The Returns to Higher Education and Public Employment*

Ravi Somani†

November 2020

*Declarations of interest: none.

†The World Bank, 1818 H St NW, Washington DC, USA. rsomani@worldbank.org. +1 202 473 2078

Abstract

This paper evaluates a rapid expansion in public universities in Ethiopia to estimate the returns to higher education in a low-income context. The empirical analysis exploits the geographical and time variation in the expansion. The estimates suggest that higher-education attainment almost doubles the probability of paid employment and almost doubles hourly wages. The returns are linked to occupational shifts to the public sector, which is characterized by attractive wages and demands for higher levels of education.

Keywords: Higher education, returns to education, public employment, occupational choice

Highlights

- I evaluate the effects of an expansion in public universities in Ethiopia to estimate the returns to higher education
- I use spatial and time variation in the expansion to identify its effects on education attainment and labor-market returns
- The expansion leads to a 4-percentage-point increase in higher-education attainment, a 31% increase relative to baseline
- Higher-education attainment almost doubles the probability of entering paid employment and increases hourly wages by 80%
- The returns to higher education are linked to occupational shifts to public administration

Acknowledgements

I would like to thank Marcos Vera-Hernandez and Antonio Cabrales for their support and excellent supervision. I am grateful to Pedro Carneiro, Jishnu Das, Florian Englmaier, Hyejin Ku, Eliana La Ferrara, David McKenzie, Fabian Postel-Vinay, Imran Rasul, Daniel Rogger and anonymous referees for valuable feedback. This paper benefited from valuable feedback from presentations at the Ethiopian Development Research Institute, National University of Singapore, Queen Mary's, University of London, The University of Oxford, The University of Sussex, The University of Warwick, University College London, and the World Bank. I am thankful to the World Bank's Bureaucracy Lab, the Ethiopian Ministry of Public Service and Human Resource Development, Elsa Araya, Tom Bundervoet, Jean-Paul Faguet, Verena Fritz, Christopher Gaukler, Qaiser Khan, and Wendmsyamregne Mekasha for support and access to data.

1 Introduction

Globally, tertiary education accounts for 20% of government expenditure on education (UNESCO, 2020). Across OECD countries, 1.5% of GDP is spent on tertiary education. The demand for tertiary education continues to grow: gross enrollment in tertiary education has increased from 19% to 38% since 2000 (World Bank, 2020a). Yet, there is little causal evidence on the returns to higher education outside of the US (Oreopoulos and Petronijevic, 2013; Peet et al., 2015; Patrinos and Psacharopoulos, 2020).

This paper exploits a rapid expansion in public universities in Ethiopia to provide some of the first causal estimates of the returns to higher education in a low-income setting. The expansion led to a significant increase in higher-education attainment (4-percentage-points) and I estimate the employment and wage returns to higher education at 90% and 80%, respectively. I find evidence that these returns are linked to occupational shifts towards public administration, which is consistent with the importance of public employment for high-skilled individuals in low-income contexts (Finan et al., 2017; World Bank, 2020b). Public-sector wage and recruitment policies may, therefore, mediate the accumulation and distribution of human capital in developing countries.

During the 2000s, the Ethiopian government engaged in a rapid expansion of public universities across the country. The number of universities increased from 8 to 33 in just over 10 years and the expansion was geographically spread due to equity and political considerations, incorporating both time and spatial variation (Ashcroft and Rayner, 2011). Moreover, the expansion occurred in a context with restricted and low levels of migration across districts, which are self-contained local economies by policy design (Federal Democratic Republic of Ethiopia, 2003; Dorosh and Schmidt, 2010).

To study the effects of the expansion, I build a new dataset comprising a large cross-section of individuals from the 2013 Labor Force Survey, a hand-coded geo-located database of universities, and census data. I categorise individuals by their level of 'exposure' to the expansion, in a difference-in-differences framework, following the approach of Card and Lemieux (2001), Duflo (2001), and Atkin (2016). I categorise individuals into exposed cohorts based on the standard age of entry into higher education. I assume that individuals over the standard age of entry face higher access costs to university, for example due to investments in job-specific human capital, family responsibilities or the difficulty of re-entering the education system (Meng and Gregory, 2002). This assumption is strongly supported by the data. I also categorise individuals into exposed districts if they reside in a district with a public university, assuming a lower cost of access through, for example, transport and accommodation costs. This assumption is driven by the context of restricted migration and well-defined district markets. I restrict the analysis to non-rural districts, which are balanced across a range of demographic characteristics and find strong evidence in support of the parallel-trends assumption.

Using this framework, I find that the expansion led to a 4-percentage-point increase in higher-education attainment ($p = 0.005$), a 31% increase relative to baseline levels. I also find that the expansion led to a 4-percentage-point increase in paid employment ($p = 0.059$, a 14% increase) and evidence of a 15% increase in monthly earnings. Using the interactions between cohort exposure and district exposure as instruments for higher-education attainment, I estimate the labor-market returns to higher education. I find that higher education almost doubles the probability of paid employment (a return of 90%) and almost doubles hourly wages (a return of 80%) after accounting for selection into paid employment. These estimates are consistent with correlational estimates in low-income countries, documented below, and imply substantial labor-market returns to higher-education attainment in such contexts.

To study the potential mechanisms behind these results, I draw on descriptive evidence from the Labor Force Survey, the World Bank's Worldwide Bureaucracy Indicators, and Finan et al. (2017) that highlight the importance of public employment for the high-skilled, especially in low-income settings. In Ethiopia, I find that 55% of all employed tertiary-educated individuals work in public administration, relative to 25% of individuals with a secondary education.[1] Public administration is also characterized by high wages and education requirements for entry: wages are approximately 50% higher in public administration than in services, the second best-paid industry; and individuals working in public administration are three-times more likely to have a tertiary education than those working in services, the second-highest.

[1]The global average of employed tertiary-educated individuals working in the public sector is 39% and the low-income average (among countries with less than $3,000 per capita in PPP terms) is 45% according to the World Bank's Worldwide Bureaucracy Indicators.

Consistent with this descriptive evidence, I find evidence that the expansion in higher education leads to a sorting of individuals towards employment in public administration (a 3.5-percentage-point increase, $p = 0.045$) and away from services. Instrumenting for higher-education attainment with the exposure interactions, I find that higher education significantly increases the likelihood of entering public employment in this context (a return of 85%, $p = 0.048$). These findings suggest that the returns to higher education are partially driven by public employment and public-sector wages.

I also investigate the interaction between the baseline public-sector wage premium and the baseline size of the public sector in the local labor market and higher-education attainment. First, I group districts by a measure of the baseline local public-wage premium and estimate the effects of the expansion on higher-education attainment for each level of the premium. Second, I group districts by a measure of the baseline proportion of individuals working in public administration. I find that the expansion leads to an increase in higher-education attainment only in areas where the public-wage premium is positive (the premium is balanced across exposed and unexposed districts) and in areas where the public sector is a relatively more important employer. These findings are suggestive of an interplay between the effects of education policies and public-sector contracting policies. As such, optimal policy formulation may require the design of these policies to be considered simultaneously.

This paper contributes to the vast literature on the potential returns to higher education, which has attracted the attention of scholars for many years (Becker, 1960; Weisbrod and Karpoff, 1968; Card, 1993; Rumberger and Thomas, 1993; Kane and Rouse, 1995; Card, 1999; Arcidiacono, 2004; Black and Smith, 2006; Machin and McNally, 2007; Brand and Xie, 2010; Altonji et al., 2012; Hout, 2012; Liu et al., 2015; Barrow and Malamud, 2015; Rzepka, 2018; Dai et al., 2018; Bettinger et al., 2019). Nevertheless, to date, there still remain few causal estimates of the returns to higher education, particularly outside the US (Peet et al., 2015; Patrinos and Psacharopoulos, 2020).[2] Methodologically, this paper is closest to studies that use variation in schooling access to estimate the effects of education on labor-market outcomes (Duflo, 2001, 2004; Filmer, 2007; Oppedisano, 2014; Li et al., 2014; Ou and Zhao, 2016; Dai

[2]Examples of causal estimates include Card (1993), Lemieux and Card (2001), Bound and Turner (2002), Stanley (2003), Angrist and Chen (2011), and Zimmerman (2014). See Oreopoulos and Petronijevic (2013) for a review.

et al., 2018; Xing et al., 2018).

The estimates of the returns to higher education in this paper are larger than the correlational estimates of 21% per year of tertiary education in Sub-Saharan Africa documented in Montenegro and Patrinos (2014) and consistent with the low-income-country average of 27% per year estimated in Patrinos and Psacharopoulos (2020). The estimates in this paper are similar to recent estimates of college returns in the USA (Goldin and Katz, 2007; Carnevale et al., 2011; Hout, 2012; Oreopoulos and Petronijevic, 2013; Barrow and Malamud, 2015; The Hamilton Project, 2017; Bleemer and Zafar, 2018); larger than older estimates from the UK at 21% for men and 39% for women (Blundell et al., 2000); and similar to those in China at 21% per each year of post-secondary education (Dai et al., 2018). The estimates are closer to the top-end of the estimates of *unconditional* returns to tertiary education in OECD and partner countries: Costa Rica (100%), Mexico (95%), Hungary (79%), Lithuania (79%), Ireland (74%), United States (72%), Czech Republic (69%), Germany (69%), and Portugal (69%); and higher than the OECD-average of 50% (OECD, 2019).

2 Context

Ethiopia is the second-most populous country in Africa with 100 million people, 10% of the population of Sub-Saharan Africa. Annual per-capita income is USD707 and Ethiopia ranks 173 out of 189 on the Human Development Index. Ethiopia has experienced rapid economic growth over the last decade, with an annual average growth rate in GDP of almost 10%. Industry, including construction, and services have accounted for the majority of recent economic growth, with substantially lower contributions from agriculture, demonstrating the growing importance of higher-skilled occupations in Ethiopia.

2.1 Education System

The education system in Ethiopia is structured as follows. Primary education consists of grades 1-8, where the 'standard' age of entry is 7. The 2015/16 national education statistics show that the gross enrolment ratio was 176%, indicating that many students start late or repeat years (Ministry of Education, Ethiopia, 2016). General secondary level consists of grades 9-10, after which there is a national examination that determines eligibility for higher secondary education. Higher secondary education consists of grades 11-12, after which there is a national examination that determines eligibility for entry into post-secondary education, based on a centrally determined threshold exam score. The 'standard' age of entry into university is 19 years of age although many enter later given the prevalence of late entry into the education system and average repetition rates of between 5% per grade.

To apply for entry into higher education, students note preferences for universities and broad fields of study. Students can prioritize their local university and over 95% of students attend a public university within their home district. This limited migration for education is related to constraints on voluntary migration across jurisdictions imposed by policy (Federal Democratic Republic of Ethiopia, 2003; Dorosh and Schmidt, 2010). The Ministry of Education allocates students to universities based on student preferences and exam scores, prioritizing students' first choices of university and field.

2.2 Expansion in Higher Education

The Ethiopian government rapidly expanded the number of public universities during the 2000s, from 8 to 33 between 1999 and 2013, shown in figure A1 (top-left) and table A1. The expansion is reflected in the growth in enrollment for both males and females and as a percentage of the working-age population (figure A1, top-right and bottom-left).

The locations of the new universities were determined by equity and political considerations, rather than economic. Universities were established based on ensuring regional equity (in terms of universities per region) and national unity, with the aim of providing broad access to higher education for local populations (such that no two universities would be 'too close') (Ashcroft and Rayner, 2011).[3] Such concerns over equity are apparent in the stated expansion plans, which illustrate that the planned intake capacities across localities were orthogonal to local economic conditions: the national Education Sector Development Program covering the period

[3]The national Education Sector Development Program covering 2005-2011 also states: "This expansion has taken into account equitable distribution of the higher learning institutions among the different regions of the county ... [and] to strengthen the decentralization process."

2005-2011 states that “each of these new institutions will have a capacity to enroll 9-10 thousand students” (Federal Democratic Republic of Ethiopia, 2005). The geographical spread of public universities is shown in figure A1 (bottom). Hence, the expansion in universities exhibits both time and spatial variation. I explore the correlates of university location below.

I find no evidence that the newer universities have a lower proportion of teachers with at least a bachelor’s degree and no evidence that newer universities have worse pupil-teacher ratios. However, there are concerns that quality is lower at these new universities in terms of equipment and other aspects of instruction quality (Ashcroft and Rayner, 2011; Semela, 2011).

3 Data and Identification

3.1 Data

The data used for the analysis come from three sources. I use the latest Labor Force Survey of 2013 for labor-market variables, including education levels, employment, and wages. I merge this with a hand-coded dataset of universities that I constructed, described below. For data on the population and rurality of districts, I use the latest census of 2007. I restrict the sample to individuals aged 21 to 65 – those of working age and old enough to attain a post-secondary education based on standard progression rates. I also restrict the sample to non-rural districts, comprising districts with less than one-third of residents as rural inhabitants (25th percentile).

To provide a sense of the differences between sample and non-sample districts, table A2 provides descriptive statistics for all districts in the LFS 2013, sample and non-sample districts. The table shows that sample districts are different from non-sample districts along several margins. Firstly, by construction sample districts are significantly less rural. They are also more likely to have a private university campus in the district (I discuss the issue of private universities in section 3.2). Secondly, the sample districts are more likely to attract migrants for work and education (I discuss migration further in sections 3.2 and 4.6). Thirdly, the populations in the sample districts are more educated at each level of education. Fourthly, the rates of paid employment and wages are significantly higher; sample districts are also much less agricultural and have higher rates of employment in manufacturing, construction and mining, public administration, and services. Hence, the findings in this paper can be interpreted as representative of non-rural parts of Ethiopia.

