Education, the Labour Market and Welfare in East Africa

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

Livini Tesha Donath

ID 14266059

Supervised by Oliver Morrissey and Trudy Owens

Nottingham, April 2021

Abstract

This thesis consists of three empirical essays on issues relating to the analysis of the link between education and the labour market in East Africa. The first essay investigates whether returns to schooling differ according to the choice of the measure of earnings and the different periods in which workers are paid (daily, weekly, and monthly). Using comparable data from the Living Standards Measurement Study (LSMS) for Malawi, Tanzania, and Uganda, and accounting for endogeneity using Gaussian Copula and Heckman selection models, we show that pooling/aggregating earnings to different common measures produce different estimates of returns to education. Estimating separately for each pay period, the analysis also reveals that returns to education differ significantly. The analysis suggests that estimating returns separately for different periods is more reliable than pooling.

The second essay employs Recentered Influence Function (RIF) Regressions to examine the distributional effect of education on earnings in East Africa. It investigates how the distributional effect of education on earnings differs according to the different periods in which workers are paid, using the same dataset as the first essay. Results show that, in all three countries, there is a significant difference in the distribution of earnings between pay periods, and thus the role of education in explaining earnings inequality differs across the pay periods. Generally, the effect is more substantial for workers reporting monthly earnings than their daily and weekly counterparts. Like for the first essay, the second essay also reiterates the need to estimate for each period separately for more reliable results.

The third essay examines whether the welfare difference between youth and adult headed households between 2001 and 2018 is attributable to differences in educational attainment following Universal Primary Education (UPE). The RIF decomposition method applied to the household budget survey (HBS) data for 2001 and 2018 reveals that the increase in youth educational attainment between 2001 and 2018 significantly explain the difference in welfare between the 2001 and 2018 youth cohorts. The findings also show that differences in educational attainment are significant factors explaining differences in welfare between youth and adults in each year. We find no evidence that the difference in welfare between the youth and adults and between youth in 2001 and their 2018 counterparts can be attributed to the difference in returns to education.

Aknowledgements

First and foremost, I am thankful to Almighty God, his endless mercy and favour protected and kept me safe throughout this journey.

I would like to express my deepest gratitude to my supervisors Professor Oliver Morrissey and Doctor Trudy Owens for their invaluable and timely advice and guidance which made this work possible.

I am grateful to the Commonwealth Scholarship Commission (CSC) in the UK and the University of Nottingham for funding my PhD studies. Without this invaluable financial support this study would not have been possible. Sincere gratitude to Professor Oliver Morrissey for his letters of recommendation that helped me to secure the scholarship.

I am forever grateful to my family for their emotional, moral and material support throughout my educational journey.

Finally, I would like to extend my appreciation to all the staff at the School of Economics for creating a wonderful learning and research atmosphere.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The importance of understanding and correctly measuring the rate of returns to education cannot be overemphasised. Since the seminal works of Becker (1964) and Mincer (1974) numerous studies have estimated returns to education, including for Africa (relevant literature is reviewed in each chapter). Various approaches have been adopted to address endogeneity of education with unobserved ability and sample selection biases—see Card (1999, 2001). Other issues include heterogeneity of returns across the earnings distribution and groups of workers. However, little to no attention has been focused on whether the worker's pay period matters in estimating returns to education, which is important in many African labour markets, as well as the distributional effects of education on earnings, and this thesis aims to explore these issues for East Africa.

Workers usually get paid over different pay periods depending on the type and duration of employment/work. Many surveys collecting information on the labour market then design their questionnaires with that in mind, and the World Bank's Living Standards and Measurement Study (LSMS) is no different. In East Africa, there are three main[1] pay periods over which workers report their earnings: daily, weekly, and monthly. Studies on returns to education usually pool the pay periods together by converting the earnings to a common period such as hourly, daily, or monthly. However, as the pay periods may indicate different labour markets, it is not clear how pooling the pay periods impacts the estimates of returns to education, the effect of education on the distribution of earnings, and gender wage gaps. Unlike previous studies, this thesis, therefore, examines whether

[1] Others are hourly, fortnightly, quarterly, semi-annually and annually, but relatively few workers.

alternative ways of converting the reported wages to a common unit/measure lead to different estimates; and analyse each pay period separately to examine whether the estimates differ by pay period.

The Malawi Integrated Household Surveys (IHS), the Tanzania National Panel Surveys (TNPS), and the Uganda National Panel Surveys (UNPS), which are part of the LSMS for Malawi, Tanzania and Uganda, respectively, provide comparable labour market data across the countries and this thesis takes advantage of that. Chapter 2 describes the surveys and data, limiting the focus to the surveys conducted between 2008 and 2017. This chapter aims to show how the variables of interest are extracted or computed, what adjustments are warranted and how the final sample is constructed from the raw data. Using the available information from the surveys, the reported earnings are converted to three different units: daily (hereafter DailyC), monthly (hereafter MonthlyC) and to annualised but expressed per month (hereafter MonthlyA)

The first empirical essay (Chapter 3) focuses on investigating whether, when pay periods are pooled, the estimates of returns to education varies depending on the choice of the common unit (i.e., DailyC, MonthlyC and MonthlyA); and whether the relationship between education and earnings varies across workers reporting earnings daily, weekly or monthly. These questions have not been addressed in the literature. If the different common units lead to different estimates then it implies that estimates from studies using different common units are not directly comparable. Furthermore, the common method of pooling the pay periods together may lead to biased estimates if the pay periods indicate segmented labour markets (Fichtenbaum, 2006).

In this Chapter, different specifications and estimation strategies are employed to test robustness. In terms of specification of the earnings function, both the completed years of schooling (with a quadratic term to capture non-linearity in education) and levels of education (with dummies for primary, secondary, and higher education and incomplete primary as the base group) are used. In terms of estimation, beginning with OLS as a baseline, the analysis employs Gaussian Copula (GC) estimation method by Park and Gupta (2012) to address the well-known endogeneity concerns from omitted variable bias. GC is an instrument free method that can recover the estimate of an endogenous variable by directly modelling the correlation between the endogenous regressor and the error term in the regression using copula functions. Considering its flexibility and reliance on less restrictive assumptions, GC is preferred to other instrument free methods, given the lack of comparable variables across the countries to be used as instruments for education or ability. The analysis employs the Heckman model (Heckman, 1979) to account for concerns from sample selection bias. The omitted variable and selection biases are simultaneously addressed by modelling the endogenous education within the sample selection model (Heckman with Gaussian Copula (HGC)).

A challenge with the GC model is that when the endogenous variable is discrete (like in our case[2]), and thus its cumulative distribution function (CDF) is a step function, it tends to produce shaky[3] estimates. As a robustness check, however, bootstrap aggregating (bagging) is employed in both GC and HGC regressions, whereby the model is estimated many times and coefficients averaged.

Malawi has a unique labour market structure, with a disproportionately large proportion of the labour force in rural areas primarily participating in agriculture and in off-own-farm short time (mainly piece rate) jobs famously locally known as *ganyu*. This group is excluded from the primary analysis and analysed separately as an extension to the chapter. Similar estimation methods to the primary analysis are employed in *ganyu*.

Chapter 4 extends the analysis in Chapter 3 to investigate the distributional effects of education on earnings. The novelty of this chapter is that even though some previous studies in Sub-Saharan Africa (SSA) have addressed this, no study

[2]Education (in years) is theoretically continuous but usually recorded as discrete (positive integers) in surveys.

[3]But the estimates normally differ only very slightly.

has investigated while taking into consideration the pay period of the worker, i.e., if and how the effect of education on the distribution of earnings and gender wage gaps varies according to workers' pay period. Specifically, the chapter examines, for each pay period, the effects a change in the distribution of education in the population has on the earnings distribution. It then further investigates how such a change in the distribution of education would affect the earnings gap between the high and the low earning workers. Finally, the chapter investigates the extent to which gender differences in educational attainment and in returns to education contribute to the gender earnings and inequality gaps in the three countries. As an extension, the analysis is also employed for *ganyu* labour in Malawi.

The analyses in Chapter 4 are based on Recentered Influence Function (RIF) regression and decomposition methods (Firpo et al., 2009, 2018), among the most recent econometric tools for analysing the effects of a change in the distribution of a variable on any statistic of interest (Rios-Avila, 2020b). The superiority of the methods over alternative strategies is, among others, its ability to compute many measures of earnings inequality to suit the purpose, such as interquantile share ratios of earnings, interquantile range and variance of earnings.

Chapter 5 focuses on a slightly different matter but within the returns to education literature. It investigates how the increased participation in education in Tanzania following the re-introduction of Universal Primary Education (UPE) in 2001 contributed to increased earnings and welfare. It is expected that the youth aged 15 – 35 years in 2018 would have more education than the older (adult) cohorts since it is the cohort that mainly benefited from the programme. However, what is not known is whether the increase in education significantly increased earnings and welfare; and whether the increase in education changed the relationship between education and earnings (that is, whether returns to education increased/decreased over this period).

The chapter, therefore, seeks to explore how much of the welfare differences between youth-headed households in 2001 (not affected by UPE) and 2018 (benefited from UPE) can be attributable to differences in educational attainment and returns to education between the two cohorts, using household budget survey data for 2001 and 2018. The chapter employs the same estimation strategy employed in Chapter 4—reweighted RIF decomposition—to decompose the welfare differences between the cohorts (welfare proxied by household consumption expenditure per adult equivalent relative to the national poverty line).

Chapter 2

The Living Standards Measurement Study in East Africa

2.1 Introduction

2.1.1 The Living Standards Measurement Study

The datasets for the analysis of the Eastern Africa labour market come from the Living Standards Measurement Study (LSMS) Household Survey, a World Bank program aimed at facilitating the design and implementation of multi-topic household surveys in developing countries (World Bank, 2020a). Through this program, since 1980, the World Bank in collaboration with country statistical offices, have been conducting numerous nationally[1] representative household surveys in developing countries. The surveys collect information on a wide range of topics including but not limited to employment and labour market participation; education (access and attainment); healthcare use and access; poverty; and housing and utilities (World Bank, 2020a).

Malawi, Tanzania, and Uganda remain the main beneficiaries of the LSMS program in the Eastern Africa region. Four cross-section surveys, the Integrated Household Survey (IHS), and a three-wave panel study, the Integrated Household Panel Survey (IHPS)[2], are available for Malawi between 1997 and 2017[3]. Tanzania has the largest number of surveys in the region conducted between 1991 and

[1]LSMS have also conducted a few surveys which are not nationally representative e.g. Human Resource Development Survey (HRDS) the Measuring Living Standards in Cities (MLSC) surveyand the Kagera Health and Development Survey (KHDS)

[2]IHPS is a subset of the large IHS with a longitudinal/panel dimension

[3]Most recently LSMS have been conducting the Malawi High-Frequency Phone Survey COVID-19 (HFPS COVID-19) on monthly basis.

2016[4]. Two of the surveys are cross-section: the Human Resource Development Survey (HRDS) 1993 and the Measuring Living Standards in Cities (MLSC) survey. Two are longitudinal studies: the Kagera Health and Development Survey (KHDS) (three waves), and the Tanzania National Panel Survey (TNPS) (detailed discussion to follow). Lastly, there are seven[5] available surveys for Uganda, namely the National Panel Surveys (UNPS). IHS, TNPS, and UNPS have been conducted relatively more frequently than the other available household surveys, making themselves among the primary sources of labour market statistics in the region.

We limit the scope of our study to Malawi, Tanzania, and Uganda and the period from 2008 through 2017. IHS, TNPS, and UNPS used similar questionnaires, thereby allowing comparability across the countries. In that regard, we find them more suitable for our study and therefore use them as our main source of data for our analyses. Figure 2.1 is the map of Africa showing the geographical location of the counties whose data are used in our analyses.

