed subgroups has less than 25 observations. Each observation is given the mean outcome value of the subgroup it belongs to. The predicted outcome variable of observation i is the mean of the B different estimates. The disadvantage of bagging is that a few strong predictors may dominate less influential predictors. Hence, the less influential predictors will never divide the sample, and we would falsely predict they are not associated with the outcome variable. The issue is resolved in random forest by considering a subset of the predictors in each regression tree rather than all predictors. Random forest and bagging are otherwise identical.

Boosting starts by predicting all observations to have an outcome variable of zero, and hence residuals are equal to the actual outcome variable. Next, a regression tree is fitted on the residuals with d splits (called the interaction depth), and observations are given the mean value of the outcome variable of the subgroup they end up in. The predicted outcome variable is multiplied by a factor called the shrinkage parameter, which in turn is subtracted from the initial residuals. Next, the procedure of fitting a regression tree starts over, this time fitting the updated residuals. By letting this procedure continue in eternity, boosting would overfit the training sample and it would not be good at predicting the test sample. We use 5-fold cross-validation to select the number of trees (iterations) in order to avoid overfitting. In addition to the interaction depth and shrinkage parameter, we tune parameters related to penalizing small improvements to overall performance (the gamma parameter), share of the training sample used for each tree, and share of predictors used for each tree. These parameters are tuned by evaluating the prediction performance of the model when changing the parameters.