3.1.1 Geo-Located University Database

To construct the database of universities in the country, I extracted data on university establishments and enrollment from the Ministry of Education Statistics Annual Abstracts from the 1999/2000 to 2013/14 academic years. I geo-located each public university using the websites of the Ministry of Education and each university.[4] This step allowed universities to be matched to district administrations. Since Addis Ababa, the capital city, has had a university since 1950, it is difficult to explore the effect of its establishment on the labor market with recent data (individuals aged 21 at the time of the establishment of Addis Ababa University would be aged 84 in the 2013 Labor Force Survey). I therefore exclude Addis Ababa from the analysis. I also exclude the Ethiopian Civil Service College, the Defense University College, and Kotebe Teachers Education College, due to their specialized nature in training existing civil servants.

Regarding private universities, I constructed an exhaustive list of all accredited private universities from the Ministry of Education Statistics Annual Abstracts and the Ministry of Education website.[5] 90% of all enrolment in private universities is in Addis Ababa and only 2% of higher-education enrolment outside of Addis Ababa is within private institutions. For all private universities outside Addis Ababa, I matched the location of private universities to district administrations, including for secondary campuses of private universities.

3.2 Identification

The aim of the empirical analysis is to identify the causal effect of the university expansion on education attainment and labor-market outcomes. The identification strategy is a difference-in-differences framework in the spirit of Card and Lemieux (2001), Duflo (2001), and Atkin (2016). I exploit variation in access to higher education across districts and ages.

I categorise districts as 'exposed' and 'unexposed', depending on whether they have a public

[4] http://www.moe.gov.et/public-universities [Accessed: 03 March , 2020].
[5] http://www.moe.gov.et/web/guest/statistics [Accessed: 03 March, 2020].

university established within the sample time-frame. Districts are large (the average population is 143,000), well-defined labor markets, containing their own public services, public-employment opportunities, and service industries. There are also restrictions on accessing public services and public employment for non-residents of districts (Dorosh and Schmidt, 2010). I restrict the analysis to non-rural areas, as education and career opportunities from adolescence to mid-career are relatively similar across non-rural areas, but less-so between rural and non-rural areas, providing a conceptual justification for the parallel-trends assumption required for identification.

Table 1 presents tests of balance across the two types of districts on demographic characteristics that are not potential outcome variables, (exposed) districts with and (unexposed) districts without public universities, across a range of dimensions. The first panel shows variables from the main dataset used for the analysis. The two categories of districts are not statistically different (at the 10% level) in terms of population, rurality, the existence of a private university, age, gender, migration, and education. While the prevalence of private universities across districts appears high, this includes all off-site campuses and only accounts for 2% of all enrolment in higher education outside Addis Ababa.[6]

The table also shows low levels of migration for education reasons (2.8%), which limits the extent to which migration may bias the estimates of the effects of the expansion on education attainment. Since the levels of migration for work reasons are higher (9.9%), this could be a concern for the estimates on labor-market outcomes.[7] I discuss this further in section 4.6. The second panel of the table shows that district types are also balanced across the same set of demographic characteristics (age, gender, migration, and education) prior to the start of the expansion using the 1999 Labor Force Survey.

The second dimension of variation that I exploit is the age of individuals at the time that their nearest university was established, using the standard age of entry into higher education as reference. The key assumption for identification is of parallel trends across cohorts between exposed and unexposed districts. Figure 1 (top-left) shows the proportion of individuals with

[6] I also run a probit regression of whether the district has a public university on these variables and find that only population is a significant predictor of receiving a public university (marginal effect=0.21, $p = 0.00$), consistent with the stated expansion policy described in section 2.2.

[7] Migration for education and work reasons represent three-quarters of migration. Family and marriage reasons represent 15%. The remainder is due to land shortages and natural disasters.

Table 1: Balance Table: Exposed and Unexposed Districts

	Means and Standard Errors		Difference and P-Value
	(1)	(2)	(3)
	Exposed Districts	Unexposed Districts	T-Test Difference (1)-(2)
Primary Dataset			
District Characteristics			
Population (Millions)	0.16	0.11	0.05
	[0.03]	[0.02]	[0.12]
Proportion Population Rural	0.11	0.13	-0.02
	[0.02]	[0.03]	[0.57]
Private University	0.63	0.52	0.11
	[0.10]	[0.15]	[0.54]
Demographic Characteristics (2013)			
Age	35.02	35.35	-0.32
	[0.27]	[0.29]	[0.42]
Female	0.52	0.53	-0.01
	[0.01]	[0.00]	[0.43]
Migrant for Work	0.10	0.09	0.01
	[0.01]	[0.01]	[0.46]
Migrant for Education	0.03	0.02	0.01
	[0.00]	[0.00]	[0.34]
Primary Education	0.32	0.30	0.01
	[0.01]	[0.01]	[0.40]
Secondary Education	0.25	0.22	0.02
	[0.01]	[0.02]	[0.29]
Observations [Clusters]	26197 [49]	14026 [60]	
1999 Labor Force Survey			
Demographic Characteristics (1999)			
Age	36.54	36.38	0.16
	[0.36]	[0.25]	[0.72]
Female	0.54	0.53	0.01
	[0.01]	[0.01]	[0.38]
Migrant for Work	0.00	0.00	0.00
	[0.00]	[0.00]	[0.83]
Migrant for Education	0.01	0.01	-0.00
	[0.00]	[0.00]	[0.41]
Primary Education	0.28	0.23	0.05**
	[0.02]	[0.02]	[0.04]
Secondary Education	0.11	0.09	0.02
	[0.02]	[0.02]	[0.44]
Observations [Clusters]	11269 [30]	9734 [29]	

Notes: The table shows the means and standard deviations of district and demographic characteristics for districts with a public university (Exposed Districts, column 1) and districts without (Unexposed Districts, column 2). Column 3 shows the difference between Exposed and Unexposed Districts and the p-value of the T-test. The top panel uses the main dataset, based on the 2013 Labor Force Survey and 2007 Census. The bottom panel uses the 1999 Labor Force Survey. A migrant is defined as an individual that has changed district of residence within the last ten years. The T-statistic is based on standard errors clustered at the district level.

a post-secondary education for each age-at-establishment in districts with and without public universities. Figure 1 (top-right) presents the proportions using a local-fitted polynomial and 90% confidence intervals. Both figures provide evidence in favour of parallel trends across district types up to aged 22 at the time the nearest university was established. Figure 1 (bottom) plots the unconditional difference-in-difference estimates and 90% confidence intervals from the following regression:

$$HE_{i,a,d} = \alpha_a + T_d + \Sigma_{k=12}^{39}(T_d \times c_{ik})\beta_k + u_{i,a,d} \tag{1}$$

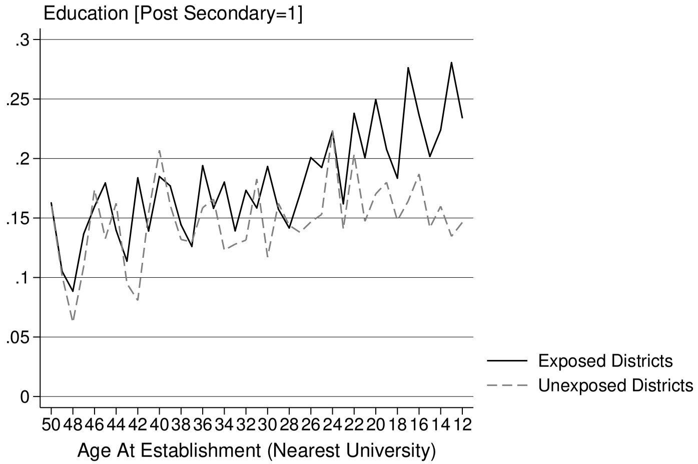
Where $HE_{i,a,d}$ is an indicator equal to one if individual i, aged a at the time the nearest university was established, in district d has a post-secondary education and zero otherwise. α_a are cohort fixed-effects; T_d is equal to one if the district has a public university; c_{ik} is a dummy equal to one if individual i was aged k at the time the nearest university was established. The figure shows that the coefficients on β_k become positive and significant (at the 10% level) after aged 20 at the time that the nearest university was established, but are approximately zero and not statistically different from zero for older cohorts.

Table A3 shows the results from a similar regression, with the addition of controls:

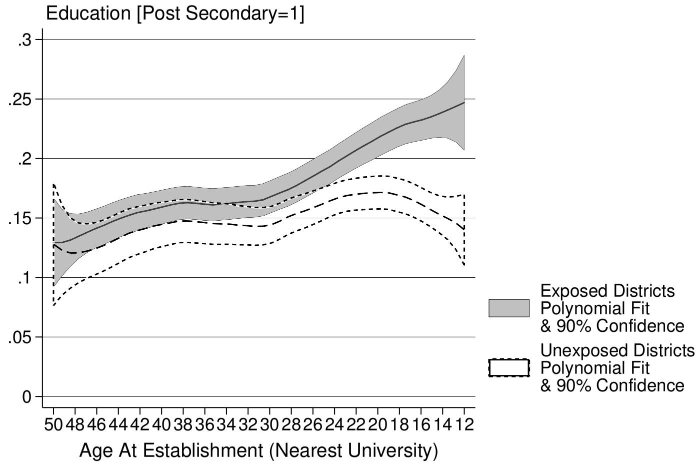
$$\begin{aligned} HE_{i,a,d} = \alpha_a + \phi_d + X_{i,a,d} + \Sigma_{k=12}^{39}(T_d \times c_{ik})\beta_k \\ + \Sigma_{k=12}^{39}(X_{i,a,d} \times c_{ik})\delta_k + u_{i,a,d} \end{aligned} \tag{2}$$

ϕ_d are district fixed-effects. $X_{i,a,d}$ are a set of district and individual controls, including the gender and current age (potential experience) of the individual, and the total population and rurality of the district (proportion of rural inhabitants).

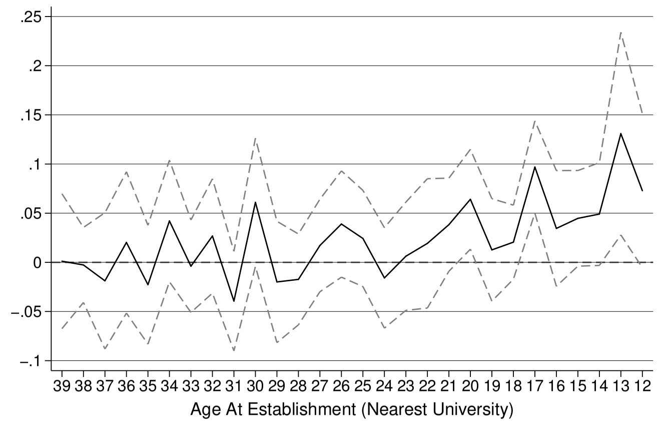
The table shows the results without the controls or district fixed-effects in column (1), with individual controls in column (2), with district controls in column (3), district fixed-effects in column (4), and interactions between controls and age-at-establishment dummies in column (5). The results in column (5) show significantly positive values for β_{17}, β_{18}, and β_{20}, and



(a) Raw Averages



(b) Polynomial



(c) Difference-in-Differences

Notes: The top-left figure shows the proportion of individuals with a post-secondary education for each age-at-establishment for districts with public universities (solid, black line) and districts without (gray, dashed line). The top-right figure shows the local fitted second-order polynomial and 90% confidence intervals for each age-at-establishment for districts with public universities (gray area, solid lines) and without (dashed lines). The bottom figure shows the coefficients (β_k) and 90% confidence intervals on the interactions between the district exposure variable (T_d) and each age-at-establishment indicator ($c_{i,k}$), from aged 39 to 12 at establishment, from a regression as per equation 1. The analysis includes all individuals aged up to 50 at establishment (those aged 40-50 form the omitted group).

Figure 1: Post-Secondary Education by Cohort and District Groups

positive and similarly large estimates for the range β_{17} to β_{21} ($\geq$ 3 percentage-points). The coefficients β_{22} to β_{29} are insignificantly different from zero across all columns.

Based on these findings and the standard age of entry into higher education, for the main analysis I group those aged 17-21 ('exposed cohorts') at the time their nearest university was established.[8] I compare these individuals to those aged 24-29 at the time the nearest university was established ('unexposed cohorts'). As shown in table A3 and in figure 1, there is no visible or significant effect on education attainment for the comparison cohorts as a result of the expansion. Furthermore, this comparison group is still relatively young so as not to have accrued vastly more human capital through work experience. As a further validity check, I run a control experiment, comparing those 24-29 at the time of establishment to those aged 34-40 at establishment, described below.

3.3 Higher Education Labor Markets and Public Employment

This section provides some background on the higher-education labor market. Figure A2 (left) shows the employment status by education level across 5 categories: paid employee; unpaid employee; self employed; non-employed and at school; non-employed and not at school. There is a noticeable distinction between those with a post-secondary education and those without. Most individuals with a tertiary education are paid employees (73%), with the next largest component non-employed and not at school (11%). This is in contrast to those with a secondary level of education, where only 36% are in paid employment, the next largest component is self employed (27%), followed by non-employed and not at school (23%).

Figure A2 (right) shows the industry of work by education level for the employed sample. There is again a stark contrast between those with a tertiary education and others. 55% of employed individuals with a tertiary education work in public administration, 30% in services and the majority of the remainder (12%) in manufacturing, mining or construction. Those with a secondary education, for comparison, work mostly in services (44%) followed by manufacturing, mining or construction (25%) and public administration (25%).

[8]Individuals over the standard age of entry are likely to face higher access costs to university, for example due to investments in job-specific human capital, family responsibilities or the difficulty of re-entering the education system (Meng and Gregory, 2002).