2.1.2 Malawi Integrated Household Survey (IHS)

IHS is a series of extensive household surveys conducted every five[6] years in Malawi. The first IHS, was conducted in 1997/98 and is commonly referred to as IHS1, the second in 2004/05 (IHS2), the third in 2010/11 (IHS3), and IHS4 in 2016/17 (National Statistical Office, 2020). A portion of IHS3 sample was selected to be re-interviewed in 2013, and 2016 (integrated into IHS4) thereby forming a longitudinal survey panel study (IHPS). However, because the number of enumeration areas for the IHPS 2016 was reduced to 102 out of 204 of the baseline enumeration areas in IHPS 2010, the sample lost representativeness at the regional and urban/rural levels although maintained representativeness at the national level. Also, the reduction of the enumeration areas by half led to fewer observations with positive wage earnings. For that reason, we prefer to use IHS instead of IHPS, and therefore unlike for the other countries, the Malawi data do

[4] Recent addition is Tanzania National Panel Survey 2019-2020 (not included in the analysis)
[5] Latest are UNPS 2018/19 and UNPS 2019/2020 (not included in the analysis).
[6] Except the most recent – the Fifth IHS (IHS5) conducted after 3 years

Figure 2.1: Map of Africa Showing Geographical Location of the Study Countries

not have a panel dimension.

As stated earlier, our analyses focus on the 2008 -2017 period, and therefore we use IHS3 conducted from March 2010 to March 2011 and IHS4 that was conducted from April 2016 to April 2017. Both surveys are based on a two-step stratified sampling from the 2008 Malawi Population and Housing Census (PHC). The samples are representative at the national, regional, and urban/rural levels. The first two rows of Table 2.1 show the distribution of the study population by survey years for Malawi.

2.1.3 Tanzania National Panel Survey (TNPS)

Five waves of TNPS are currently available and four are used in this study. The first wave was conducted from October 2008 to September 2009, the second from October 2010 to September 2011, the third from October 2012 to November 2013, and the fourth from October 2014 to January 2016. The panel was refreshed in the fourth round. Consequently, the fourth wave included only 784 households (out of the 4,036 households interviewed) that were present in the previous rounds (URT, 2017). The surveys for all four waves are based on a two-step stratified sampling. The first three waves sampling are based on the 2002 Tanzania PHC while the fourth is based on the 2012 Tanzania PHC. The samples are representative at the national, regional, and urban/rural levels. The distribution of the study population by survey rounds for Tanzania is shown in rows 2 - 6 of Table 2.1.

2.1.4 Uganda National Panel Survey (UNPS)

We use five of the seven available waves of the UNPS. The first wave, conducted as a follow-up survey of 3,123 households that had been visited by the Uganda National Household Survey (UNHS) in 2005-06, was conducted from September 2009 to August 2010; the second from October 2010 to September 2011, the third from November 2011 to November 2012, the fourth from September 2013 to August 2014 and the fifth from March 2015 to March 2016. After following up the surveyed households in three consecutive waves, UNPS panel replaces some household by new ones. Since, like for Tanzania, the survey is intended to trace households and not individuals, the panel of individuals in the labour force is small (this is explained later in the chapter). The last five rows of Table 2.1 show the distribution of the study population by survey rounds for Uganda.

Table 2.1: Distribution of Study Population by Country

Country	Survey Year	Households	Individuals
Malawi	2010/11	12,271	56,409
Malawi	2016/17	12,447	53,885
Tanzania	2008/09	3,265	16,709
Tanzania	2010/11	3,846	20,559
Tanzania	2012/13	5,010	25,412
Tanzania	2014/15	4,036	21,027
Uganda	2009/10	2,607	17,511
Uganda	2010/11	2,564	18,810
Uganda	2011/12	2,356	16,139
Uganda	2013/14	3,118	17,495
Uganda	2015/16	3,300	16,748

Source: Respective IHS, TNPS and UNPS reports.

2.2 Data Management

2.2.1 Data Cleaning

The LSMS data are accessible and freely downloadable from the websites of the World Bank[7] and the countries' statistical offices. After obtaining the required survey data, we proceeded to data cleaning and construction and preparation of the variables for our analysis. One of the common issues with survey data is missing data for some variables caused by, among others, non-response during the survey. For Tanzania and Uganda where we had some panel dimension, as a first step, we utilised information from the other survey waves or other survey questions to deduce the possible missing information. For instance, we replaced the missing data on education by the education level reported in the other waves; or the previous year's grade for those who were in school in the year preceding the survey year. In the second step, we imputed the missing data by replacing them by median values (computed by gender, location, age group and pay period). The variables that required some imputations due to missing values included hours (and days for Uganda) worked in the last seven days, usual working weeks in a

[7]These versions of the data were downloaded on 10^{th} May 2019 from http://surveys.worldbank.org/lsms

month and months in a year. The potential problem with this approach is that the imputation may introduce bias on the mean and variance of the affected variables.

Furthermore, some questions were only asked in some rounds of the survey. For instance, while the last three waves of TNPS collected information on the number of weeks per month and the number of months each individual usually worked in the job during the last 12 months, the first wave did not. For the first wave, we imputed the missing data by replacing them by the median values (again computed by gender, location, age group and pay period). As the median values were obtained within the sample, they may be less prone to outliers. This helped in keeping the first wave in our analysis, however, like stated earlier, it may introduce some bias to the final result. We understand that there are alternative imputation methods that could have been used in this case, but we do not pursuit them here. See Appendix 2B for the number of cases imputed for each country.

Another issue was that we identified inaccurate information for some variables such as changes of the time-invariant individual characteristics (e.g. gender and year of birth) across surveys, inconsistencies in the reported/recorded education, and outlier wages. For the time-invariant variables, we corrected these inconsistencies as follows: firstly, for those observed three or more times (Tanzania and Uganda) we took the value occurring most times to be the correct one. Secondly, for those observed multiple times but no value was reported more than the other, we took the one reported during their first survey as the correct value (this is an arbitrary assumption though as we could have as well treated them as different individuals). We also used information from the other survey questions and from other waves (for Tanzania and Uganda) to correct the inconsistencies in education for those with contradicting years of schooling. We then deleted all individuals whom we could not infer their education either from the other waves or from the other survey questions.

For TNPS, some of the reported wages were either too high or too low such that they could not feasibly reflect the pay periods. Cross-examination across waves also revealed that these wages were different from those reported in the other waves. We concluded that there might be errors in recording the pay period or the wages (such as errors due to addition/omission of zeros). We, therefore, adjusted (116 cases out of 14,444 individuals with valid values of wages) through the utilisation of information on wages and pay periods from other waves of the survey. See Appendix 2A for all correction and adjustments made to key variables.

2.2.2 Construction of Earnings Variables

The surveys collected information about individuals wage earnings and the frequency of payment. The payment periods for Malawi were daily, weekly, and monthly; for Tanzania hourly, daily, weekly, fortnightly, monthly, quarterly, semi-annually, and annually; and for Uganda hourly, daily, weekly, and monthly. Note that the payment periods may not necessarily imply the same period/duration of employment, i.e., being paid daily or weekly does not always mean that employment last only for a day or a week. Each of the conversion methods is discussed below.

2.2.2.1 Aggregating to Daily earnings (DailyC)

Wages were converted to daily wages as follows:

(a) Hourly to daily (Tanzania and Uganda)

The hourly wage multiplied by the number of hours assuming nine (9) working hours a day.

(b) Weekly to daily

Weekly wage divided by the total number of days worked per week (unless otherwise stated in the survey, days were inferred from the total weekly hours).

(c) Fortnightly to daily (Tanzania)

Fortnightly wage divided by two and then divided by the total number of days worked per week.

(d) Monthly to daily

Monthly wage divided by 22 (assuming those earning monthly wage worked 22 days in any month).

(e) Quarterly

Quarterly wage divided by 66 (since assumption 22 working days in any month).

(f) Semi-annual

Semi-annual wage divided by 132 working days.

(g) Annual

Annual wage divided by 264 working days.

2.2.2.2 Aggregating to Monthly Earnings (MonthlyC)

Monthly wages were constructed from the reported wages as follows:

(a) Hourly to monthly (Tanzania and Uganda)

Total number of hours worked over the last seven (7) days multiplied by hourly wage and number of weeks worked in the job in a typical month[8].

(b) Daily to monthly

For Uganda, the number of days the individual worked over the last seven days was available from the data. For Malawi and Tanzania, since the number of days was not available, we used the total number of hours in a week to infer days. Assuming nine (9) working hours per days, we obtained the proxy for days by dividing the total hours by nine. We then constructed the monthly wage as a product of the daily wage, days worked, and the number of weeks worked per month.

(c) Weekly to monthly

Weekly wage multiplied by the number of weeks worked per month

[8] As stated earlier, TNPS (except for the first wave) asked how many weeks per month did the individual usually work in the job during the last 12 months. For the first wave we replaced it by the median values of the sample for each pay period.

(d) Fortnightly to monthly (Tanzania)

Fortnightly wage multiplied by two (2).

(e) Quarterly to monthly (Tanzania)

Quarterly wage divided by three (3).

(f) Semi-annually to monthly (Tanzania)

Semi-annual wage divided by six (6).

(g) Annually to monthly

Annual wage divided by 12.

2.2.2.3 Aggregating to Annualised Earnings (MonthlyA)

Lastly, the reported wages were annualised as follows:

(a) Hourly to annual (Tanzania and Uganda)

The product of hourly wage and hours per week, weeks per month and months worked over the last 12 months.

(b) Daily to annual

The product of daily wage and days per week, weeks per month, and months worked over the last 12 months.

(c) Weekly to annual

The product of weekly wage, weeks worked per month and months worked over the last 12 months.

(d) Fortnightly to annual (Tanzania)

Fortnight wage divided by two then multiplied by weeks worked per month and months worked over the last 12 months.

(e) Monthly to annual

Monthly wage multiplied by the number of months worked over the last 12 months.

(f) Quarterly to annual

Quarterly wage divided by three, then multiplied by the number of months worked over the last 12 months.

(g) Semi-annually to annual

Semi-annual wage divided by six, multiplied by the number of months worked over the last 12 months.

Note: we express the annualised wages monthly by dividing by 12, which gives the average monthly earnings from over the last 12 months. This may not be identical to our constructed measure MonthlyC (except for those paid monthly who worked 12 months last year). After the construction of our wage/earnings variables, we observed a small number of cases with very low and very high MonthlyA, likely errors in recording the wage or variables used to construct the aggregated wages. We then resorted to trimming the bottom and top one percent of MonthlyA to get rid of the outliers.

2.2.3 Construction of Explanatory Variables

2.2.3.1 Years and Levels of Education

In all three countries, each grade requires a year to complete. The IHS, TNPS and UNPS used a closed-ended question to capture the highest grade completed by each member of the household. Therefore, we utilised the information on the grades completed to calculate the respondent's years of schooling assuming that each additional grade corresponds to an additional year of schooling. Note, however, that there was no information on the number of years the individual took to complete their highest grade. Hence, the calculation of years of schooling assumed no repetitions or skipping of grades.

Primary education is compulsory in all three countries, and it runs for eight years in Malawi and seven years in Tanzania[9] and Uganda. In Malawi, secondary

[9] Before 1969 primary education in Tanzania ran for eight years. An education reform act in late 1960s eliminated the 8th grade thereby reducing the primary school years from eight to seven. We used individuals' years of birth to infer whether the individual obtained seven or eight

education lasts for four years and until 2015 consisted of two sets of two years. The first two years lead to the Junior Certificate of Education (JCE) (which was abolished in 2015) and the second two years to the Malawi Certificate of Secondary Education (MCSE). Admission to (non-university) technical college education such as diplomas in vocational training including nursing, primary teacher training and agriculture requires a minimum of JCE and run for two, three or four years. Admission to university requires the MCSE, with a minimum of three years required to earn a university degree.

In Tanzania and Uganda, secondary education consists of six years in two levels: ordinary level (O-level) and advanced level (A-level) which run for four and two years, respectively. Diploma education is two years for those enrolled after A-level and three or four years for those enrolled after O-level (in our analysis, we use three years for those enrolled after O-level). University education is three to five years, depending on the programme of study. Note that individuals can enrol in technical/vocational education after completing primary or secondary education. This form of education can take less than a year to more than two years. For simplicity, in our calculation, we assume this level does not constitute an additional year of schooling[10].

Since the surveys reported the highest grade of schooling completed (for each level of education) assigning individuals into dummy variables for the highest completed levels was straight forward. Accordingly, we constructed the following dummy variables:

(a) noeduc: educational dummy, 1 if incomplete primary school education and 0 otherwise.

(b) primary: educational dummy, 1 if completed primary school education and 0 otherwise.

years of primary schooling. We assumed all individuals who completed eight years of primary education started school at age seven and were born before 1956. Any error or misreporting of the birth year would then affect years of schooling, especially those with post-secondary education.

[10]In Tanzania and Uganda a total of 724 and 404 wage employees had vocational education of unspecified duration, accounting for 6.5% and 8.7% of the samples respectively.

(c) secondary: educational dummy, 1 if completed ordinary/advanced secondary school education and 0 otherwise.

(d) higher: educational dummy, 1 if completed diploma/university education and 0 otherwise.

2.2.3.2 Other Explanatory Variables and Exclusion Restrictions

(a) age: After correcting the inconsistencies in the year of birth[11], we calculated age as the difference between the survey year, and the year of birth (taking into consideration the month of birth).

(b) female: The variable female is a gender dummy = 1 for females and 0 otherwise.

(c) rural: The variable rural is a location dummy = 1 for rural areas and 0 for urban. It was readily available in the datasets.

(d) panel: Tanzania and Uganda only, panel is a dummy variable = 1 for the individuals observed multiple times and 0 otherwise.

(e) year: Malawi only, year is a dummy variable = 1 if the year of the survey is 2016 and 0 if 2010.

(f) married: The survey question for marital status consisted of seven responses: monogamous married, polygamous married, living together, separated, divorced, never married, and widow(er). We made a dummy variable = 1 if married or living together and 0 otherwise.

(g) kids5: It is the proportion of children aged five and under in the household calculated as the ratio of the number of children aged five years and younger to the total number of household members.

[11]For TNPS the year of birth was not available in the fourth wave, instead, we calculated it from the age of the respondents.

(h) kids14: It is the proportion of children aged 6 to 14 years of age in the household calculated as the ratio of children aged between 6 and 14 years inclusively to the total number of household members.

2.2.4 Characteristics of the Wage Sample

2.2.4.1 Final Samples

After cleaning the data, we remained with samples of labour force of 45,494, for Malawi, 38,857 for Tanzania and 29,188 for Uganda. Of these samples 5,816[12], 11,215, and 4,631 individuals had valid values of earnings for Malawi, Tanzania and Uganda, respectively on which we focus our analysis. As mentioned earlier, the samples for Tanzania and Uganda have some individuals who were surveyed more than once. Therefore, the 11,215 observations for Tanzania consist of a total of 8,210 individuals of which 6,016 were surveyed only once, 1,458 twice, 661 three times and 75 four times. Likewise, the 4,631 observations for Uganda consist of a total of 2,491 individuals of which 1,929 were surveyed once, 704 twice, 207 three times, 98 four times and 58 five times. Thus, although the TNPS and UNPS were intended as panels, due to attrition and refreshing very few individuals (only 27% for TNPS and 23% for UNPS) are observed at least twice in the sample. As a result, as explained earlier, in our analyses we include a variable for individuals with repeat observations but otherwise pool and treat observations as independent.

We treat *ganyu* separately because it is a specific (segmented) labour market, and the measure of earnings (as explained in section 2.3.2) is different to other labour. It is not correct that this is equivalent to the informal sector; non-*ganyu* workers are in formal and informal segments of the market and, as for Uganda and Tanzania, the distinction of pay periods provided in the data does not clearly correspond to occupation or formal/informal distinctions. The survey questions were framed as follows: (non- *ganyu*): In the last 12 months, did you work as an

[12] Not including off-own-farm labour (*ganyu*), this category of labour is covered in a separate section. *Ganyu* is treated separately because it is a specific (segmented) labour market, and the measure of earnings (as will be explained later in this chapter) is different.

employee for a wage, salary, commission, or any payment in kind: including doing paid apprenticeship, domestic work or paid farm work, excluding *ganyu*, even if only for one hour? (*Ganyu*): In the last 12 months, did you engage in casual, parttime or *ganyu* labour, even if only for one hour? The questions were also asked for the last 7 days.

2.2.4.2 Shares of Workers that do not Work Full Periods

In each of the payment periods, there are categories of workers that did not work full periods. For instance, workers reporting monthly earnings that did not work all 12 months last year; workers reporting weekly earnings that did not work for four weeks (a full month) last month or that did not work for 48 weeks last year; and workers reporting daily earnings that did not work the whole week, month or full-year over. Table 2.2 shows the share of workers that work full period. Table 2.2 shows that the share of workers that work full periods is higher for workers reporting monthly earnings compared to those reporting daily and weekly earnings, suggesting the presence of a larger number of casual workers in daily and weekly compared to monthly earners. Also, the second panel of Table 2.2 shows that in Tanzania the shares of daily and weekly workers that work over the whole year is substantially lower than other countries, indicating that these categories constitute large shares of seasonal workers that only work some weeks or months in a year.

2.2.4.3 Description of Employment Type and Labour Market Characteristics

Table 2.3 shows, for each county and pay period, the employer of the primary job. As expected, the private sector (not necessarily equivalent to formal) dominates employment in all countries. Compared to Tanzania and Uganda, the proportion of workers employed by the government and paid daily and weekly is relatively high in Malawi due to the exclusion of *ganyu* workers. For Malawi and Uganda, the private sector could be broken down into private enterprise and private household (the latter are likely to be informal, but the former may not all be formal). Table 2.3

Table 2.2: Shares of Workers that Work Full Periods (Percentage)

	Daily	Weekly	Monthly	Pooled
Malawi				
Five or more days a week	43.62	61.15	66.66	65.35
Four weeks a month	91.44	94.51	97.66	97.15
Twelve months a year	47.35	45.55	57.04	55.6
Observations	182	505	5,129	5,816
Tanzania				
Five or more days a week	33.55	41.73	87.01	55.52
Four weeks a month	39.72	38.41	95.25	60.08
Twelve months a year	15.09	13.29	51.18	27.89
Observations	3,738	1,929	4,830	11,215
Uganda				
Five or more days a week	67.03	68.33	87.3	79.5
Four weeks a month	78.68	75.22	93.35	86.8
Twelve months a year	54.38	57.41	58.85	57.48
Observations	1,262	589	2,765	4,631

Source: Author's computations from IHS, TNPS and UNPS. Note: off-own-farm (*ganyu*) labour excluded for Malawi. Distribution adjusted by survey weights.

shows a decline in the share of workers employed by households as the pay period increases from daily to monthly, with the most substantial decline in Uganda.

Table 2.4 shows the proportion of workers with a job contract in each pay period. As *ganyu* are excluded for Malawi, consistent with the large proportion of daily and weekly workers employed in the government sector (Table 2.3), Table 2.4 shows that the proportion of workers with a job contract is significantly greater than in Tanzania and Uganda. Only about 2% of the workers paid daily and weekly in Tanzania have a job contract compared to 46% of workers paid monthly. For Uganda, 3% and 13% of daily and weekly paid workers respectively have a job contract compared to 52% of workers paid monthly. In both countries, on this definition about half of monthly paid workers are informal. The small proportion of workers paid daily and weekly who have a job contract clearly implies a high degree of informality in daily and weekly samples in the two countries.

Employment duration (permanent, fixed term or temporary) is described in Table 2.5. Again, given exclusion of *ganyu*, Malawi has a significantly higher

Table 2.3: The Distribution of Workers by Employer (% shares)

Country & Employer	Daily	Weekly	Monthly	Pooled
Malawi				
Government	29.19	41.41	33.03	33.95
Private enterprise	26.04	48.31	35.32	36.27
Private household	37.63	35.85	33.95	34.26
Other	7.14	3.32	6.05	5.83
Observations	182	505	5,129	5,816
'Full' sample	182	505	5,129	5,816
Tanzania				
Government	1.46	1.8	22.3	9.43
Private	96.25	95.56	71.09	86.6
Other	2.29	2.64	6.61	3.97
Observations	2,453	1,217	2,823	6,859
'Full' sample	3,738	1,929	4,830	11,215
Uganda				
Government	0.56	5.26	30.63	19.11
Private enterprise	45.99	44.51	42.64	43.72
Private household	52.72	45.93	21.62	33.34
Other	0.74	4.31	5.11	3.83
Observations	1,259	589	2,760	4,623
'Full' sample	1,262	598	2,765	4,631

Source: Author's computations from IHS, TNPS and UNPS. Note: off-own-farm (*ganyu*) labour excluded for Malawi. Distribution adjusted by survey weights.

Table 2.4: proportion of workers with a job contract (%)

	Daily	Weekly	Monthly	Pooled
Malawi				
Contract	34.64	41.41	33.03	33.95
Observations	134	284	2,237	2,655
'Full' sample	182	505	5,129	5,816
Tanzania				
Contract	2.23	2.24	45.7	18.55
Observations	2,453	1,217	2,823	6,859
'Full' sample	3,738	1,929	4,830	11,215
Uganda				
Contract	2.61	12.71	51.62	33.16
Observations	1,259	589	2,760	4,623
'Full' sample	1,262	598	2,765	4,631

Source: Author's computations from IHS, TNPS and UNPS. Note: off-own-farm (*ganyu*) labour excluded for Malawi. Distribution adjusted by survey weights.

proportion of permanent employees paid daily and weekly compared to Tanzania and Uganda. The small share of permanent employees paid daily and weekly for Tanzania and Uganda indicates that these pay periods comprise large shares of informal paid employment. Surprisingly, many workers paid monthly in Malawi consider themselves to have permanent employment, although they do not have a job contract. Nonetheless, this clearly indicates that, even for Malawi, we cannot treat the monthly sample as equivalent to the formal sector.

Table 2.5: Employment Duration (% shares)

Country & Employer	Daily	Weekly	Monthly	Pooled
Malawi				
Permanent	54.54	38.51	65.47	62.23
Fixed>year	4.67	12.97	7.12	7.6
Temporary/freelance	40.8	48.52	27.41	30.17
Observations	129	276	2,151	2,556
'Full' sample	182	505	5,129	5,816
Tanzania				
Permanent	0.39	0.35	20.92	8.37
Fixed>year	6.44	4.29	22.1	12.12
Temporary/freelance	93.17	95.35	56.98	79.05
Observations	1,114	630	1,550	3,475
'Full' sample	3,738	1,929	4,830	11,215
Uganda				
Permanent	0.36	3.85	25.31	15.68
Fixed>year	8.84	14.03	7.81	8.93
Temporary/freelance	90.8	82.13	66.88	75.39
Observations	1,258	589	2,757	4,619
'Full' sample	1,262	598	2,765	4,631

Source: Author's computations from IHS, TNPS and UNPS. Note: off-own-farm (*ganyu*) labour excluded for Malawi. Distribution adjusted by survey weights.