Figure A3 presents the proportion of employed individuals with a tertiary education in each occupational industry. The left sub-figure shows this for the whole sample and the right sub-figure restricts the sample to those aged 30 and above at the time their nearest university was established, for a 'baseline' picture. The figures show that public administration attracts the greatest proportion of individuals with a tertiary education (around 45%) relative to other industries (the next highest is services with approximately 15%). Figure A4 presents the average wages in each industry, for all individuals and for those aged 30 and above at establishment. Again, there is a noticeable distinction between public administration and others: those in public administration earn approximately 1.4 USD per hour, 0.5 USD more than in services, the next highest. This presents the public sector as highly attractive and restricted to the most educated, consistent with the existing literature (Finan et al., 2017).[9]

4 Results

4.1 Education Attainment

Motivated by the above, I run the following regression:

$$HE_{i,a,d} = \alpha_a + \phi_d + \beta(T_d \times I_{i,a,d}) + \delta(X_{i,a,d} \times I_{i,a,d}) + u_{i,a,d} \tag{3}$$

Where $I_{i,a,d}$ is an indicator equal to one if the individual was aged 17-21 at the time that the nearest university was established (exposed cohorts) and the omitted group is those aged 24-29 at establishment.

Table 2 presents the results. Column (1) shows the results without any of the controls in $X_{i,a,d}$. Column (2) includes individual controls; column (3) adds district controls; column (4) includes

[9]This link between higher education and public employment is apparent in other settings. The Worldwide Bureaucracy Indicators presents the proportion of employed tertiary-educated individuals working in the public sector. The average across the 101 countries in the sample is 39% and tends to be larger for poorer countries – the average is 45% among countries with a GDP per capita below $3,000 in PPP terms (Ethiopia's GDP per capita in PPP terms is $2,311). Ethiopia also has a relatively more attractive public sector than in other settings. Relative to the formal private sector, I find that the public-sector wage premium in Ethiopia is 13%, conditional on age, education indicators, and gender. In the World Bank's Worldwide Bureaucracy Indicators, the corresponding global average relative to the formal private sector is 7% and among low-income countries is 9.8%.

district fixed-effects; column (5) includes interactions between individual controls and $I_{i,a,d}$; column (6) adds interactions between district controls and $I_{i,a,d}$; column (7) adds interactions between region fixed-effects and $I_{i,a,d}$ to allow for region-specific trends. Column (8) uses an alternative measure of district exposure: the number of kilometers (in 100km) from the nearest public university.

Across the different specifications there is consistent evidence of a significant increase in higher-education attainment as a result of the expansion. In the main specification in column (7), the estimated effect is a 4-percentage-point increase in education attainment, significant at the 1% level. The proportion of individuals with a post-secondary education is 17% in the sample and 13% among those aged 30 and above at the time of establishment, a 'baseline' comparison. Relative to this baseline level, the expansion led to a 31% increase in higher-education attainment. Relative to the sample average and the national average (10%), the expansion led to a 24% increase and a 40% increase, respectively.

The results in column (8) show that access, in terms of distance, is a key driver of district-level exposure. For each 100km reduction in distance to the nearest public university, higher-education attainment increases by 4-percentage-points. The average distance to the nearest public university in the sample is 28km and the mean across all districts in the country (including rural districts) is 50km.

4.1.1 Control Experiment

As a further check of the identifying assumption of parallel trends across cohorts between the district types, I run a control experiment akin to Duflo (2001). I run the same regression as equation 3 but run the comparison between two older cohort groups that should be unaffected by the university expansion. I use those aged 24-29 at the time of establishment and those aged 34-40. Table A4, columns (1)-(7) show the results for the same set of specifications as table 2 but using this control set of cohort groups. The coefficients in all of these specifications are close to zero in magnitude and not statistically different from zero.

Table 2: Difference in Differences: 17-21 v 24-29

	(1) Baseline	(2) Individual Controls	(3) District Controls	(4) District Fixed Effects	(5) Individual x Cohort Trends	(6) District x Cohort Trends	(7) Region x Cohort Trends	(8) Distance Exposure
Exposed Cohort (17-21) x Exposed District	0.041*	0.039*	0.033	0.035*	0.040**	0.038*	0.042***	
	(0.024)	(0.022)	(0.020)	(0.018)	(0.019)	(0.020)	(0.015)	
Exposed Cohort (17-21) x Distance (100km)								-0.044***
								(0.0091)
Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	No	No	Yes	No	No	No	No	No
District Fixed Effects	No	No	No	Yes	Yes	Yes	Yes	Yes
Individual Controls x Exposed Cohort	No	No	No	No	Yes	Yes	Yes	Yes
District Controls x Exposed Cohort	No	No	No	No	No	Yes	Yes	Yes
Region FE x Exposed Cohort	No	No	No	No	No	No	Yes	Yes
Adjusted R-Squared	0.0074	0.020	0.035	0.056	0.057	0.056	0.058	0.058
Observations	14065	14065	14065	14065	14065	14065	14065	14065
Clusters	109	109	109	109	109	109	109	109
Dependent Variable Mean								0.19
Dependent Variable Std Dev								0.39

Notes: Clustered standard errors in parentheses. Standard errors clustered at the district level. The dependent variable is an indicator equal to one if the individual has a post-secondary education. Exposed Cohort (17-21) is equal to one if the individual is aged 17-21 at the time their nearest university was established. The comparison cohort is those aged 24-29 at establishment. Exposed District is equal to one if the district has a public university. Distance (100km) is a measure of the distance to the nearest public university in 100km units. Individual controls include gender and age. District controls include population and rurality.

4.1.2 Marriage, Migration, and Lower Levels of Education

Table A5 presents the effects of the expansion on alternative outcomes, using the main specification as per equation 3 and column (7) of table 2. The alternative outcomes used are: secondary-education attainment (column 1); an indicator for whether the individual is still at school (column 2); whether the individual has migrated for work (column 3); whether the individual has migrated for education (column 4); and whether the individual is married (column 5). The coefficients are close to zero and not statistically significantly different from zero in each of the columns, suggesting that the expansion did not (yet) directly affect lower levels of education attainment, or migration and marriage decisions.

4.1.3 Robustness to Individual Districts, Rurality and Cohort Definitions

To ensure that the main results are robust to influential district observations, for example not driven by districts that experience large one-off shocks to public employment, I conduct two robustness checks. First, I rerun the main specification, as per column (7) of table 2, but dropping each district in the sample in turn. Hence, I run 109 regressions and show the estimated coefficients and 90% confidence intervals for each in figure A5 (top-left). In each case, the estimate of the coefficient of interest (β) is significantly positive at the 10% level and the average magnitude of the coefficient is 0.042. In figure A5 (top-right), I undertake a similar exercise, re-running the main specification 100 times, but dropping two random districts in each case. For each case, the estimate of β is significantly positive and is 0.042 on average.

To ensure that the main results are robust to the rurality threshold set for the sample (less than one-third of residents as rural inhabitants), I rerun the main specification as per column (7) of table 2, for different rurality thresholds in figure A5 (bottom). The first specification presented in the plot restricts the sample to districts with only less than ten percent of residents as rural inhabitants (the left-most observation in the figure) and the last restricts the sample to districts with less than forty percent of the residents as rural inhabitants (right-most). In each case, the estimate of β is significantly positive and is 0.048 on average.

To ensure that the main results are robust to different comparison cohorts, in table A6, I re-run the main specification as per column (7) of table 2 using different comparison cohorts. The effect of the expansion on higher-education attainment is robust across different comparison cohort groups (24-29; 24-32; 24-36; and 24-40).

4.2 Employment, Wages and Earnings

I explore the direct effects of the expansion on labor-market outcomes in table 3. I run the same regression specification as equation 3 with an indicator for whether the individual is employed as the dependent variable in column (1), whether the individual is in paid employment in column (2), monthly earnings in USD in column (3), the inverse hyperbolic sine of monthly earnings in column (4), the hourly wage in USD in column (5) and the log hourly wage in column (6). For the estimates on earnings, I initially assume that the non-employed and self-employed receive a monthly wage of zero. Table A7 presents estimates under different scenarios, described below. The parallel-trend plots for employment and wages are presented in figure A6 and table A8 presents the results of the control experiment for these variables, suggesting no significant evidence of pre-trends.

The results in columns 1 and 2 suggest that the expansion had a large and significant effect on employment, driven by an increase in paid employment of 4 percentage-points (significant at the 10% level), which is a 14% increase relative to a 'baseline' of 28% for those aged 30 and above at the time of establishment in the sample. The results for earnings suggest an average increase of 2.56 USD per month in nominal terms (column 3) and an increase of 0.16 in inverse-hyperbolic-sine-transformed earnings (column 4), both significant at the 10% level. These effects translate to an increase in earnings of approximately 15%. The effects on hourly wages and log hourly wages are not significant at the 10% level (columns 5 and 6).

Table A7 shows the estimates on total earnings as per columns (2) and (3) in table 3 but under different assumptions for the earnings for the non-employed and self-employed. I impute the earnings of both the non-employed and self-employed as: zero for columns (1) and (8) of table A7; equal to the 1st percentile of positive earnings in columns (2) and (9); the 5th percentile of positive earnings in columns (3) and (10); the 10th percentile of positive earnings in columns (4) and (11); and the 25th percentile of positive earnings in columns (5) and (12). In columns (6),

Table 3: Employment, Earnings and Wages

	(1) Employed	(2) Paid Employed	(3) Monthly Earnings USD	(4) IHS Monthly Earnings	(5) Hourly Wage USD	(6) Log Hourly Wage
Exposed Cohort (17-21) x Exposed District	0.037*	0.037*	2.56*	0.16*	-0.0024	-0.0026
	(0.020)	(0.019)	(1.42)	(0.088)	(0.10)	(0.033)
Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes
District Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes
Region FE x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.13	0.096	0.077	0.10	0.11	0.21
Observations	14065	14065	14065	14065	5471	5471
Clusters	109	109	109	109	109	109
Dependent Variable Mean	0.74	0.34	16.5	1.41	1.01	0.56
Dependent Variable Std Dev	0.44	0.47	37.3	2.07	1.32	0.48

Notes: Clustered standard errors in parentheses. Standard errors clustered at the district level. The dependent variable is an indicator equal to one if the individual is currently employed in column (1); an indicator equal to one if the individual is in paid employment in column (2); the monthly earnings in USD in column (3); the inverse hyperbolic sine of monthly earnings in column (4); the hourly wage in USD in column (5); and the log hourly wage in column (6) Exposed Cohort (17-21) is equal to one if the individual is aged 17-21 at the time their nearest university was established. The comparison cohort is those aged 24-29 at establishment. Exposed District is equal to one if the district has a public university. Individual controls include gender and age. District controls include population and rurality.

(7), (13) and (14) I impute different earnings for the self-employed and non-employed. In these columns, I impute the self-employed earnings as the median of positive earnings and I impute the earnings of the non-employed as equal to zero in columns (6) and (13); and equal to the 25th percentile of earnings in column (7) and (14). Across the specifications, there is consistent evidence that the expansion had a positive effect on nominal earnings in USD. The results are less precise for inverse-hyperbolic-sine-transformed earnings, with significant effects (at the 10% level) only in columns (8) and (9), but no consistent evidence across the specifications of a significant increase.

Table A9 provides greater detail of the effects on employment status with finer categories of employment status as dependent variables in the regressions. The increase in employment and paid employment found in columns 1 and 2 of table 3 appear to mirror an almost-equal reduction in non-employment (column 4, table A9), rather than a shift from unpaid employment or self-employment (columns 2 and 3, table A9). These findings reflect the better employment prospects of higher-educated individuals, discussed in section 3.3.

4.3 Industry Selection

I investigate whether the expansion in higher education led to a change in the distribution of individuals across occupations by running a set of regressions as per equation 3 with industry-of-employment classifications as dependent variables. The results are presented in table A10. Columns (1) to (4) of table A10 present the results for the full sample of individuals and columns (5) to (8) present the results for the employed sample.

The results suggest an increase in the proportion of all individuals working in public administration (column 3), with a coefficient of 0.036, significant at the 5% level, and an insignificant decrease in the proportion working in services (column 4), with a coefficient of -0.02 ($p = 0.24$). A similar shift is also apparent when restricting the sample to employed individuals only, with an increase in the proportion in public administration (0.035, $p = 0.12$, column 7) and a decrease in services (-0.037, $p = 0.076$, column 8).

These results suggest that the higher-education expansion led to an occupational shift towards public administration, consistent with the descriptive evidence presented in section 3.3. There is no evidence that this increase in public employment is mechanically related to job opportunities through universities, as shown through the results of the control experiment in column 7 of table A8. In addition, employment in higher-education institutions makes up less than 2% of public employment according to the latest official figures (Ministry of Education, Ethiopia, 2016; Federal Civil Service Commission, Ethiopia, 2018). To further alleviate concerns that the effects of the expansion are driven by differential public-sector labor-demand trends in these areas, I find that the overall propensity of public employment is balanced, at 17%, across exposed and unexposed districts (difference=0.0052, $p = 0.75$); I find that the propensity of public employment for the exposed cohort is also balanced across the two types of districts (difference=0.022, $p = 0.16$), as is the public-employment propensity for the unexposed cohort (difference=0.0080, $p = 0.71$). I also find that the intensive margin of the expansion (in terms of original student-intake capacity) is weakly correlated with the proportion of individuals employed in public administration at 'baseline' (for those aged 30 and above at the time that their nearest university was established) in the district in figure A7. Lastly, I augment the probit regression described in section 3.2 including also each of the industry-of-employment classification variables. I find that the marginal effect on the indicator for employment in public administration is 0.017 ($p = 0.477$) and that a joint test of the industry-of-employment classification variables gives $\chi^2 = 2.9$, $p = 0.573$.

4.4 Returns to Higher Education

I use the plausibly exogenous variation in higher-education attainment, resulting from the expansion in universities, to estimate the returns to higher education in table 4. I run the following regressions:

$$Employed_{i,a,d} = \alpha_{1,a} + \beta_1 HE_{i,a,d} + \delta_1 X_{i,a,d} + \epsilon_{i,a,d} \tag{4}$$

$$w_{i,a,d} = \alpha_{2,a} + \beta_2 HE_{i,a,d} + \delta_2 X_{i,a,d} + u_{i,a,d} \quad if\ Employed_{i,a,d} = 1 \tag{5}$$

Where $w_{i,a,d}$ is the log hourly wage of the individual and $Employed_{i,a,d}$ is an indicator equal to one if the individual is in paid employment. In addition to gender and age, $X_{i,a,d}$ includes an indicator for whether the individual has less than primary education, hence the estimated returns are relative to those with at least primary education. Consistent with the above, I also control for cohort fixed-effects, district fixed-effects, and interactions between the district exposure variable and individual controls, district controls, and region fixed-effects.

In column (1) of table 4, I present the OLS estimate of β_1 from equation 4. The OLS estimate suggests a large employment return to higher education of 0.43 ($p = 0.00$). The OLS estimate may be biased through correlation between $\epsilon_{i,a,d}$ and $HE_{i,a,d}$; for example, if (unobserved) high-ability individuals are both more likely to attain a higher education and more job offers.

In column (2), I instrument for $HE_{i,a,d}$ with $T_d \times I_{i,a,d}$ and show the 2SLS coefficient for β_1. I show the F-statistic (7.7) and corresponding p-value (0.0065) from the first stage in square and curly brackets. The 2SLS estimate for β_1 is 0.90 ($p = 0.026$). This estimate suggests an almost doubling of paid-employment propensity as a result of higher-education attainment, consistent with the evidence presented in section 3.3.

In column (3), I present the OLS estimate of β_2. This estimate suggests a high wage return to higher education of 0.47 (p =0.00). Again, the OLS estimate may be biased through correlation between $u_{i,a,d}$ and $HE_{i,a,d}$ if, for example, high-ability individuals are both more likely to attain a higher education and demonstrate a higher level of productivity. In addition, the wage estimates rely on a selected sample into paid employment, which may also introduce bias in the OLS estimates (through correlation between $u_{i,a,d}$ and $\epsilon_{i,a,d}$). To deal with the latter, I estimate a Heckman selection model to explore the role of selection into paid employment on the estimates of the wage returns to higher education in column (4). I instrument the employment decision with $T_d \times I_{i,a,d}$. The top panel shows the results for β_2 (0.81, $p = 0.00$) and the bottom panel the marginal effects relating to β_1 (0.34, $p = 0.00$). These results suggest that higher-education attainment almost doubles hourly wages accounting for the selection into paid employment.

Given the strong link between higher education and public employment (section 3.3 and table A10), in columns (5) and (6) I estimate equation 5 with an indicator for whether the individual is employed in public administration as the dependent variable. Column (5) presents the OLS estimate for β_1 (0.32, $p = 0.00$) from such a regression and column (6) instruments for $HE_{i,a,d}$

Table 4: Returns to Higher Education

	(1) OLS Paid Employment	(2) 2SLS Paid Employment	(3) OLS Wage	(4) Heckman Selection Model	(5) OLS Public Employment	(6) 2SlS Public Employment
Log Hourly Wage						
Post Secondary	0.43***	0.90**	0.47***	0.81***	0.32***	0.86**
	(0.013)	(0.41)	(0.016)	(0.022)	(0.016)	(0.43)
		[7.67]				[7.67]
		{0.0066}				{0.0066}
Selection into Employment						
Post Secondary (Marginal Effect)				0.34***		
				(0.035)		
Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes
District Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes
Region FE x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	14065	14065	5471	13698	14065	14065
Clusters	109	109	109	109	109	109

Notes: Clustered standard errors in parentheses. Standard errors clustered at the district level. The dependent variable is an indicator equal to one if the individual is in paid employment in columns (1) and (2); the log hourly wage in columns (3) and (4); and an indicator equal to one if the individual is employed in public administration in columns (5) and (6). The square brackets in columns (2) and (6) present the F-statistic from the first stage of the 2SLS estimation and the curly brackets present the associated p-value. The bottom panel shows the marginal effects from the Probit estimation of selection into employment, the first stage of the Heckman estimation, in column (4). The top panel shows the results from the main equation or second stage in all columns. Individual controls include gender, age, and an indicator for whether the individual has less than primary education. District controls include population and rurality.

with $T_d \times I_{i,a,d}$ giving an estimate of 0.86 ($p = 0.048$). The results provide further evidence that public employment is a core driver of the returns to higher education.

4.5 Public Wages and Education Decisions

As the public sector is the major employer of higher-educated labor in the economy (section 3.3), I explore how the baseline public-wage premium and the baseline size of the public sector in the local labor market interact with education attainment.

To investigate the former, I first calculate the average log hourly wage of those working in public administration and aged 30 and above at the time their nearest university was established (cohorts unaffected by the expansion) for each district. I do the same for the other industries and use the difference between the average log hourly wage in public administration and the other industries as a proxy for the district-specific public-wage premium. The difference between the public-wage premium in districts with and without public universities is 0.077 and a T-test of equality gives a p-value of 0.29.

I then group districts into bins of the public-wage premium, using bins of size 0.05 and 0.1 and run the main difference-in-differences regression, as per equation 3 and column (7) of table 2 for each public-wage-premium bin.[10] Figure A8 (top panel) displays the results. I find that the expansion leads to a significant (at the 10% level) increase in education attainment only in areas where the public-wage premium is positive.

To investigate the latter, I undertake a similar exercise, but instead of grouping districts by the public-wage premium, I group districts by the baseline size of the public sector in the local labor market; that is, the proportion of individuals aged 30 and above at the time of establishment that are working in public administration. Figure A8 (bottom) displays the results, suggesting that the expansion leads to a significant increase in education attainment only in areas where the public sector is a relative more important employer.

Together, these results provide suggestive evidence that public-sector hiring and wage policies

[10] I do this by creating a variable that calculates the deviation from each bin value and interacting $(T_d \times I_{i,a,d})$ with this deviation, so that the coefficient on $(T_d \times I_{i,a,d})$ represents the education effect at the bin value. As the district dummies capture the average public-wage premium, I do not include district fixed-effects in these regressions, but include region fixed-effects.

mediate education decisions and the effectiveness of education policies in the economy.

4.6 Migration

Although there are no direct effects of the expansion on migration decisions (table A5), migration is non-zero, particularly for work reasons. Below, I explore the role of migration as a source of bias in the results.

I present the main analyses restricting the sample to non-migrants in table A11. The sample is restricted to individuals that have not migrated for work or education reasons in columns (1) to (6) and to individuals that have not migrated for any reason in columns (7) to (12). The results and conclusions of the effects of the expansion are consistent with those in tables 2 and 3, suggesting that endogenous migration decisions are not biasing the main estimates.

This finding is consistent with the notion that entry into the local public administration – the main entry point into a public-administration career and the major sector of employment for those with a higher education – is primarily restricted to local residents. This is part of a broader set of constraints to voluntary migration in this context (Federal Democratic Republic of Ethiopia, 2003; Dorosh and Schmidt, 2010)

5 Conclusion

This paper presents some of the first causal estimates of the labor-market returns to higher education in a low-income setting by exploiting a large-scale expansion of tertiary-education institutions in Ethiopia. The expansion leads to a significant 4-percentage-point increase in higher-education attainment, a 31% increase relative to baseline levels. Using measures of an individual's exposure to the expansion as instruments for higher-education attainment, I estimate the labor-market returns to higher education. I find that higher education almost doubles the propensity to enter paid employment and increases wages by 80%, in line with correlational estimates from similar contexts (Montenegro and Patrinos, 2014; Patrinos and Psacharopoulos, 2020). These estimates imply substantial private benefits from higher-education attainment in low-income contexts, such as Ethiopia.

This paper points to the relatively attractive pecuniary conditions and higher education requirements of public-sector employment – which are especially apparent in low-income countries (Finan et al., 2017) – as potential driving factors behind these high returns. Ethiopia demonstrates relatively high rates of public employment for tertiary-educated individuals and relatively attractive conditions of public employment as compared to other (low-income) settings. In my sample the proportion of employed tertiary-educated individuals in public employment is 55%, whereas in the World Bank's Worldwide Bureaucracy Indicators the global average is 39% and the average among low-income countries (those with less than $3,000 per capita in PPP terms; Ethiopia's GDP per capita in PPP terms is $2,311) is 45%. In addition, relative to the formal private sector, I find that the public-sector wage premium in Ethiopia is 13%, conditional on age, education indicators, and gender, while in the World Bank's Worldwide Bureaucracy Indicators, the corresponding global average relative to the formal private sector is 7% and among low-income countries is 9.8%. These factors could be behind the higher returns to higher education found in Ethiopia relative to other low-income economies (Peet et al., 2015).

The existence of an interaction between the effects of education policies and public-sector contracting policies implies that their designs should be considered simultaneously for optimal policy formation. While this paper can only provide suggestive evidence in this domain (without exogenous variation in public-sector hiring policies and wages), this is an important area for further research.

References

Altonji, J. G., Blom, E., and Meghir, C. (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *Annual Review of Economics*, 4(1):185–223.

Angrist, J. D. and Chen, S. H. (2011). Schooling and the Vietnam-era GI bill: Evidence from the draft lottery. *American Economic Journal: Applied Economics*, 3(2):96–118.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375.

Ashcroft, K. and Rayner, P. (2011). *Higher Education in Development: Lessons from Sub Saharan Africa*. IAP.

Atkin, D. (2016). Endogenous skill acquisition and export manufacturing in Mexico. *The American Economic Review*, 106(8):2046–2085.

Barrow, L. and Malamud, O. (2015). Is college a worthwhile investment? *Annual Review of Economics*, 7(1):519–555.

Becker, G. S. (1960). Underinvestment in college education? *The American Economic Review*, 50(2):346–354.

Bettinger, E., Gurantz, O., Kawano, L., Sacerdote, B., and Stevens, M. (2019). The long-run impacts of financial aid: Evidence from California's cal grant. *American Economic Journal: Economic Policy*, 11(1):64–94.

Black, D. A. and Smith, J. A. (2006). Estimating the returns to college quality with multiple proxies for quality. *Journal of Labor Economics*, 24(3):701–728.

Bleemer, Z. and Zafar, B. (2018). Intended college attendance: Evidence from an experiment on college returns and costs. *Journal of Public Economics*, 157:184–211.

Blundell, R., Dearden, L., Goodman, A., and Reed, H. (2000). The returns to higher education in Britain: evidence from a British cohort. *The Economic Journal*, 110(461):F82–F99.

Bound, J. and Turner, S. (2002). Going to war and going to college: Did World War II and the

GI bill increase educational attainment for returning veterans? *Journal of Labor economics*, 20(4):784–815.

Brand, J. E. and Xie, Y. (2010). Who benefits most from college? evidence for negative selection in heterogeneous economic returns to higher education. *American Sociological Review*, 75(2):273–302.

Card, D. (1993). Using geographic variation in college proximity to estimate the return to schooling. *National Bureau of Economic Research Working Paper.*

Card, D. (1999). The causal effect of education on earnings. In *Handbook of Labor Economics*, volume 3, pages 1801–1863. Elsevier.

Card, D. and Lemieux, T. (2001). Can falling supply explain the rising return to college for younger men? A cohort-based analysis. *The Quarterly Journal of Economics*, 116(2):705–746.

Carnevale, A. P., Cheah, B., and Rose, S. J. (2011). The college pay off. Technical report, Georgetown University Center on Education and the Workforce.

Dai, F., Cai, F., and Zhu, Y. (2018). Returns to higher education in China: Evidence from the 1999 higher education expansion using fuzzy regression discontinuity. *IZA Discussion Paper.*

Dorosh, P. and Schmidt, E. (2010). The rural-urban transformation in Ethiopia. *ESSP II Working Paper.*

Duflo, E. (2001). Schooling and labor market consequences of school construction in indonesia: Evidence from an unusual policy experiment. *American Economic Review*, 91(4):795–813.

Duflo, E. (2004). The medium run effects of educational expansion: Evidence from a large school construction program in indonesia. *Journal of Development Economics*, 74(1):163–197.

Federal Civil Service Commission, Ethiopia (2018). National civil service human resource statistics 2017/18.

Federal Democratic Republic of Ethiopia (2003). Rural development policy and strategies. Technical report, Federal Democratic Republic of Ethiopia.

Federal Democratic Republic of Ethiopia (2005). Education sector development program iii. Technical report, Federal Democratic Republic of Ethiopia.

Filmer, D. (2007). If you build it, will they come? School availability and school enrolment in 21 poor countries. *The Journal of Development Studies*, 43(5):901–928.

Finan, F., Olken, B. A., and Pande, R. (2017). The personnel economics of the developing state. In *Handbook of Economic Field Experiments*, volume 2, pages 467–514. Elsevier.

Goldin, C. and Katz, L. F. (2007). The race between education and technology: The evolution of us educational wage differentials, 1890 to 2005. *National Bureau of Economic Research*.

Hout, M. (2012). Social and economic returns to college education in the United States. *Annual Review of Sociology*, 38:379–400.

Kane, T. J. and Rouse, C. E. (1995). Labor-market returns to two-and four-year college. *The American Economic Review*, 85(3):600–614.

Lemieux, T. and Card, D. (2001). Education, earnings, and the 'Canadian GI Bill'. *Canadian Journal of Economics/Revue canadienne d'économique*, 34(2):313–344.

Li, S., Whalley, J., and Xing, C. (2014). China's higher education expansion and unemployment of college graduates. *China Economic Review*, 30:567–582.

Liu, V. Y., Belfield, C. R., and Trimble, M. J. (2015). The medium-term labor market returns to community college awards: Evidence from North Carolina. *Economics of Education Review*, 44:42–55.

Machin, S. and McNally, S. (2007). Tertiary education systems and labour markets. *Education and Training Policy Division, OECD*, 6.

Meng, X. and Gregory, R. G. (2002). The impact of interrupted education on subsequent educational attainment: A cost of the Chinese cultural revolution. *Economic Development and Cultural Change*, 50(4):935–959.

Ministry of Education, Ethiopia (2016). Education statistics annual abstract 2015/16.