Table 2.6 shows the share of workers enrolled in pension schemes by country and pay period. Again, consistent with the previous results, Malawi has a substantially higher proportion of workers enrolled in a pension scheme than Tanzania and Uganda (about 32% compared to 13-15%). However, the enrolment rate significantly differs across the pay periods. In Malawi, workers paid daily have the highest rate, while in Tanzania and Uganda the workers paid monthly have the highest enrolment rate. Compared to Malawi, workers paid daily and weekly in Tanzania and Uganda have a substantially lower enrolment rate with daily workers exhibiting the lowest rate. As enrolment in a pension scheme is usually associated with formal employment, it appears that there are significant shares of formal employment paid daily and weekly in Malawi but not in Tanzania and Uganda. With at most one-third of the workers paid monthly in all three countries enrolled in a pension, Table 2.6 reiterates that being paid monthly does not necessarily imply having formal employment.

Table 2.6: Proportion (%) of Workers Enrolled in a Pension Scheme

	Daily	Weekly	Monthly	Pooled
Malawi				
Enrolled	36.07	19.5	33.82	32.49
Observations	134	284	2,237	2,655
'Full' sample	182	505	5,129	5,816
Tanzania				
Enrolled	1.06	1.19	31.69	13.06
Observations	1,114	630	1,550	3,475
'Full' sample	3,738	1,929	4,830	11,215
Uganda				
Enrolled	0.73	4.92	24.41	15.4
Observations	1,254	587	2,750	4,606
'Full' sample	1,262	598	2,765	4,631

Source: Author's computations from IHS, TNPS and UNPS. Note: off-own-farm (*ganyu*) labour excluded for Malawi. Distribution adjusted by survey weights.

Classification of workers by occupation is shown in Table 2.7. Agriculture, livestock keeping, forestry, fishing and hunting remain the dominant occupations for workers paid daily and weekly in Tanzania and Uganda, accounting for more than half of the samples.

Table 2.7: General Occupation of Workers (%)

	Daily	Weekly	Monthly	Pooled
Malawi				
Professional & Technical	30.5	8.93	20.61	19.81
Service Workers	26.94	29.81	30.97	30.72
Agriculture, forestry, fishing & hunting	12.96	23.69	9.82	11.27
Other	29.6	37.57	38.6	38.2
Observations	181	504	5,127	5,812
'Full' sample	182	505	5,129	5,816
Tanzania				
Professional & Technical	12.68	10.76	27.58	18.41
Service Workers	5.35	4.11	22.43	11.35
Agriculture, forestry, fishing & hunting	55.73	58.62	10.27	39.08
Other	26.24	26.51	39.72	31.16
Observations	3,738	1,929	4,830	11,215
'Full' sample	3,738	1,929	4,830	11,215
Uganda				
Professional & Technical	1.87	8.06	43.34	27.85
Service Workers	17.32	28.62	8.19	8.43
Agriculture, forestry, fishing & hunting	65.16	51.02	9.63	14.28
Other	15.65	12.3	38.84	49.44
Observations	1,143	541	2,668	4,367
'Full' sample	1,262	598	2,765	4,631

Source: Author's computations from IHS, TNPS and UNPS. Note: off-own-farm (*ganyu*) labour excluded for Malawi. Distribution adjusted by survey weights.

In addition, monthly paid workers constitute the largest proportion of professional and technical occupations in the two countries, suggesting that attractive employment of this category of workers are more likely to last longer and pay monthly. In Malawi, on the other hand, professional and technical occupations account for the largest share of workers paid daily (consistent with relatively high earnings, see Table 2.10), while service workers account for the largest share of workers paid weekly and monthly. As *ganyu* workers are excluded, it is not surprising that Malawi has a smaller proportion of workers in agriculture and related compared to Tanzania and Uganda.

Table 2.8 shows the industry in which the worker is employed. Community, social and personal services account for the most employment in Malawi, whereas in Tanzania and Uganda, agriculture and related account for the most employment. In all three countries, the share of monthly paid workers in agriculture and related is substantially smaller than for daily and weekly. The small share of monthly paid workers in this industry is more pronounced in Tanzania than in the other two countries.

Table 2.8: Industry of the Employment (Employer's Business) (%)

	Daily	Weekly	Monthly	Pooled
Malawi				
Agriculture, forestry, fishing & hunting	12.75	19.07	5.70	7.20
Manufacturing	15.80	20.62	17.19	17.48
Education	14.06	2.10	6.67	6.47
Community, Social & Personal Services	32.04	30.64	40.50	39.29
Other	25.35	27.57	29.94	29.56
Observations	168	473	4,913	5,554
'Full' sample	182	505	5,129	5,816
Tanzania				
Agriculture, forestry, fishing & hunting	62.29	65.98	12.36	44.42
Manufacturing	4.29	4.37	5.70	4.73
Education	0.60	1.09	11.98	4.94
Community, Social & Personal Services	1.83	1.28	6.33	3.45
Other	30.99	27.28	63.63	42.46
Observations	3,214	1,701	4,040	9,508
'Full' sample	3,738	1,929	4,830	11,215
Uganda				
Agriculture, forestry, fishing & hunting	28.18	34.01	14.20	20.68
Manufacturing	5.82	7.18	6.07	6.13
Education	0.12	3.27	29.04	17.80
Community, Social & Personal Services	12.14	11.53	8.08	9.77
Other	53.74	44.01	42.61	45.62
Observations	1,228	567	2,695	4,505
'Full' sample	1,262	598	2,765	4,631

Source: Author's computations from IHS, TNPS and UNPS. Note: off-own-farm (*ganyu*) labour excluded for Malawi. Distribution adjusted by survey weights.

Table 2.9 show the proportion of workers whose occupations can be considered as white-collar jobs. While the survey questionnaire included the relevant question for Tanzania, unfortunately, the responses were not published with the data; hence, the comparison is only for Malawi and Uganda. About 26% of daily paid workers in Malawi have a white-collar job, which is in line with earlier results that showed that the daily sample constitutes a relatively larger proportion of professional and technical workers and workers with a job contract. In Uganda, on the other hand, a minimal share (less than 1%) of workers paid daily have a white-collar job compared to 36% of the workers paid monthly. Although the monthly samples constitute a relatively large share of workers in a white-collar job, blue-collar jobs still dominate.

Table 2.9: Proportion (%) of workers in White-collar jobs

	Daily	Weekly	Monthly	Pooled
Malawi				
White collar	25.57	5.63	19.41	18.28
Observations	182	505	5,129	5,816
'Full' sample	182	505	5,129	5,816
Uganda				
White collar	0.38	5.81	35.95	22.29
Observations	1,258	584	2,744	4,601
'Full' sample	1,262	598	2,765	4,631

Source: Author's computations from IHS and UNPS. Note: off-own-farm (*ganyu*) labour excluded for Malawi. Distribution adjusted by survey weights.

Table 2.10 shows the distribution of earnings for workers with various levels of education in each pay period. Because workers with similar levels of education are paid differently depending on their pay period, this suggests that the pay periods indicate different segments of the labour markets.

2.2.4.4 Distribution of Earnings and other Variables by Payment Period

Table 2.11 shows the distribution of earnings by payment period using our three measures of earnings. We present the earnings in both US dollars \$US and local currency units (LCU). The columns for monthly and annualised are directly comparable since both present earnings per month (calculated in different ways). For the earnings in the daily column to be comparable to the other columns, they need to be multiplied by a factor of 22 (since the assumption is 22 working days

Table 2.10: The distribution of earnings for workers with various levels of education (%)

	Daily	Weekly	Monthly	Pooled
Malawi				
No education	163.28	119.73	89.61	96.93
Primary	126.51	144.31	141.07	140.98
Secondary	239.81	344.11	247.88	251.53
Higher	407.98	806.97	692.81	687.67
Observations	182	505	5,129	5,816
Tanzania				
No education	24.21	16.95	37.49	25.61
Primary	45.24	38.36	78.79	54.56
Secondary	84.95	87.79	197.24	171.89
Higher	0.00	0.00	370.71	371.46
Observations	3,738	1,929	4,830	11,215
Uganda				
No education	58.56	50.56	41.29	50.1
Primary	88.98	85.53	84.08	85.95
Secondary	134.07	95.61	108.18	112.15
Higher	118.72	158.48	202.06	196.86
Observations	1,262	589	2,765	4,631

Source: Author's computations from IHS, TNPS and UNPS. Note: off-own-farm (*ganyu*) labour excluded for Malawi. Distribution adjusted by survey weights.

in any month).

Table 2.11 shows that DailyC and MonthlyC give larger average monthly earnings compared to MonthlyA. Importantly, DailyC will overestimate monthly earnings because it does not consider the number of days the worker worked in a week, and the number of weeks worked in a month. Table 2.11 shows that regardless of the measures of earning used, the earnings differ by payment period. Nonetheless, regarding which period has the highest/lowest earnings, it depends on the measure of earnings used.

Table 2.12 shows the distribution of the explanatory variables to be used in the wage equation. For Tanzania and Uganda, the variable 'panel' shows the proportion of workers with repeated (panel) observations. The small proportion (45%) for Tanzania is explained by the fact that the last round of the survey was a refresh wave. There are a few issues that could potentially affect our results. Workers paid monthly have more education than their daily and weekly counterparts in all three countries. In Tanzania, there are no workers with higher education reporting earnings daily or weekly. In Malawi, only 12% of the workers reporting daily earnings and 4% of those reporting weekly earnings have higher education, while in Uganda 3% and 8% of the workers reporting daily and weekly earnings respectively have higher education. Overall, workers in Malawi have more years of schooling on average compared to their Tanzania and Uganda counterparts.

Table 2.11: Distribution of Earnings by Different Earnings Measures ($US and LCU)

Country & Period	Obs.	$ daily Mean	$ daily SD	$ monthly Mean	$ monthly SD	$ annualised Mean	$ annualised SD	LCU daily Mean	LCU daily SD	LCU monthly Mean	LCU monthly SD	LCU annualised Mean	LCU annualised SD
Malawi													
Daily	182	18.77	21.60	258.16	318.11	206.42	251.48	2,650.15	3,049.41	36,443.80	44,907.13	29,140.24	35,501.37
Weekly	503	22.10	50.94	223.54	309.10	174.26	289.04	3,120.01	7,190.90	31,557.13	43,635.62	24,599.68	40,804.40
Monthly	5,129	12.59	18.01	264.82	377.90	226.84	339.20	1,777.72	2,542.08	37,384.43	53,347.49	32,022.80	47,884.38
Pooled	5,816	13.72	23.62	260.58	370.04	221.04	332.38	1,937.44	3,334.45	36,786.58	52,237.98	31,204.69	46,921.56
Tanzania													
Daily	3,738	5.18	5.50	59.08	100.24	32.54	85.08	6,587.81	7,664.81	78,580.79	133,314.68	43,283.94	113,156.05
Weekly	1,929	4.95	5.76	60.98	100.97	32.30	84.51	6,287.49	7,481.10	81,109.67	134,287.74	42,962.47	112,393.41
Monthly	4,830	6.82	8.02	150.02	176.51	123.90	161.28	9,069.55	10,670.72	199,530.00	234,755.77	164,793.41	214,501.08
Pooled	11,215	6.00	7.55	98.14	147.47	69.10	126.88	7,973.95	10,045.83	130,519.70	196,133.61	91,905.45	168,748.00
Uganda													
Daily	1,262	4.64	5.00	91.47	109.51	78.92	103.44	9,429.51	10,159.08	185,739.05	222,364.60	160,242.78	210,038.33
Weekly	589	5.53	8.06	91.60	122.62	76.34	109.07	11,221.61	16,358.95	185,987.04	248,982.52	155,016.12	221,470.95
Monthly	2,765	5.88	6.71	129.33	147.56	114.19	139.02	11,936.67	13,618.94	262,606.79	299,616.72	231,865.71	282,275.95
Pooled	4,631	5.53	6.74	114.11	136.35	99.66	127.84	11,227.57	13,688.05	231,705.66	276,867.30	202,361.64	259,583.03

Source: Author's computations from IHS, TNPS and UNPS. Note: Earnings in $ are accounted for inflation using exchange rates in 2009 (1$= 141.17 Malawi Kwacha; 1$ = 1330 Tsh; and 1$= 2030.49 UGX). Off-own-farm (*ganyu*) labour excluded for Malawi. Distribution adjusted by survey weights.