Montenegro, C. E. and Patrinos, H. A. (2014). *Comparable estimates of returns to schooling around the world*. The World Bank.

OECD (2019). *Education at a Glance 2019*. OECD Paris.

Oppedisano, V. (2014). Higher education expansion and unskilled labour market outcomes. *Economics of Education Review*, 40:205–220.

Oreopoulos, P. and Petronijevic, U. (2013). Making college worth it: A review of research on the returns to higher education. *National Bureau of Economic Research*.

Ou, D. and Zhao, Z. (2016). Higher education expansion and labor market outcomes for young college graduates.

Patrinos, H. A. and Psacharopoulos, G. (2020). Returns to education in developing countries. *The Economics of Education*, pages 53–64.

Peet, E. D., Fink, G., and Fawzi, W. (2015). Returns to education in developing countries: Evidence from the living standards and measurement study surveys. *Economics of Education Review*, 49:69–90.

Rumberger, R. W. and Thomas, S. L. (1993). The economic returns to college major, quality and performance: A multilevel analysis of recent graduates. *Economics of Education Review*, 12(1):1–19.

Rzepka, S. (2018). Labor market returns to college education with vocational qualifications. *Education Economics*, 26(4):411–431.

Semela, T. (2011). Breakneck expansion and quality assurance in Ethiopian higher education: Ideological rationales and economic impediments. *Higher Education Policy*, 24(3):399–425.

Stanley, M. (2003). College education and the midcentury GI bills. *The Quarterly Journal of Economics*, 118(2):671–708.

The Hamilton Project (2017). The education wage premium contributes to wage inequality. Accessed: 25th February, 2020.

UNESCO (2020). Unesco institute for statistics. Accessed: 03 March, 2020.

Weisbrod, B. A. and Karpoff, P. (1968). Monetary returns to college education, student ability, and college quality. *The Review of Economics and Statistics*, pages 491–497.

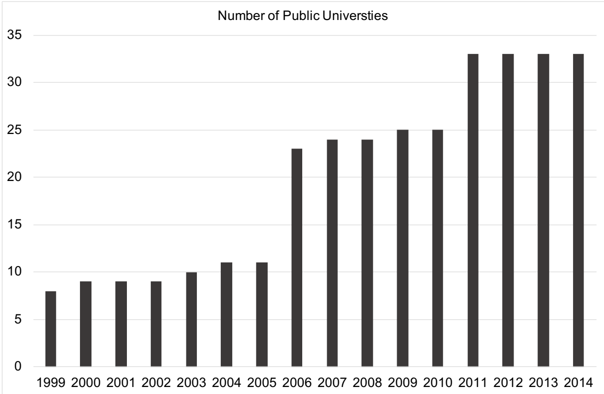
World Bank (2020a). World bank development indicators. Accessed: 25 February, 2020.

World Bank (2020b). Worldwide bureaucracy indicators. Accessed: 03 March, 2020.

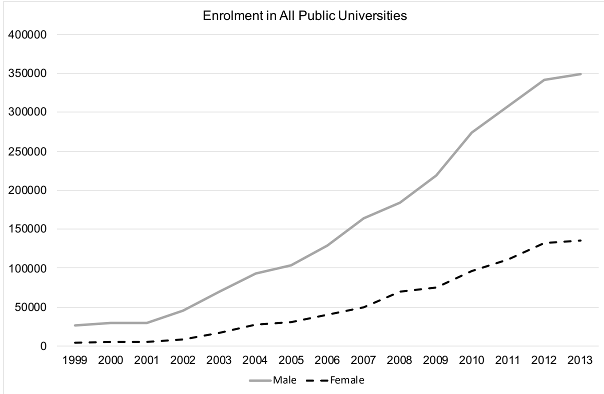
Xing, C., Yang, P., and Li, Z. (2018). The medium-run effect of China's higher education expansion on the unemployment of college graduates. *China Economic Review*, 51:181–193.

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4):711–754.

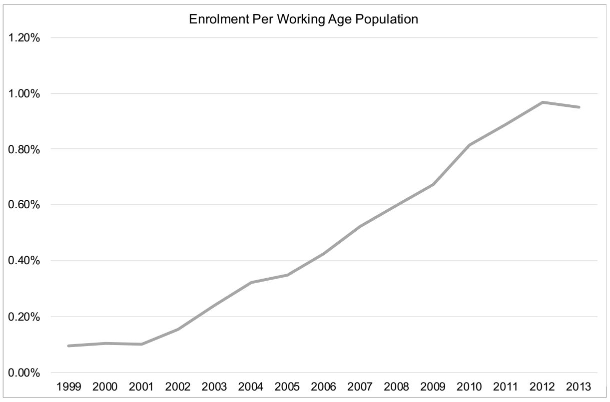
ONLINE APPENDIX



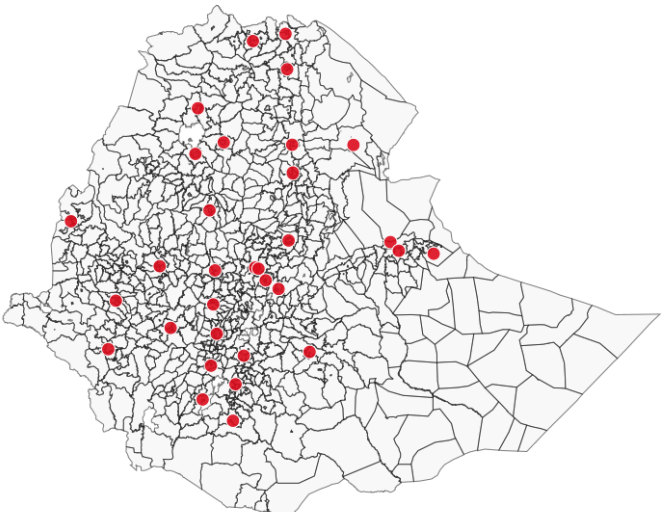
(a) Number of Universities



(b) Enrolment



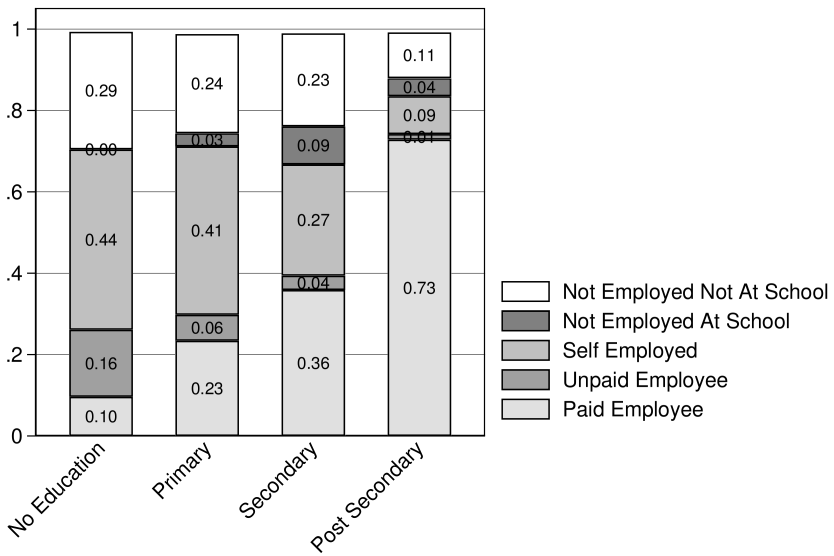
(c) Enrolment per Working Age Population



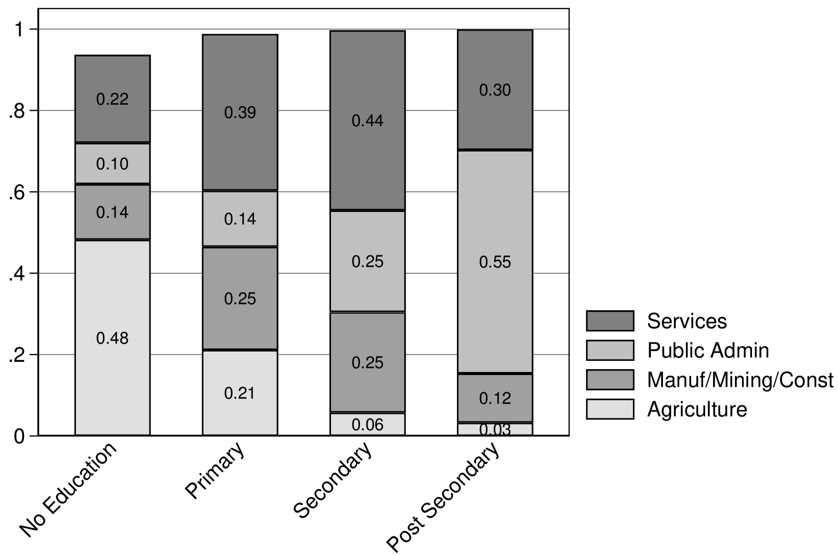
(d) Locations

Notes: The top-left figure shows the total number of public universities between 1999 and 2014 academic years. The top-right figure shows the enrollment in public universities by gender for the period 1999 to 2013 academic years. The botton-right figure shows the total enrollment in public universities as a percentage of the working age population for 1999 to 2013. The black dashed line shows female enrollment and the solid grey line shows male enrollment. The bottom figure shows the location of each public university in Ethiopia as of 2016. Source: Ministry of Education, Ethiopia; the working-age population data is sourced from the World Bank.

Figure A1: Number of Public Universities and Enrollment



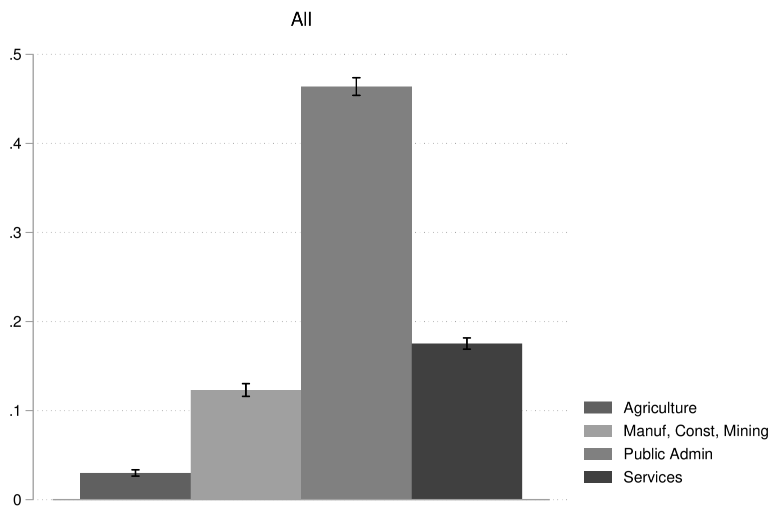
(a) Employment Status



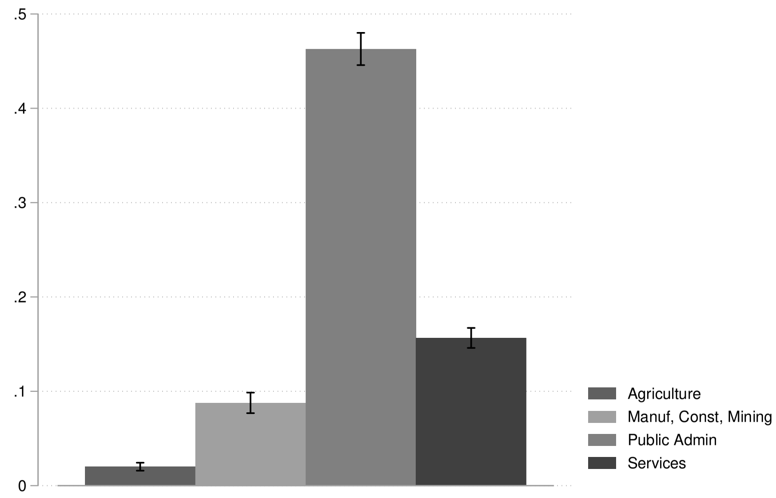
(b) Industry of Work

Notes: The left figure shows the employment status for all individuals in the sample by education group. The right figure shows the industry of work by education level for all employed individuals in the sample. Source: 2013 Labor Force Survey.

Figure A2: Employment Status and Industry of Work by Education



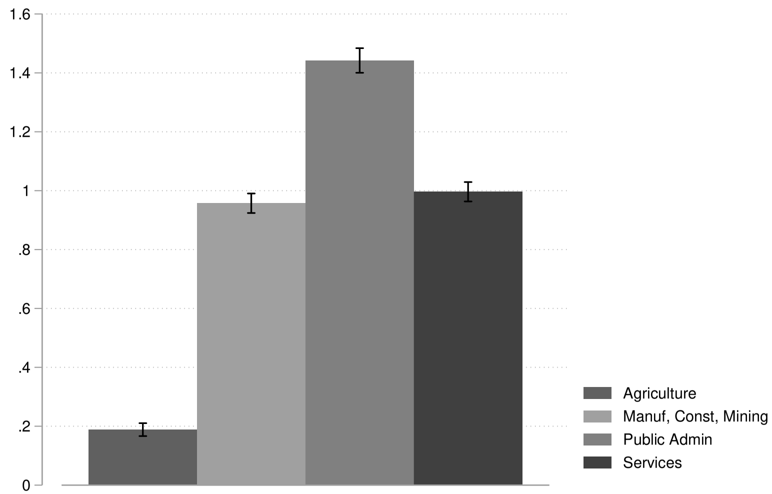
(a) All



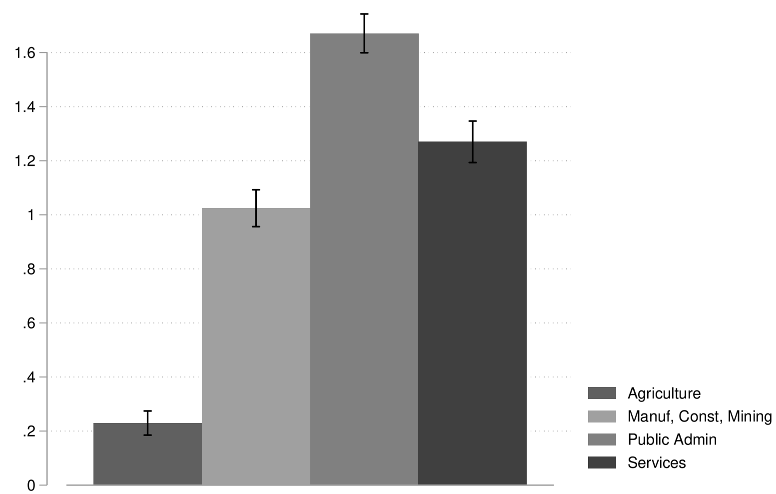
(b) "Unexposed" Cohorts Only – Above 30 At Establishment

Notes: The left figure shows the proportion of individuals (bars) and 90% confidence intervals (capped lines) with a post-secondary education by industry of work, using the full sample of individuals aged between 21 and 65. The right figure restricts the sample to only those aged 30 and above at the time that their nearest university was established. Source: 2013 Labor Force Survey and Geo-Located University Database.