Table 2.12: Summary Statistics for the Explanatory Variables

Country & Sample	Obs.	sch		age		weeks		primary	secondary	higher	female	rural	panel	year
		Mean	SD	Mean	SD	Mean	SD	%	%	%	%	%	%	
Malawi														
Daily	182	9.03	4.51	36.8	10.2	36.75	13.99	20	29	12	26	67	NA	74
Weekly	505	7.15	3.91	34.94	10.35	35.51	14.71	24	10	4	30	66	NA	53
Monthly	5,129	9.26	4.12	35.85	10.71	39.63	12.69	26	24	14	25	49	NA	50
Pooled	5,816	9.05	4.16	35.79	10.66	39.13	13.01	26	23	13	25	52	NA	51
Ganyu	16,528	4.77	3.52	33.42	11.89	15.64	12.66	14	2	0	51	92	NA	64
Tanzania														
Daily	3,738	5.26	3.21	33.14	11.92	15.32	15.39	57	4	0	40	78	25	NA
Week	1,929	5.28	3.24	33.64	12.08	13.62	15.16	56	5	0	36	82	19	NA
Monthly	4,830	8.16	3.69	33.34	11.89	34.95	16	51	27	7	38	53	41	NA
Pooled	11,215	6.35	3.67	33.37	11.93	22.1	18.38	55	13	3	38	70	45	NA
Uganda														
Daily	1,262	6.35	3.53	30.99	10.84	36.04	14.57	34	11	3	20	64	39	NA
Weekly	589	6.97	3.81	32.77	11.44	35.27	14.87	32	12	8	29	72	15	NA
Monthly	2,765	9.85	4.39	34.71	11.06	39.38	12.45	29	16	31	37	56	56	NA
Pooled	4,631	8.51	4.41	33.46	11.18	37.89	13.54	31	14	20	31	61	57	NA

Source: Author's computations from IHS, TNPS and UNPS. The last two columns show % observed multiple times for Tanzania and Uganda (panel) and % in 2016 for Malawi (year) respectively. Distribution adjusted by survey weights

2.3 Off-own-farm Labour (*Ganyu*) in Malawi

2.3.1 Introduction

Malawi is one of the poorest countries in the world, with 70% of its population below the international poverty line in 2016 (World Bank, 2020b). Compared to other neighbouring countries, a very high proportion (84% as of 2018) of its population of about 18 million[13] resides in rural areas, and about 73% of the labour force is employed in agriculture (National Statistical Office, 2019) (National Statistical Office, 2019). In addition, Malawi experiences only one rainy season in a year which affects agricultural productivity as most farmers own small scale farms which depend heavily on rainfall. Because of low productivity in agriculture, many households are unable to sustain their livelihood through own production alone necessitating extra income that is obtained through off-own-farm casual labour (Bryceson, 2006).

Short time off-own-farm labour in Malawi is commonly referred to as *ganyu*[14]. Although traditionally the term applied to rural farm activities, its definition extends to include both farm and off-farm tasks in which labour is usually hired for a short time, mainly daily and weekly. This kind of labour (hereafter *ganyu*) is mainly piece rate, including labour for farm tasks (such as, planting, weeding, ridging, and harvesting) and non-agricultural tasks like building houses, fetching water, and helping in construction. There are various kinds of *ganyu* such as non-wage *ganyu* (*Chipere*) whereby neighbours or relatives work for each other without pay; ganyu during food shortages where households supply their labour in exchange for food; non-agricultural *ganyu* where short time labour is supplied to non-agricultural activities; and cross-border *ganyu* where short time labour is supplied to neighbouring countries especially by households close to the border with Mozambique (Whiteside, 2000). In the IHS the in-kind *ganyu* payments were

[13] According to the 2018 Malawi Population and Housing Census (PHC), Malawi had a population of 17,563,749 in 2018.

[14] *Ganyu* is a Chichewa word meaning hire or part-time job.

converted to an equivalent cash amount, and therefore in this study, we treat all the types of *ganyu* labour payments as cash wages/earnings.

Due to its unique labour market structure, with a disproportionately large proportion of the labour force in rural areas primarily participating in agriculture, the IHS has a separate section in its household questionnaire to collect information on *ganyu* labour. Since the kind of activities carried out by *ganyu* workers varies by location and season of the year, the frequency individuals participate in *ganyu* may range from a day to a year. In that regard, the survey collected information on all individuals who participated in *ganyu* labour even if just for a single day over the last 12 months. It is worth noting, however, that *ganyu* labour supply is likely to be underreported because it is stigmatised: some individuals perceive it as an admission of poverty and thus shameful to divulge(Whiteside, 2000).

2.3.2 *Ganyu* Labour in Our Labour Force Sample

After the initial data cleaning, we remained with a total of 17,849 individuals who participated in *ganyu* labour over the last 12 months. Out of this total, 847 individuals did both *ganyu* and other wage employment activities (27 in daily, 120 in weekly, and 700 in monthly). To simplify the analysis, for those who did both, we incorporated them into their primary wage employment by adding their *ganyu* earnings to their wage earnings and then exclude them from the *ganyu* sample. After excluding those who did both *ganyu* and other wage employment activities, we opted to trim off the bottom and top 1% as a way of getting rid of the outliers. The final sample consists of 16,528 individuals who participated only in *ganyu* labour as their primary source of labour earnings. This sample suggests that *ganyu* workers accounted for 79.3% of all earners in 2016 compared to 70% in 2010 implying an increase of 13.3%[15] over the survey period.

Figure 2.2 shows the spatial distribution of the sample across the surveys. As expected, a disproportionately large proportion of *ganyu* workers reside in rural areas (91.7% rural vs 8.3% urban), and the distribution has remained relatively

[15]Figures are adjusted using survey weights throughout this section.

stable over the survey years (91.0% rural vs 9.0% urban in 2010; and 92.1% rural vs 7.9% urban in 2016). However, as mentioned earlier, there were relatively more *ganyu* workers in 2016 than 2010 (63.4% of the *ganyu* workers are from 2016 survey, and 36.6% are from the 2010 survey). Figure 2.3 shows further desegregation by regions of residence. The Northern region accounts for the smallest share of *ganyu* workers in all years. This is not surprising as the region is less populated than the other regions. The Southern region accounts for the largest proportion of *ganyu* workers in 2010 while the Central region accounts for the largest proportion of *ganyu* workers in 2016.

Figure 2.2: Distribution of *Ganyu* Workers by Location of Residence

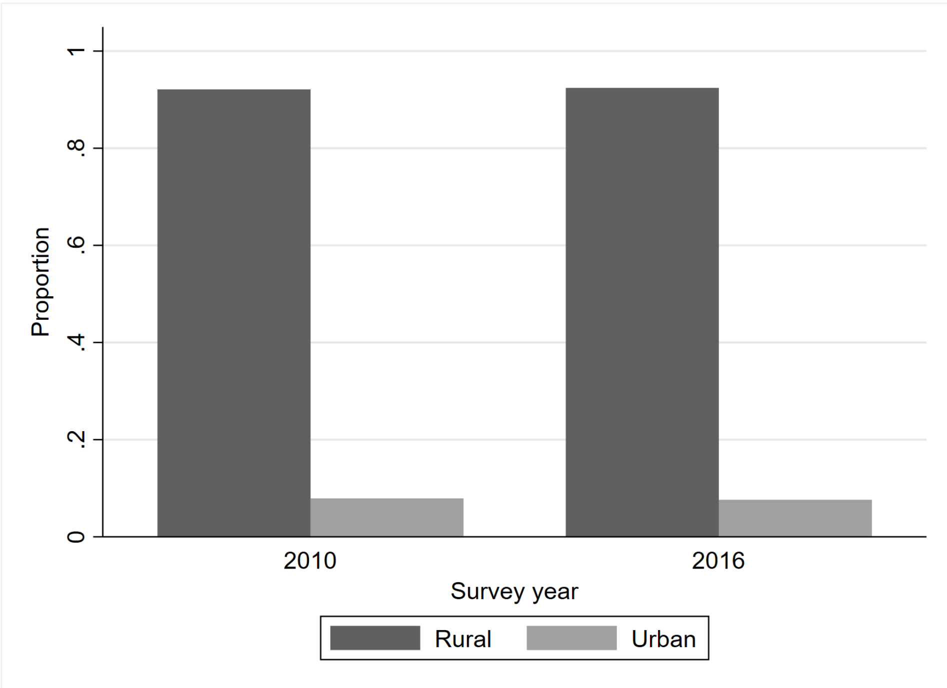


Figure 2.4 shows the distribution of *ganyu* workers by gender and survey year. Generally, there is no significant difference between male and female participation in *ganyu* labour (50.7% male compared to 49.3% female) although the gender balance reversed over the survey years (50.5% vs 49.5% in 2010 and 51.4% vs 48.6% in 2016).

In terms of educational attainment, about 21% of ganyu workers had never gone to school, and less than 1% had more than secondary education (more than 12 years of education). Figure 2.5 shows the distribution of *ganyu* workers by years