Figure A3: Proportion with Higher Education by Industry of Work



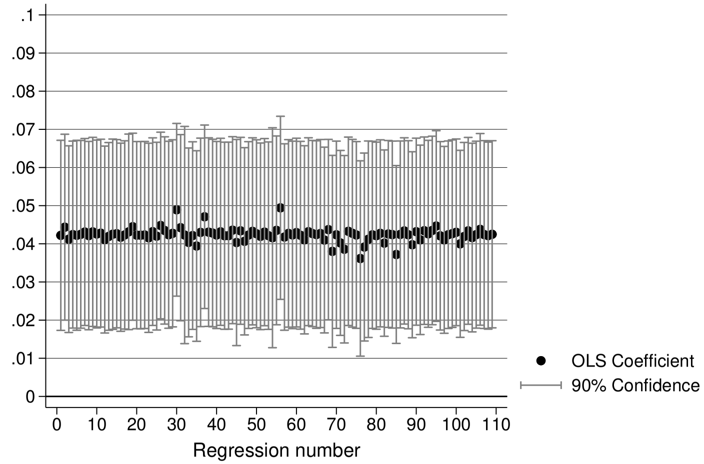
(a) All



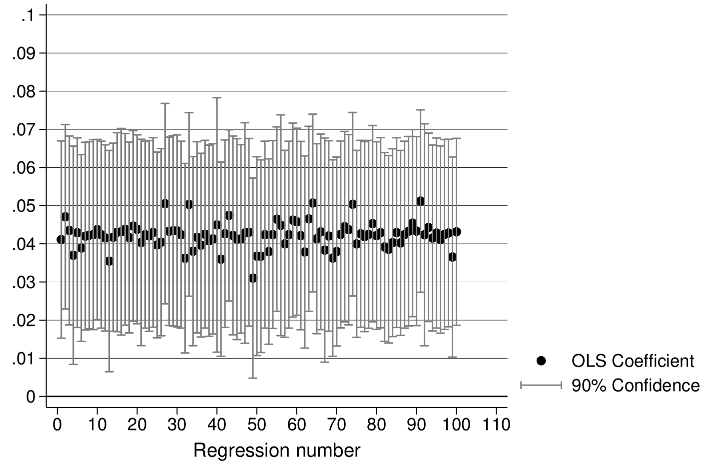
(b) "Unexposed" Cohorts Only – Above 30 At Establishment

Notes: The left figure shows the average hourly wage in USD (bars) and 90% confidence intervals (capped lines) by industry of work, using the full sample of individuals aged between 21 and 65. The right figure restricts the sample to only those aged 30 and above at the time that their nearest university was established. Source: 2013 Labor Force Survey and Geo-Located University Database.

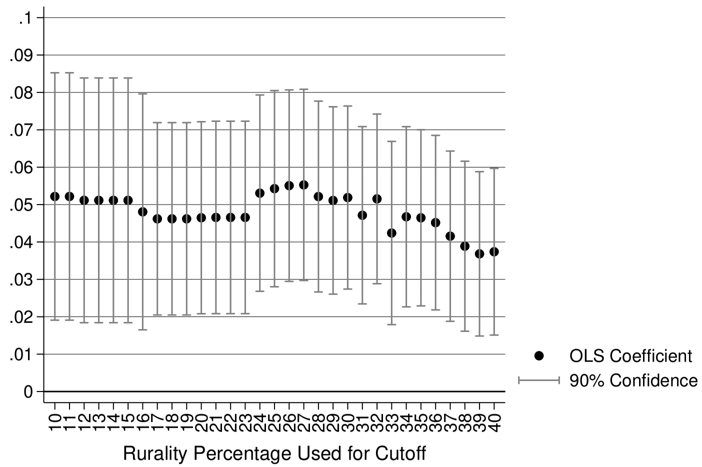
Figure A4: Average Wages by Industry of Work



(a) Dropping Each District



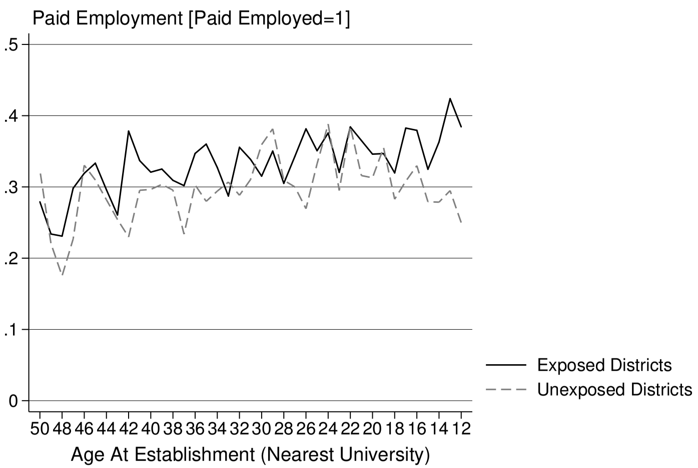
(b) Dropping Random District Pairs



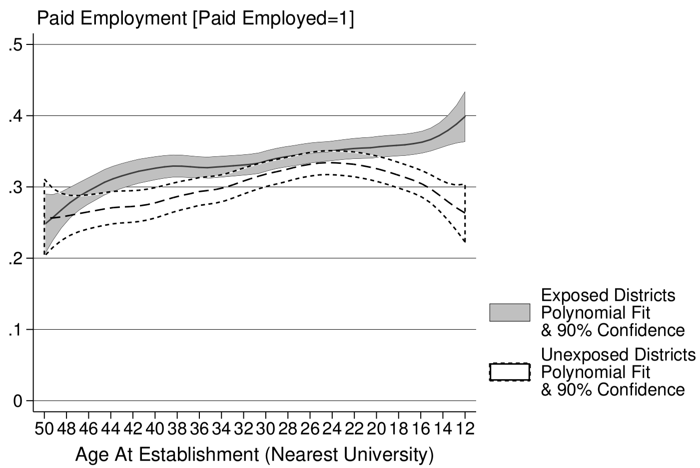
(c) Varying the Rurality of the Sample

Notes: The top-left figure shows the OLS coefficients (black circles) and 90% confidence intervals (gray lines) from a set of regressions, where the regression specification is as per column (7) of table 2 but each single district from the sample is dropped in each regression. The top-right figure shows the OLS coefficients (black circles) and 90% confidence intervals (gray lines) from a set of regressions, where the regression specification is as per column (7) of table 2 but random pairs of districts are dropped from the sample in each regression. The bottom-right figure shows the OLS coefficients (black circles) and 90% confidence intervals (gray lines) from a set of regressions, where the regression specification is as per column (7) of table 2 but alters the study sample to include districts with less than different percentages of rural inhabitants, as indicated by the x-axis labels.

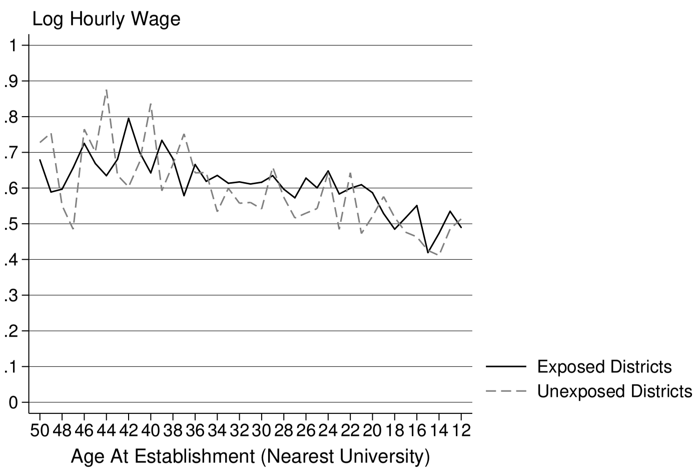
Figure A5: Robustness to Robustness to Individual Districts and Rurality Thresholds



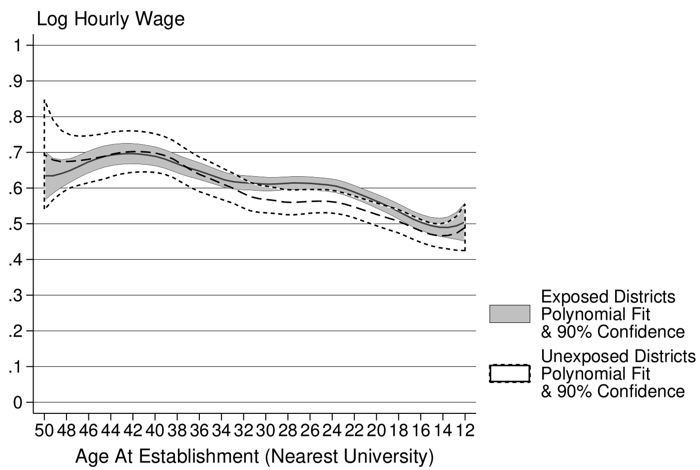
(a) Raw Averages: Paid Employment



(b) Polynomial: Paid Employment



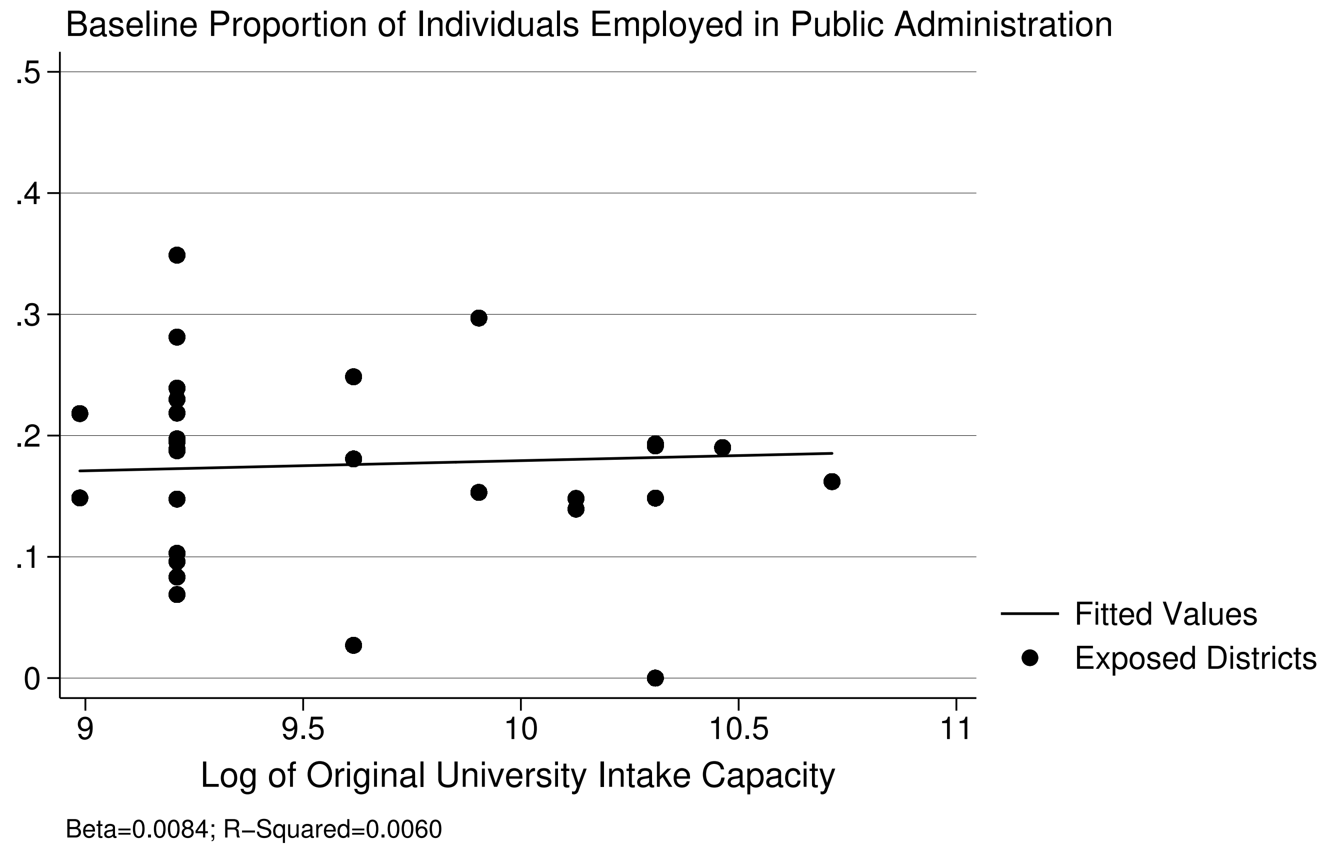
(c) Raw Averages: Log Hourly Wage



(d) Polynomial: Log Hourly Wage

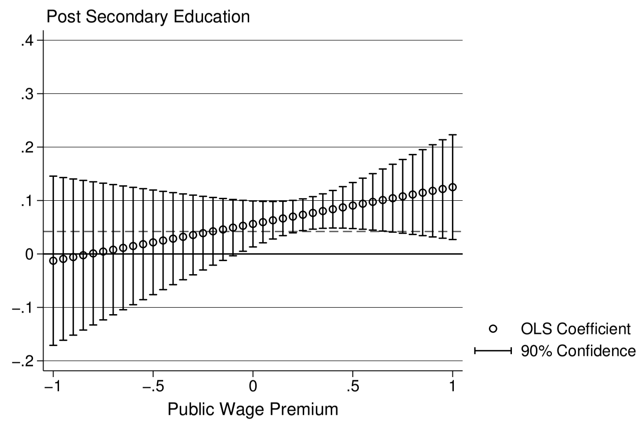
Notes: The top-left figure shows the proportion of individuals in paid employment for each age-at-establishment for districts with public universities (solid, black line) and districts without (gray, dashed line). The top-right figure shows the local fitted second-order polynomial of the proportion of individuals in paid employment and 90% confidence intervals for each age-at-establishment for districts with public universities (gray area, solid lines) and without (dashed lines). The bottom-left figure shows the average log hourly wage for each age-at-establishment for districts with public universities (solid, black line) and districts without (gray, dashed line). The bottom-right figure shows the local fitted second-order polynomial of average log hourly wages and 90% confidence intervals for each age-at-establishment for districts with public universities (gray area, solid lines) and without (dashed lines).