Figure 2.3: Distribution of *Ganyu* Workers by Region of Residence

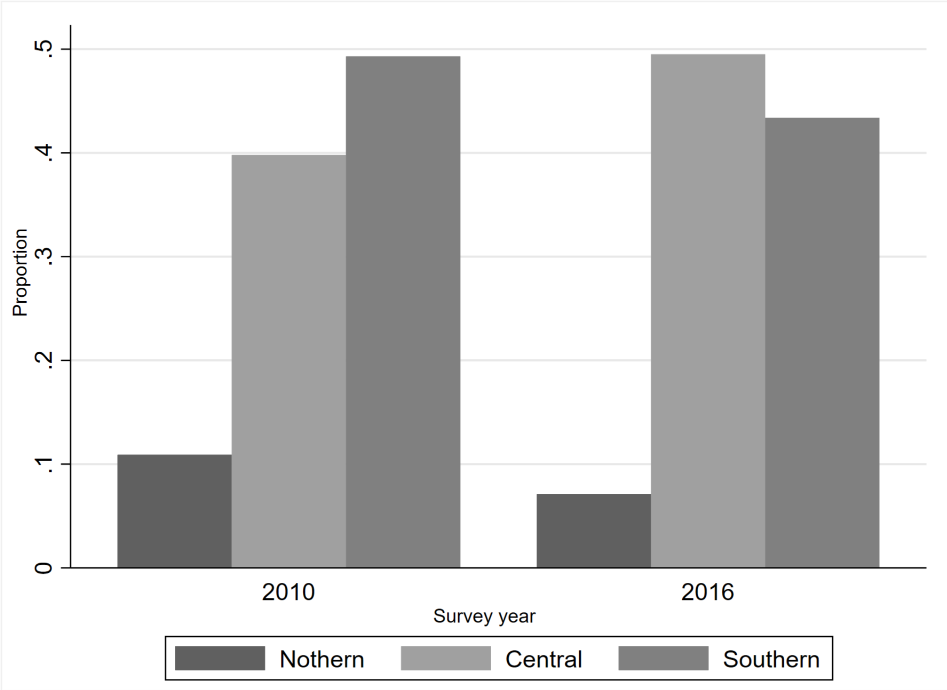
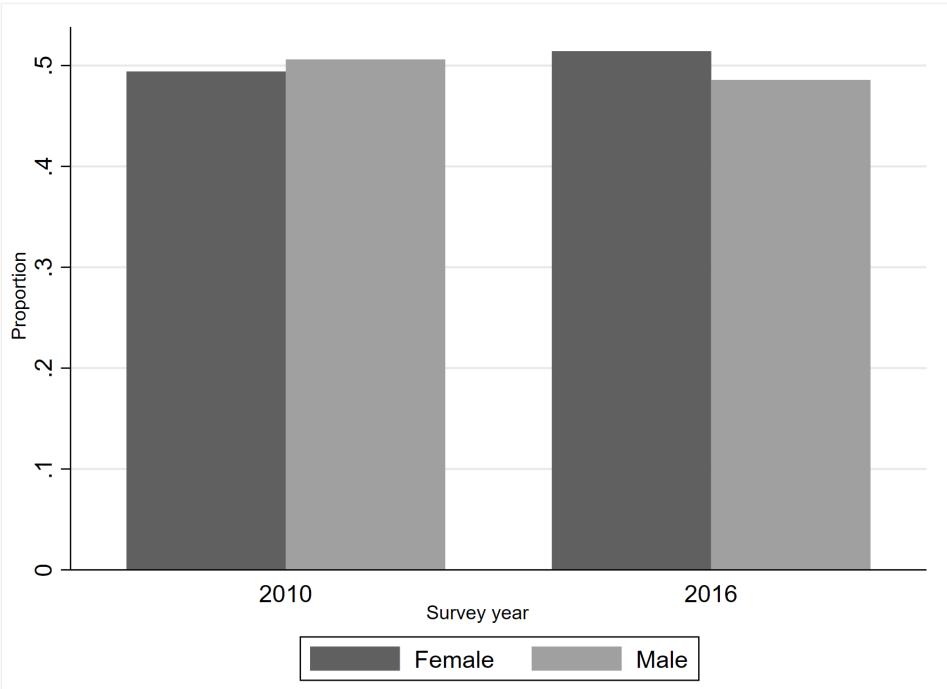


Figure 2.4: Distribution of *Ganyu* Workers by Gender and Survey Year



of education. Generally, more schooling reduces the likelihood of participating in *ganyu* labour. Nonetheless, Figure 2.6 shows that in terms of *ganyu* labour supply, as measured by the number of days participated in *ganyu* over the past 12 months, education does not matter. Figure 2.7 shows the relationship between education and *ganyu* earnings (using the MonthlyA). As expected, more education is associated with more earnings. It is important to note, however, that because there are very few individuals with more than secondary education in the sample, the relationship for those with more than 12 years of education may be imprecise, which may explain the unusually high level of earnings for those with 16 years of education.

Figure 2.5: Distribution of *Ganyu* Workers by Education Attainment

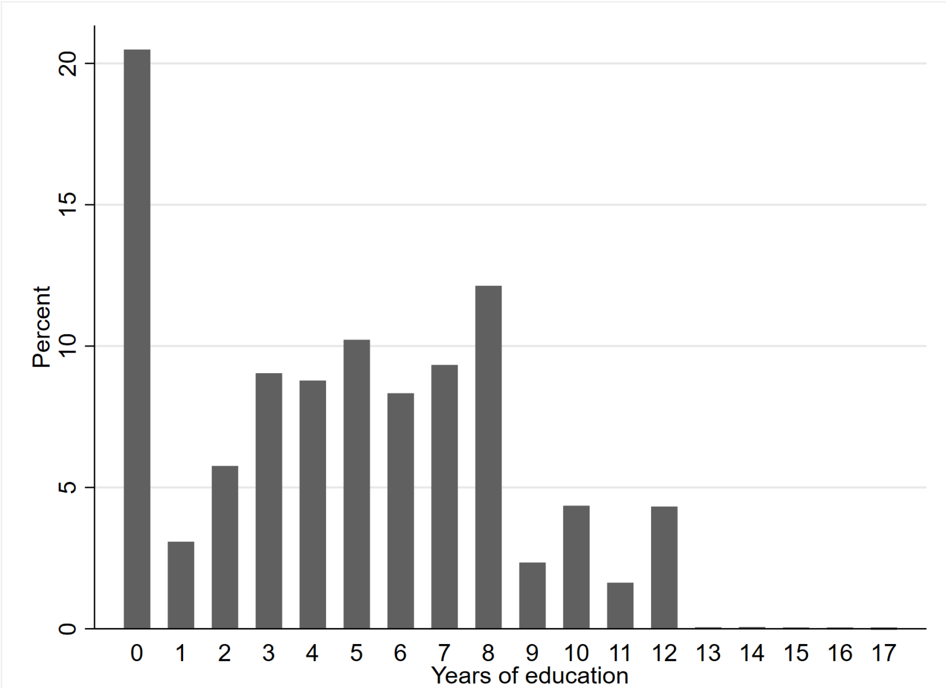


Figure 2.8 shows the association between annual *ganyu* labour supply and monthly per capita income. High *ganyu* labour supply is associated with low-income household members and vice versa. This is consistent with the assertion that *ganyu* is supplied more for survival than for earning wage income. As stated earlier, the kind of activities carried out by *ganyu* workers varies by location and season, thereby determining the individuals' total *ganyu* labour supply in a year. We thus also show the share of *ganyu* workers that supplied their labour for five days a week, four weeks a month and for all months over the last 12 months. Table 2.13 shows the share of *ganyu* workers that supplied their labour for a full period. Only about 10% of the *ganyu* workers reported supplying *ganyu* labour for 12 months in the year before the survey. This suggests that *ganyu* labour might be

Figure 2.6: *Ganyu* Labour Supply by Education Attainment

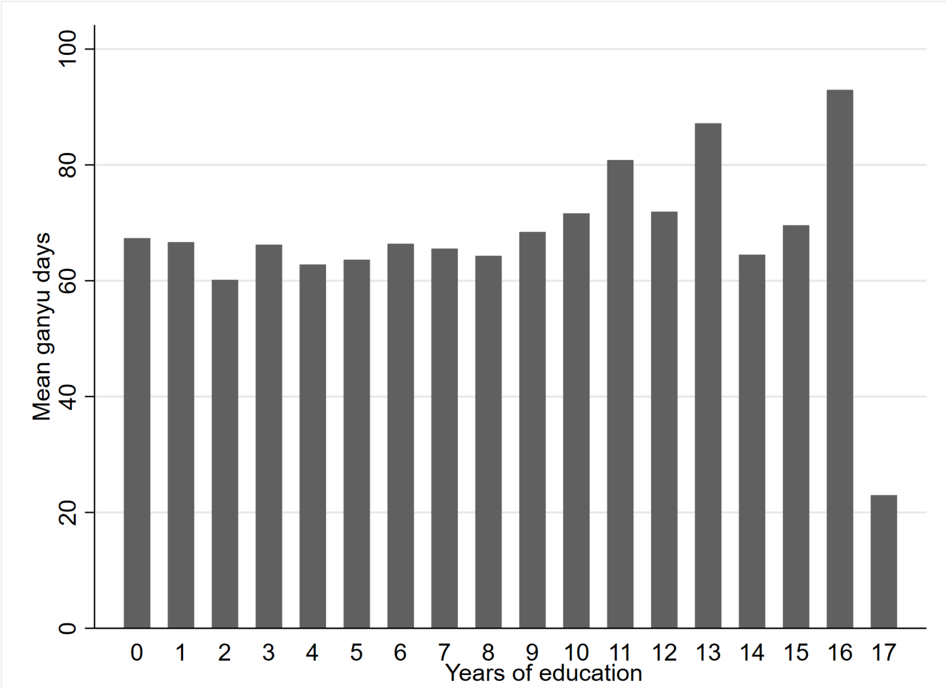
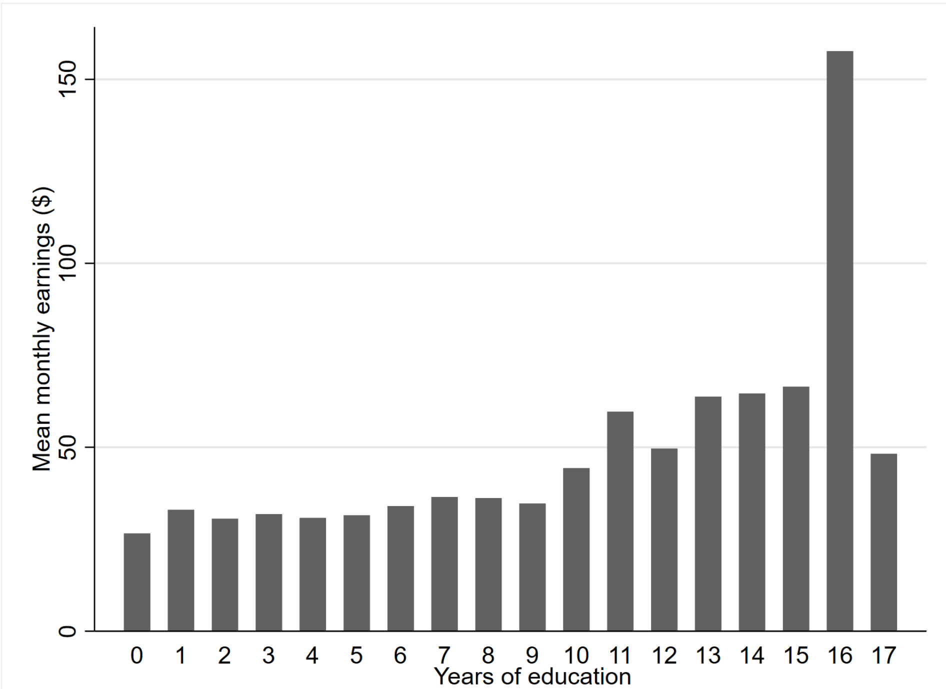


Figure 2.7: Education and *Ganyu* Earnings



mainly supplied to meet short time livelihood needs, especially the period between the harvesting seasons.

Table 2.14 shows the distribution of *ganyu* earnings by the different earnings measures discussed earlier. In the surveys, *ganyu* earnings were measured daily,

Figure 2.8: Association between Income and *Ganyu* Labour Supply

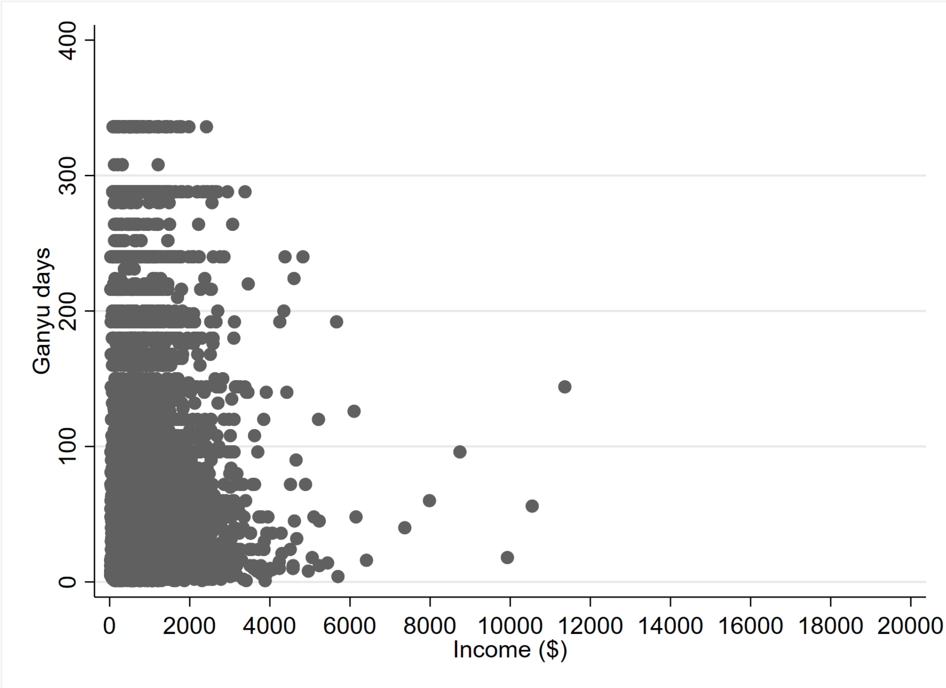


Table 2.13: Shares of *Ganyu* Workers that Worked Full Periods

	Percent
Five or more days a week	40.25
Four weeks a month	36.62
Twelve months a year	10.5
Observations	16,528

Source: Authors' Computation from IHS 2010, 2016

that is the average payment per day in cash or in-kind that the individual received for the days they supplied *ganyu* labour over the last 12 months. The daily wages were then converted to monthly by first converting them to weekly multiplying them by the number of days they supplied *ganyu* labour per week, and then to monthly by multiplying again by the number of weeks per month worked in *ganyu*. The annualised *ganyu* earnings were obtained by multiplying the monthly *ganyu* earnings by the number of months worked in *ganyu* over the last 12 months and then expressed in monthly basis by dividing by 12. Due to the nature of *ganyu* labour supply, it should be expected that converting to daily or monthly would very likely overestimate *ganyu* earnings because *ganyu* workers rarely work for the whole week or month. Figure 2.9 shows a kernel distribution of earnings corresponding to Table 2.14. We observe a more complicated distribution when the actual reported daily wages are used, and more smooth distribution when annualised earnings are used.

Table 2.14: Distribution of *Ganyu* Earnings ($US and LCU)

lPeriod & Unit	Obs.	Daily		Monthly		Annualised	
		Mean	SD	Mean	SD	Mean	SD
$	16,528	5.76	6.93	69.41	101.72	33.76	53.52
LCU	16,528	812.51	977.83	9797.99	14,360.21	4,766.48	7,555.73

Source: Author's computations from IHS. Earnings in $ are in constant 2009 exchange rate (1$= 141.17 Malawi Kwacha)

Figure 2.9: Distribution of *Ganyu* Earnings by Different Earnings Measures

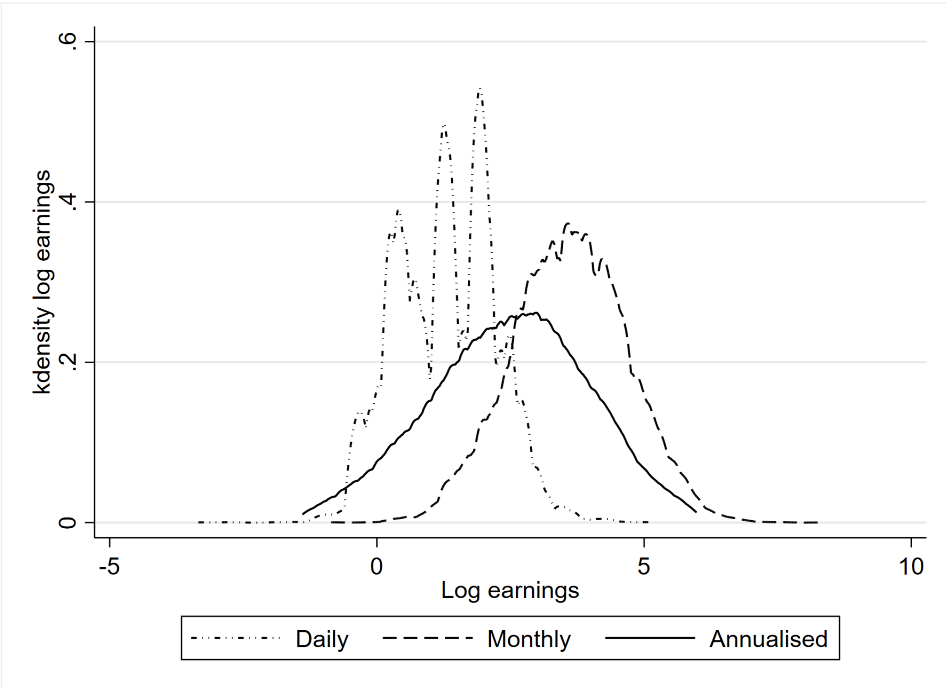


Table 2.15 shows the average (raw) earnings and education attainment by gender (Panel A) and location of residence (Panel B). The first three columns of Table 2.15 show the differences in earnings using the three measures of earnings while the last shows the corresponding difference for education attainment (measured by completed years of schooling). Regardless of the measure of earnings, males in *ganyu* labour are better paid than females, and urban workers are better paid than their rural counterparts. However, assuming a multiplication factor of

22 for the first column to be comparable to the next two columns, aggregating to daily gives a larger gap compared to aggregating to monthly or annualised. In terms of education endowment, males and urban *ganyu* workers have more years of schooling than their female and rural counterparts.

Table 2.15: Gender Differences in Earnings and Education Attainment

	Earnings Measure ($)			Education (years)
	Daily	Monthly	Annualised	
A. Gender				
Male	6.61	85.04	45.12	5.51
Female	4.89	53.8	22.48	4.06
Difference	1.72	31.24	22.64	1.45
Obs. Male	8,344	8,344	8,344	8,344
Obs. Female	8,285	8,285	8,285	8,285
Obs. Total	16,629	16,629	16,629	16,629
B. Location				
Urban	8.08	114.55	65.45	6.56
Rural	5.53	65.11	30.77	4.61
Difference	2.82	49.44	34.68	1.95
Obs. Urban	1,285	1,285	1,285	1,285
Obs. Rural	15,344	15,344	15,344	15,344
Obs. Total	16,629	16,629	16,629	16,629

Note: Differences significant at 1% level.

2.4 Conclusion

In this chapter we provided a detailed description of the dataset and variables to be used in the analyses for the next two emperical chapters. The description focused on the survey data cleaning, extraction of key variables from the surveys, construction of the variables for analysis, and how we arrived at the final samples.

It is worth noting that the available information on occupation and the employment industry reveals that pay periods do not correspond to a particular occupation/industry but a range of different occupations/industries. To the extent that the pay period may indicate different labour markets, labour market analyses by pay period may help to capture labour market segmentation that cannot otherwise be identified in the data. For example, while there is complete information on the pay period, data on formality is incomplete. We do not include these various indicators of job type as variables in the analysis because there are typically significant numbers of missing observations for at least one country. We showed that not all workers paid monthly are white-collar, and whilst it may be the case that almost all workers paid daily or weekly in Tanzania and Uganda can be considered informal, not all monthly are formal.

In addition, we provide a detailed description of the off-own-farm casual labour market (*ganyu*), a peculiar source of labour earnings in Malawi. The reason to treat *ganyu* separately is that it is a specific (segmented) labour market, and the measure of earnings is different to other labour. In addition, the screening questions in the questionnaire employed to extract labour market information were different and less detailed for the *ganyu* labour. We acknowledge that the exclusion of *ganyu* workers from Malawi's sample makes the data less comparable to the other two. Nonetheless, the samples represent the 'standard' labour force and focus of the analyses is not comparison across countries rather a comparison of the pay periods within each country.

Appendices

Appendix 2A: Correction and Adjustments Made to Key Variables

The following corrections and adjustments were made to key variables utilizing information from other waves, other survey questions or both. Note that the corrections and adjustments were made to the initial samples (containing all survey individuals) so the number of cases with corrections/adjustments may be significantly lower in the final samples.

Tanzania

Age: 4,855 cases with inconsistencies corrected using information from other waves.

Sex: 244 cases with inconsistencies corrected using information from other waves.

Education:

- 968 cases with inconsistencies in education when education was lower than previous survey(s) were corrected.
- 66 cases with missing schooling replaced by values using information from other education variables.
- 171 cases of individuals born after 1956 with eight instead of seven years of primary school level of education were corrected.

Earnings: 116 adjustments were made to wages (99 cases to reported pay periods and 17 errors caused by additional/omission of zeros 116)

Uganda

Age: 3683 cases with inconsistencies corrected using information from other waves.

Sex: 141 cases with inconsistencies in sex corrected using information from other waves.

Education: 3420 cases with missing values of education replaced by highest education grades from other waves.

Appendix 2B: Imputation of Time Normally Worked

Note that the median values used for imputation in all cases were computed by gender, location (rural vs urban), age group (15-35 (youth) vs 36-65 (adults) and pay period.

Malawi

- The period of time the reported salaries cover: (91 cases replaced by 1)
- Number of hours worked last 7 days (3 cases)
- Number of months worked in the last 12 months (17 cases)
- Number of weeks worked in the last 12 months (2 cases)
- The number of days worked in *ganyu* for *ganyu* workers (38 cases)
- Number of weeks in *ganyu* (26 case)

Tanzania

- Some reported 0 hours in 12 months, assumed error in recording and replaced by median values (63)
- Some reported 0 months in 12 months, assumed error in recording and replace by median values (198 cases)
- Some reported 0 weeks in 12 months, assumed error in recording and replace by median values (35)

- Missing values of number of months in 12 months replaced by median (30)

- Missing values of number of weeks in the last 22 months (34 cases)

- Wave 1 does not have information on days, weeks or months worked over the last 12 months, but the correlation between hours worked last week and typical number of hours in a week over past 12 months is high (0.86) so we use hours last week to proxy working hours in any week. For those reporting 0 hours over the last week, we replaced by median values (175 cases). The total number of hours in a week were then used to calculate the number of days (assuming 9 hours a day)

- Weeks worked in a month and over the last 12 months, and the number of months worked over the last 12 months for wave 1 were then imputed by the corresponding medians from the other waves.

Uganda

- The number of weeks worked last month and over the last 12 months was missing in the 2009 survey, so we imputed them using medians values from the other waves.

- The number of months worked over the last 12 months for those with missing values (67 cases)

- The number of days in the last week if missing days or did not work last week (289 cases).

Chapter 3

Does the Pay Period Matter in Estimating Returns to Schooling? Evidence from East Africa

3.1 Introduction

It is important to understand the rate of returns to education because it is one of the significant determinants of willingness to invest in education. Part of the justification for public investment in education is that it adds to human capital, skills, and productivity; this should generate a social benefit in addition to the private benefit of increased earnings by more productive workers. People will be willing to pay for education if it increases their earnings (Borjas, 2016) and parents' willingness to invest resources in their children's education depends on how they value future benefits that the children will get after acquiring education (Schultz, 2004). In developing countries expected returns to education may also be an important determinant of child labour—high returns increase school attendance and tend to reduce the likelihood of child labour (Kuepié and Nordman, 2016).

Most studies on returns to education over the past five decades have concentrated on developed countries (Psacharopoulos and Patrinos, 2018, 2004). Nonetheless, there is an emerging body of literature estimating returns to education in Sub-Saharan Africa (SSA) countries, including Malawi, Tanzania and Uganda, with a broad consensus that since around 2000 returns to secondary education have exceeded those of primary education. While coefficient estimates vary, returns are increasing with the level of education (and generally also with years of education). Given limitations in the data, there are weaknesses in the existing evidence. This essay addresses some of these.