Figure A6: Employment and Wages by Cohort and District Groups

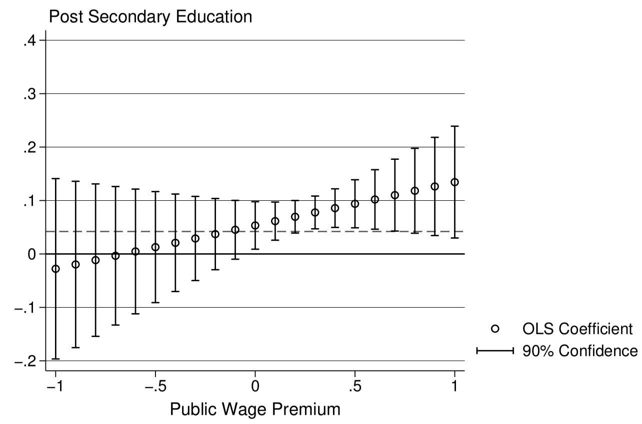


Notes: The figure shows a scatter plot and line of best fit of the proportion of those aged 30 and above at the time that their nearest university was established that are employed in public administration (y-axis) and the log of the original student-intake capacity for each district with a public university (x-axis). The figure also presents the OLS coefficient and R-Squared from an unconditional regression of the y-variable on the x-variable.

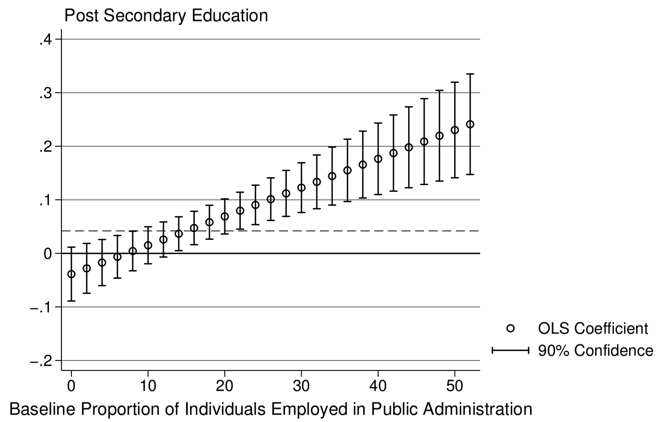
Figure A7: Correlation between Original Student-Intake Capacity and the Baseline Proportion of Individuals Employed in Public Administration



(a) Public Wage Premium: Bin Size=0.05



(b) Public Wage Premium: Bin Size=0.1



(c) Proportion Working in Public Admin

Notes: The top figure shows the coefficient β from a regression as per equation 3, with region fixed-effects instead of district fixed-effects, for each bin of the public-wage premium, measured as the average log-wage difference between the public sector and private sector in the district for those aged 30 and above at the time that their nearest university was established.. The top-left figure groups districts by their public-wage premium using bins of size 0.05 and the top-right figure uses bins of size 0.1. The bottom figure show the coefficient β from the same regression, for each bin of the size of the public sector in the local labor market, measured as the proportion of individuals working in the public sector among those aged 30 and above at the time that their nearest university was established. I create a variable that calculates the deviation from each bin value and interact $(T_d \times I_{i,a,d})$ with this deviation, so that the coefficient on $(T_d \times I_{i,a,d})$ represents the education effect at the bin value, which are displayed in the graphs. The horizontal gray-dashed line represents the point estimate of β in column (7) of table 2.

Figure A8: Education Effects by Baseline Public Wage Premium and Size of Local Public Sector

Table A1: Details of Public Universities

University	Region	Academic Year Established	Year of First Graduates	Special Purpose University
Adama University	Oromiya	2006	2009	No
Addis Ababa University	Addis Ababa	1950	1953	No
Adigrat University	Tigray	2011	2014	No
Aksum University	Tigray	2006	2009	No
Ambo University	Oromiya	2009	2012	No
Arba Minch University	SNNP	2004	2007	No
Assosa university	Benishangul Gumuz	2011	2014	No
Bahir Dar University	Amhara	1999	2002	No
Bule Hora University	Oromiya	2011	2014	No
Debre Birhan University	Amhara	2006	2009	No
Debre Markos University	Amhara	2006	2009	No
Debre Tabor University	Amhara	2011	2014	No
Dilla University	Amhara	2006	2009	No
Dire Dawa University	Dire Dawa	2006	2009	No
Ethiopian Civil Service College	Addis Ababa	1995	1998	Yes
FDRE Defense University College	Oromiya	2000	2003	Yes
Gondar University	Amhara	2003	2006	No
Haramaya University	Oromiya	1995	1998	No
Hawassa University	SNNP	1999	2002	No
Jijiga University	Somali	2006	2009	No
Jimma University	Oromiya	1999	2002	No
Kotebe Teachers Education College	Addis Ababa	1979	1982	Yes
Madawalabu University	Oromiya	2006	2009	No
Mekelle University	Tigray	1999	2002	No
Metu University	Oromiya	2011	2014	No
Mizan-Tepi University	SNNP	2006	2009	No
Samara University	Afar	2007	2010	No
Wachemo University	SNNP	2011	2014	No
Wolaita Sodo University	SNNP	2006	2009	No
Woldia University	Amhara	2011	2014	No
Wolkite University	SNNP	2011	2014	No
Wollega University	Oromiya	2006	2009	No
Wollo University	Amhara	2006	2009	No

Notes: The table shows the list of public universities as of 2014, based on official data from the Ministry of Education, Ethiopia.

Table A2: Sample and Non-Sample Districts

	Means and Standard Errors			Differences and P-Values
	(1)	(2)	(3)	(4)
	All Districts in LFS 2013	Non-Sample Districts	Sample Districts	T-Test Difference (2)-(3)
Population (Millions)	0.13	0.12	0.14	-0.03
	0.079	[0.00]	[0.02]	[0.21]
Proportion Population Rural	0.30	0.45	0.12	0.33
	0.20	[0.00]	[0.02]	[0.00]
Private University	0.28	0.01	0.60	-0.59
	0.45	[0.00]	[0.08]	[0.00]
Age	35.9	36.61	35.14	1.48
	11.5	[0.11]	[0.21]	[0.00]
Female	0.52	0.52	0.52	-0.01
	0.50	[0.00]	[0.00]	[0.21]
Migrant for Work	0.070	0.05	0.10	-0.05
	0.26	[0.00]	[0.01]	[0.00]
Migrant for Education	0.016	0.01	0.03	-0.02
	0.12	[0.00]	[0.00]	[0.00]
Primary Education	0.30	0.29	0.31	-0.02
	0.46	[0.01]	[0.01]	[0.03]
Secondary Education	0.15	0.08	0.24	-0.16
	0.36	[0.00]	[0.01]	[0.00]
Post Secondary	0.10	0.05	0.17	-0.13
	0.30	[0.00]	[0.01]	[0.00]
Employed: Paid Employee	0.20	0.10	0.32	-0.22
	0.40	[0.01]	[0.02]	[0.00]
Employed: Unpaid	0.18	0.27	0.08	0.19
	0.38	[0.01]	[0.01]	[0.00]
Employed: Self Employed	0.42	0.49	0.33	0.16
	0.49	[0.01]	[0.01]	[0.00]
Not Employed	0.20	0.14	0.27	-0.13
	0.40	[0.01]	[0.01]	[0.00]
Hourly Wage USD	0.65	0.32	1.01	-0.69
	1.31	[0.03]	[0.05]	[0.00]
Industry: Agriculture	0.40	0.60	0.16	0.45
	0.49	[0.01]	[0.02]	[0.00]
Industry: Manuf; Const; Min.	0.10	0.07	0.14	-0.08
	0.30	[0.00]	[0.01]	[0.00]
Industry: Public Administration	0.12	0.07	0.17	-0.10
	0.32	[0.00]	[0.01]	[0.00]
Industry: Services	0.16	0.09	0.24	-0.15
	0.37	[0.01]	[0.01]	[0.00]

Notes: The table shows the means and standard deviations for all districts districts in the 2013 Labor Force Survey column (1); for non-sample districts in column (2); and sample districts in column (3). Columns (4) shows the difference between sample and non-sample districts and the p-value for the T-tests of the difference in brackets. The T-statistic is based on standard errors clustered at the district level.

Table A3: Education Effects: Cohort by Cohort

	(1) Baseline	(2) Individual Controls	(3) District Controls	(4) District Fixed Effects	(5) Controls x Cohort Trends
12 At Establishment x Exposed District	0.073	0.068	0.014	0.0034	0.011
	(0.047)	(0.048)	(0.030)	(0.031)	(0.036)
13 At Establishment x Exposed District	0.13**	0.13**	0.098**	0.064	0.027
	(0.062)	(0.061)	(0.045)	(0.051)	(0.073)
14 At Establishment x Exposed District	0.049	0.047	0.036	0.019	0.045
	(0.031)	(0.031)	(0.026)	(0.026)	(0.030)
15 At Establishment x Exposed District	0.045	0.045	0.046*	0.044*	0.054*
	(0.029)	(0.028)	(0.024)	(0.025)	(0.028)
16 At Establishment x Exposed District	0.035	0.036	0.036	0.020	0.0093
	(0.035)	(0.034)	(0.034)	(0.033)	(0.035)
17 At Establishment x Exposed District	0.097***	0.083***	0.073***	0.074***	0.070***
	(0.028)	(0.025)	(0.022)	(0.022)	(0.026)
18 At Establishment x Exposed District	0.021	0.020	0.023	0.021	0.034*
	(0.023)	(0.021)	(0.017)	(0.017)	(0.020)
19 At Establishment x Exposed District	0.013	0.011	0.012	0.0092	0.025
	(0.031)	(0.031)	(0.030)	(0.030)	(0.026)
20 At Establishment x Exposed District	0.064**	0.059*	0.063*	0.066**	0.086***
	(0.031)	(0.031)	(0.033)	(0.033)	(0.025)
21 At Establishment x Exposed District	0.038	0.034	0.040	0.030	0.035
	(0.029)	(0.028)	(0.026)	(0.027)	(0.029)
22 At Establishment x Exposed District	0.019	0.014	0.0064	0.0063	-0.014
	(0.040)	(0.037)	(0.035)	(0.034)	(0.033)
23 At Establishment x Exposed District	0.0062	0.0021	0.014	0.019	0.028
	(0.033)	(0.033)	(0.034)	(0.032)	(0.030)
24 At Establishment x Exposed District	-0.016	-0.015	-0.016	-0.037	-0.043
	(0.031)	(0.032)	(0.031)	(0.028)	(0.031)
25 At Establishment x Exposed District	0.024	0.020	0.025	0.032	0.041
	(0.029)	(0.029)	(0.029)	(0.028)	(0.029)
26 At Establishment x Exposed District	0.039	0.037	0.042	0.034	0.045
	(0.033)	(0.032)	(0.031)	(0.029)	(0.029)
27 At Establishment x Exposed District	0.017	0.0064	0.0017	0.00097	-0.00055
	(0.028)	(0.029)	(0.028)	(0.026)	(0.032)
28 At Establishment x Exposed District	-0.017	-0.016	-0.0034	0.00037	0.0076
	(0.028)	(0.025)	(0.024)	(0.023)	(0.025)
29 At Establishment x Exposed District	-0.020	-0.021	-0.016	-0.012	-0.028
	(0.037)	(0.037)	(0.036)	(0.035)	(0.036)
Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	Yes	Yes
District Controls	No	No	Yes	No	No
District Fixed Effects	No	No	No	Yes	Yes
Individual Controls x Cohort FE	No	No	No	No	Yes
District Controls x Cohort FE	No	No	No	No	Yes
Adjusted R-Squared	0.0098	0.027	0.050	0.067	0.064
Observations	34869	34869	34869	34869	34869
Clusters	109	109	109	109	109
Dependent Variable Mean					0.17
Dependent Variable Std Dev					0.38

Notes: Clustered standard errors in parentheses. Standard errors clustered at the district level. The dependent variable is an indicator equal to one if the individual has a post-secondary education. The table presents the results from a regression as per equation 1. The table presents the coefficients (β_k) on the interactions between the district exposure variable (T_d) and each age-at-establishment indicator ($c_{i,k}$). The regression includes all interactions from aged 12 to 39 at establishment; only the coefficients from aged 12 to 29 are shown for presentation purposes. The analysis includes the full sample of individuals aged up to 50 at the time of establishment. Individual controls include gender and age. District controls include population and rurality.

Table A4: Control Experiment: 24-29 v 34-40

	(1) Baseline	(2) Individual Controls	(3) District Controls	(4) District Fixed Effects	(5) Individual x Cohort Trends	(6) District x Cohort Trends	(7) Region x Cohort Trends
Older Cohort (24-29) x Exposed District	0.0048	0.0058	0.0091	0.0075	0.0095	0.0083	0.00065
	(0.016)	(0.016)	(0.016)	(0.017)	(0.017)	(0.018)	(0.020)
Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	No	Yes	Yes	Yes	Yes	Yes	Yes
District Controls	No	No	Yes	No	No	No	No
District Fixed Effects	No	No	No	Yes	Yes	Yes	Yes
Individual Controls x Older Cohort	No	No	No	No	Yes	Yes	Yes
District Controls x Older Cohort	No	No	No	No	No	Yes	Yes
Region FE x Older Cohort	No	No	No	No	No	No	Yes
Adjusted R-Squared	0.0035	0.025	0.059	0.081	0.082	0.081	0.082
Observations	10287	10287	10287	10287	10287	10287	10287
Clusters	109	109	109	109	109	109	109
Dependent Variable Mean							0.17
Dependent Variable Std Dev							0.37

Notes: Clustered standard errors in parentheses. Standard errors clustered at the district level. The dependent variable is an indicator equal to one if the individual has a post-secondary education. Older Cohort (24-29) is equal to one if the individual is aged 24-29 at the time their nearest university was established. The comparison cohort is those aged 34-40 at establishment. Exposed District is equal to one if the district has a public university. Individual controls include gender and age. District controls include population and rurality.

Table A5: Lower-Level Education Outcomes, Migration and Marriage

	(1) Secondary Education	(2) At School	(3) Migrant for Work	(4) Migrant for Education	(5) Married
Exposed Cohort (17-21) x Exposed District	-0.0079	0.0016	0.016	0.016	-0.025
	(0.029)	(0.016)	(0.021)	(0.013)	(0.020)
Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes
Individual Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes
District Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes
Region FE x Exposed Cohort	Yes	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.056	0.062	0.071	0.046	0.11
Observations	14065	14065	14065	14065	14065
Clusters	109	109	109	109	109

Notes: Clustered standard errors in parentheses. Standard errors clustered at the district level. The dependent variable is an indicator equal to one if the individual has a secondary level of education column (1); an indicator equal to one if the individual is still in school in column (2); an indicator equal to one if the individual has migrated for work reasons in column (3); an indicator equal to one if the individual has migrated for education reasons in column (4); and an indicator equal to one if the individual is married in column (5). Exposed Cohort (17-21) is equal to one if the individual is aged 17-21 at the time their nearest university was established. The comparison cohort is those aged 24-29 at establishment. Exposed District is equal to one if the district has a public university. Individual controls include gender and age. District controls include population and rurality.

Table A6: Robustness: Comparison Cohorts

	(1) Baseline	(2) Comparison Cohort 24-32 at Establishment	(3) Comparison Cohort 24-36 at Establishment	(4) Comparison Cohort 24-40 at Establishment
Exposed Cohort (17-21) x Exposed District	0.042***	0.035**	0.033**	0.035**
	(0.015)	(0.015)	(0.014)	(0.015)
Cohort Fixed Effects	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes
Individual Controls x Exposed Cohort	Yes	Yes	Yes	Yes
District Controls x Exposed Cohort	Yes	Yes	Yes	Yes
Region FE x Exposed Cohort	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.058	0.062	0.066	0.070
Observations	14065	16378	19647	21943
Clusters	109	109	109	109

Notes: Clustered standard errors in parentheses. Standard errors clustered at the district level. The dependent variable is an indicator equal to one if the individual has a post-secondary education. Exposed Cohort (17-21) is equal to one if the individual is aged 17-21 at the time their nearest university was established. The comparison cohort is those aged 24-29 at establishment in column (1); those aged 24-32 at establishment in column (2); those aged 24-36 in column (3); and those aged 24-40 in column (4). Exposed District is equal to one if the district has a public university. Individual controls include gender and age. District controls include population and rurality.

Table A7: Earnings Results with Different Imputation Assumptions

	Earnings in USD							Inverse Hyperbolic Sine and Logs						
	(1) Earnings USD Unemp=0	(2) Earnings USD Unemp=p1	(3) Earnings USD Unemp=p5	(4) Earnings USD Unemp=p10	(5) Earnings USD Unemp=p25	(6) Earnings USD Unemp=0 Self Emp=p50	(7) Earnings USD Unemp=p25 Self Emp=p50	(8) Earnings IHS Unemp=0	(9) Earnings IHS Unemp=p1	(10) Earnings IHS Unemp=p5	(11) Earnings IHS Unemp=p10	(12) Earnings IHS Unemp=p25	(13) Earnings IHS Unemp=0 Self Emp=p50	(14) Earnings IHS Unemp=p25 Self Emp=p50
Exposed Cohort (17-21) x Exposed District	2.56*	2.43*	2.21*	2.10*	1.87	2.24*	1.71	0.16*	0.089*	0.053	0.043	0.028	0.12	0.022
	(1.42)	(1.37)	(1.28)	(1.24)	(1.15)	(1.25)	(1.05)	(0.088)	(0.051)	(0.034)	(0.030)	(0.026)	(0.080)	(0.023)
Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.077	0.074	0.069	0.066	0.060	0.12	0.081	0.10	0.11	0.11	0.11	0.093	0.18	0.14
Observations	14065	14065	14065	14065	14065	14065	14065	14065	14065	14065	14065	14065	14065	14065
Clusters	109	109	109	109	109	109	109	109	109	109	109	109	109	109

Notes: Clustered standard errors in parentheses. Standard errors clustered at the district level. The dependent variable is the monthly earnings in USD in columns (1) to (7); the inverse hyperbolic sine of monthly earnings in columns (8) to (14). Columns (1) and (8) impute monthly earnings for the non-employed as zero; columns (2) and (9) impute monthly earnings for the non-employed as equal to the 1st percentile of positive earnings; columns (3) and (10) impute monthly earnings for the non-employed as equal to the 5th percentile of positive earnings; columns (4) and (11) impute monthly earnings for the non-employed as equal to the 10th percentile of positive earnings; and columns (5) and (12) impute monthly earnings for the non-employed as equal to the 25th percentile of positive earnings; columns (6) and (13) impute monthly earnings for the non-employed as zero and monthly earnings for the self-employed as the median of positive earnings; columns (7) and (14) impute monthly earnings for the non-employed as equal to the 25th percentile of positive earnings and monthly earnings for the self-employed as the median of positive earnings. Exposed Cohort (17-21) is equal to one if the individual is aged 17-21 at the time their nearest university was established. The comparison cohort is those aged 24-29 at establishment. Exposed District is equal to one if the district has a public university. Individual controls include gender and age. District controls include population and rurality.

Table A8: Control Experiment: Employment and Earnings: 24-29 v 34-40

	(1) Employed	(2) Paid Employed	(3) Monthly Earnings USD	(4) IHS Monthly Earnings	(5) Hourly Wage USD	(6) Log Hourly Wage	(7) Industry Public Admin
Older Cohort (24-29) x Exposed District	-0.0061	-0.028	-1.61	-0.12	0.11	0.0088	-0.0030
	(0.023)	(0.024)	(1.68)	(0.10)	(0.11)	(0.040)	(0.024)
Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.13	0.13	0.090	0.14	0.11	0.24	0.046
Observations	10287	10287	10287	10287	3861	3861	10287
Clusters	109	109	109	109	106	106	109
Dependent Variable Mean	0.79	0.33	18.2	1.41	1.17	0.62	0.19
Dependent Variable Std Dev	0.41	0.47	41.7	2.11	1.55	0.51	0.39

Notes: Clustered standard errors in parentheses. Standard errors clustered at the district level. The dependent variable is an indicator equal to one if the individual is currently employed in column (1); an indicator equal to one if the individual is in paid employment in column (2); the monthly earnings in USD in column (3); the inverse hyperbolic sine of monthly earnings in column (4); the hourly wage in USD in column (5); the log hourly wage in column (6); and an indicator equal to one if the individual is employed in public administration in column (7). Older Cohort (24-29) is equal to one if the individual is aged 24-29 at the time their nearest university was established. The comparison cohort is those aged 34-40 at establishment. Exposed District is equal to one if the district has a public university. Individual controls include gender and age. District controls include population and rurality.

Table A9: Employment Status

	(1) Employed: Paid Employee	(2) Employed: Unpaid	(3) Employed: Self-Employed	(4) Not Employed
Exposed Cohort (17-21) x Exposed District	0.037*	0.011	-0.0084	-0.037*
	(0.019)	(0.013)	(0.018)	(0.020)
Cohort Fixed Effects	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes
Individual Controls x Exposed Cohort	Yes	Yes	Yes	Yes
District Controls x Exposed Cohort	Yes	Yes	Yes	Yes
Region FE x Exposed Cohort	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.096	0.14	0.073	0.13
Observations	14065	14065	14065	14065
Clusters	109	109	109	109
Dependent Variable Mean	0.34	0.079	0.32	0.26
Dependent Variable Std Dev	0.47	0.27	0.47	0.44

Notes: Clustered standard errors in parentheses. Standard errors clustered at the district level. The dependent variable is an indicator equal to one if the individual is in paid employment in column (1); an indicator equal to one if the individual is in unpaid employment in column (2); an indicator equal to one if the individual is in self-employment in column (3); and an indicator equal to one if the individual is not in employment in column (4). Exposed Cohort (17-21) is equal to one if the individual is aged 17-21 at the time their nearest university was established. The comparison cohort is those aged 24-29 at establishment. Exposed District is equal to one if the district has a public university. Individual controls include gender and age. District controls include population and rurality.

Table A10: Selection into Industry of Employment

	All Individuals				Employed Sample			
	(1) Agriculture	(2) Manufacturing, Construction, Mining	(3) Public Admin	(4) Services	(5) Agriculture	(6) Manufacturing, Construction, Mining	(7) Public Admin	(8) Services
Exposed Cohort (17-21) x Exposed District	0.011	0.011	0.036**	-0.020	-0.00094	0.0074	0.035	-0.037*
	(0.015)	(0.019)	(0.018)	(0.017)	(0.017)	(0.022)	(0.022)	(0.021)
Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.26	0.063	0.026	0.036	0.29	0.062	0.041	0.056
Observations	14065	14065	14065	14065	10452	10452	10452	10452
Clusters	109	109	109	109	109	109	109	109
Dependent Variable Mean	0.13	0.15	0.18	0.26	0.18	0.21	0.24	0.35
Dependent Variable Std Dev	0.34	0.36	0.38	0.44	0.38	0.40	0.43	0.48

Notes: Clustered standard errors in parentheses. Standard errors clustered at the district level. The dependent variable is an indicator equal to one if the individual is employed in agriculture in columns (1) and (5); an indicator equal to one if the individual is employed in manufacturing, construction or mining in columns (2) and (6); an indicator equal to one if the individual is employed in public administration in columns (3) and (7); an indicator equal to one if the individual is employed in services in columns (4) and (8). Exposed Cohort (17-21) is equal to one if the individual is aged 17-21 at the time their nearest university was established. The comparison cohort is those aged 24-29 at establishment. Exposed District is equal to one if the district has a public university. Individual controls include gender and age. District controls include population and rurality. The sample includes employed and non-employed individuals in columns (1) to (4) and only employed individuals in columns (5) to (8).

Table A11: Robustness: Migration

	Non Migrants (Work and Education)						Non Migrants (All)					
	(1) Post Secondary	(2) Employed	(3) Paid Employed	(4) Monthly Earnings USD	(5) IHS Monthly Earnings	(6) Log Hourly Wage	(7) Post Secondary	(8) Employed	(9) Paid Employed	(10) Monthly Earnings USD	(11) IHS Monthly Earnings	(12) Log Hourly Wage
Exposed Cohort (17-21) x Exposed District	0.043**	0.031	0.038**	2.97*	0.18**	0.0012	0.057***	0.029	0.059***	2.84**	0.25***	-0.013
	(0.017)	(0.021)	(0.018)	(1.51)	(0.082)	(0.035)	(0.017)	(0.022)	(0.020)	(1.40)	(0.081)	(0.035)
Cohort Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Individual Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District Controls x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Region FE x Exposed Cohort	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R-Squared	0.060	0.14	0.099	0.076	0.11	0.24	0.053	0.12	0.085	0.062	0.091	0.25
Observations	11857	11857	11857	11857	11857	4413	9883	9883	9883	9883	9883	3628
Clusters	109	109	109	109	109	109	109	109	109	109	109	107
Dependent Variable Mean	0.17	0.73	0.31	15.4	1.30	0.55	0.15	0.75	0.29	13.3	1.20	0.49
Dependent Variable Std Dev	0.38	0.44	0.46	37.4	2.03	0.48	0.36	0.44	0.46	33.9	1.95	0.45

Notes: Clustered standard errors in parentheses. Standard errors clustered at the district level. The dependent variable is an indicator equal to one if the individual has a post-secondary level of education columns (1) and (7); an indicator equal to one if the individual is employed in columns (2) and (8); an indicator equal to one if the individual is in paid employment in columns (3) and (9); monthly earnings in USD in columns (4) and (10); the inverse hyperbolic sine of monthly earnings in columns (5) and (11); and the log hourly wage in columns (6) and (12). Exposed Cohort (17-21) is equal to one if the individual is aged 17-21 at the time their nearest university was established. The comparison cohort is those aged 24-29 at establishment. Exposed District is equal to one if the district has a public university. Individual controls include gender and age. District controls include population and rurality. The sample is restricted to individuals that have not migrated for work or education reasons in columns (1) to (6). The sample is restricted to individuals that have not migrated for any reason in columns (7) to (12).