Studies on returns to education are mainly based on nationally representative surveys conducted by government statistical agencies and in some cases on surveys by private researchers. These surveys usually collect data on earnings by different pay periods (typically daily, weekly and monthly) which often reflects the type of employment. Most studies then measure earnings by converting these to a common period, normally hourly or monthly (see for example Nikolov and Jimi (2018), Mishra and Smyth (2015), Peet et al. (2015), and Serneels et al. (2017)). For instance, daily and weekly rates are converted to monthly rates by multiplying by a factor of 22 or 4, respectively. A concern of this approach is the possible introduction of measurement errors into the data (e.g., a person paid daily may not work 5 or 6 days each week), leading to inaccurate estimates on the returns to education. Measurement error which might arise because of conversion of the reported earnings to a common period leads to less efficient estimates of returns to education (Bound et al., 1994; Pischke, 1995). In addition, Card (2001) shows that using different measures of earnings yields different estimates of returns to education. However, we view pay period as indicating segmented labour markets not otherwise identified in the data. As explained in Chapter 2 (see descriptive in Table 2.10), workers with the same level of education earn different wages depending on their pay periods. Example, workers with secondary and higher education in Malawi earn higher if they work and get paid weekly. To the extent that the pay period may indicate different labour markets, pooling across pay periods leads to biased estimates of the rate of return to education (Fichtenbaum, 2006).

This essay analyses returns to education in Malawi, Tanzania and Uganda by answering two key questions. Firstly, when earnings are aggregated to a common unit, do different units give different estimates of returns to schooling? Secondly, does the pay period matter in estimating returns to schooling in East Africa?

Benefiting from relatively large and recent nationally representative datasets, this essay tests the unexplored hypotheses that estimates of returns to schooling depend on (i.e., vary according to) the period of measurement of the earnings and that different conversions may lead to different estimates. To the best of our knowledge, there have been no studies on Africa which have explored this issue. Given the absence of good and comparable instruments for education in the three countries to account for selection and endogeneity, we estimate returns by applying the Gaussian Copula (GC) instrument free method proposed in Park and Gupta (2012), combined with Heckman model for selection into employment categories.

Our findings suggest that returns to education differ by pay period and that pooling the periods together may lead to imprecise estimates. Specifically, in Malawi the returns for workers paid daily are the highest, followed by monthly and then weekly. In Tanzania, the returns for workers paid weekly are not only the highest but also increase at a higher rate than for the other pay periods. In Uganda, returns are highest for weekly earners followed by monthly and then daily. Our results also show that pooling/aggregating earnings to different common measures produces different returns and that estimates are generally closer to those from the pay period that constitutes the largest proportion of the sample. In this regard, our analysis suggests that estimating returns separately for workers paid over different periods is more reliable than pooling.

As explained earlier in Chapter 2, due to its unique labour market structure, with a disproportionately large proportion of the labour force in rural areas primarily participating in agriculture and or *ganyu*, we chose to analyse this group separately. The findings reveal that, generally, converting *ganyu* earnings to monthly yields larger estimates with larger standard errors than converting to daily or annualised suggesting that converting *ganyu* earnings to monthly gives less efficient estimates of returns to education.

The rest of the chapter is organised as follows: Section 3.2 provides an overview of the related literature. Section 3.3 describes the employed empirical methodology, followed by Section 3.4 on data and description. Section 3.5 presents the results and discussions, and Section 3.6 concludes.

3.2 Literature Review

Analysis of the relationship between education and earnings goes back to 1960s after the seminal works of Schultz (1961) and Becker (1964). In his theory of investment in human capital, Becker (1964) assumed that education raises earnings and productivity and that individuals choose the level of education that maximises the expected present value of their lifetime earnings net of costs for acquiring such levels of education. This theoretical analysis almost immediately triggered many empirical examinations with a debate about the true effect of education on earnings. The increase in earnings associated with an additional year of education is known as private returns to education or simply returns to education.

Mincer (1974), based on the human capital investment theory, developed a model for analysing the effect of education on wage earnings called the human capital earnings function (or Mincer wage function). This approach models the logarithm of wage earnings as a linear function of an individual's years of schooling, experience and experience squared. That is:

$$logW = a + bS + cE + dE^2 + \varepsilon \tag{3.1}$$

Where W is wage earnings, S is years of schooling completed, E is labour market experience, a, b,c and d are parameters, and ε is an error term. Since its formulation, this model has become the standard model for analysing returns to education, with many studies extending it to include more variables that affect wage earnings such as gender, race and work-related characteristics (Card, 1999, 2001; Patrinos and Psacharopoulos, 2010; Peet et al., 2015).

As an alternative, some studies employ a non-parametric (full discounting) method to estimate returns to education (see for example Heckman et al. (2006), Heckman et al. (2008), Heckman and Vytlacil (2007, 2005)). This method is relatively data demanding, which limits its applicability, especially in developing countries such as those in Sub-Saharan Africa (SSA). In that regard, it is beyond the scope of this section to describe the literature on this approach in detail.

One of the challenges with the Mincer model is how to estimate the causal effect of education on earnings with the endogeneity of education given unobserved ability. The consensus in the literature is that without controlling for individual ability, OLS on 3.1 gives inconsistent estimates (Cameron and Trivedi, 2005; Wooldridge, 2010). Economists have adopted various methods to address this problem. The most widely used solution for addressing this issue is to use instrumental variables (IV) based on either two-stage least squares (2SLS) or a control function (Cameron and Trivedi, 2009; Card, 1999, 2001). The method requires an additional variable (instrument) that affects an individual's education but is not correlated with their wage. Studies have employed different instruments, most frequently family background characteristics such as parental education, parents' occupation and spouse education; and school system features such as proximity to school, tuition fees, quality of the school, and (change in) compulsory schooling laws in minimum years of basic education (Card, 2001). To a large extent, the choice of instrument is dictated by availability of data.

Card (2001) reviewed 11 studies from developed countries conducted between 1990 and 2000 that relied on IV as the source of exogenous variation in education to obtain identification. Although in theory OLS estimates should be larger than their IV counterparts, that is, ability should bias the returns upwards (Card, 1999), the review concluded that estimates of returns to education from IV are larger than those from OLS. Similar conclusions were reached in a global review of literature on returns to education (Psacharopoulos and Patrinos, 2004, 2018). However, one of the critical limitations of the IV approach is that it is difficult to find variables that can generate exogenous variation in education in the study population.

An alternative solution is using instrument free methods, that is, methods which do not require any external instruments. They include latent instrumental variables (Ebbes et al., 2005); methods that use heteroscedasticity to obtain identification (Farré et al., 2013; Klein and Vella, 2009, 2010; Lewbel, 2012); and

Gaussian Copula (Park and Gupta, 2012). These methods are particularly useful when there are no (good) instruments in the data, for example, when a researcher uses survey data collected for other purposes and hence appropriate instruments have not been included in the questionnaire. To date, these methods have not yet been widely applied in the returns to education literature.

Ebbes et al. (2005) developed the Latent Instrumental Variables (LIV) method, which provides a means of obtaining consistent estimates in the presence of endogeneity without relying on external instruments. In this model, the instruments are unobserved and therefore estimated from the data. The model is also capable of testing if the regressor is correlated with the error term using the Hausman test. To demonstrate the superiority of the LIV approach over the traditional IV approaches, the study compared the estimates of returns to education from the two approaches using three datasets: the US National Longitudinal Survey of Young Men (NLSY), Brabant data and the University of Michigan Panel Study of Income Dynamics (PSID). In all datasets, the LIV approach found upward bias of OLS estimates of around 7% (close to 10% from studies of twins) while the IV approach gave different sizes of the bias for each dataset.

Klein and Vella (2009) formulated an IV free strategy and used it to estimate the causal effect of education on wages by utilising the presence of heteroscedasticity in the data to obtain identification. Their model consists of two non-parametric equations one for (the determinants of) education and the other for (the determinants of) wage. All determinants of education are used as regressors in the wage equation and the causal effect then obtained through heteroscedasticity (Klein and Vella, 2009). Applying the model to a sample of Australian workers from the 2001 Household Income and Labour Dynamics in Australia (HILDA), the strategy estimated returns at 10% compared to 6% for OLS.

Building on Klein and Vella (2009, 2010), Farré et al. (2013) formulated a parametric approach for estimating returns to education with endogenous education in the absence of internal instrumental variables. Applying the method to a subsample of American youth from the National Longitudinal Survey of Youth 1979, they estimated returns at 11.2% compared to 6.8% for OLS.

Lewbel (2012) introduced another instrument free method that utilises the presence of heteroscedasticity in the data to restrict the correlation between the regressors and the (product of the) errors in the regression model. Mishra and Smyth (2015) employed this method to estimate returns to schooling in China using two datasets: matched worker-firm data from Minhang Shanghai and the China Household Finance Survey (CHFS) 2011. For the first dataset, estimated returns to education were 7.4% using OLS and 25.7% using the Lewbel method. For the second dataset, the returns were estimated at 8.6% for OLS, 18.9% for traditional IV, 12.9% for Lewbel and 19.1% for Lewbel + traditional IV. In studies using heteroscedasticity-type instruments, OLS appeared to bias the returns downwards.

Another key challenge with the Mincer model is how to deal with sample selection bias. Sample selection arises because wages are observed only for individuals in wage employment who report positive values of wage during data collection. The wages of the wage earners might not reflect the wages of the non-wage income earners (for example, the unemployed, self-employed, agricultural workers) had they worked in wage employment. If the exclusion of these individuals from the analysis is not random, without controlling for how individuals select into wage employment the OLS estimator will give inconsistent estimates (Cameron and Trivedi, 2005; Heckman, 1979; Verbeek, 2004). The standard solution for this problem is to use the Heckman Two-step Sample Selection Model formulated by Heckman (1979). The model recovers consistent estimates by running OLS in two steps where the exclusion from the sample is modelled as an omitted variable. A detailed discussion of the method is provided in section 3.

While the debate on the issues/challenges in estimating the true causal effect of education on earnings remains, in the last few years many studies have emerged on developing countries including SSA. In a recent review of global literature on returns to education, Psacharopoulos and Patrinos (2018) document that in developing countries returns to an extra year of schooling averages about 9.2% compared to 8% in developed countries. These studies, however, widely differ in terms of methods (including OLS, traditional IV, propensity scores matching, and Heckman sample selection models) and data (such as nationally representative, regional or sectoral level data) making it difficult to directly compare the estimates of returns across countries and studies (see Appendix 3A. for the detailed analysis of selected studies on developing countries).

Within the IV and or sample selection literature in the developing countries, recent contributions include Nikolov and Jimi (2018), Kuepié and Nordman (2016), Wang (2013) and Aslam et al. (2012). Nikolov and Jimi (2018) used data from the 2014 Integrated Labour Force Survey (ILFS) and estimated returns to an extra year of schooling in Tanzania at 7% using the quarter of birth as an IV. Kuepié and Nordman (2016) employed a control function based IV approach (father's education and occupation/professional status as instruments) in their study on returns to education in two cities of the Republic of Congo. Data for the analysis came from the Employment and Informal Sector Survey (EESIC) 2009. They found that primary education had no effect on earnings in either of the cities, while returns for lower secondary, upper secondary and higher education were respectively 9%, 5% and 12% for Brazzaville and 9%, 14% and 13% for Pointe-Noire. In a study on China, Wang (2013) used data from the urban sample of the China Household Income Project (CHIP) 1995 and 2002 to examine the pattern of returns over time. The study employed family background characteristics (parental and spouse education) as instruments for education and found that returns increased over the two survey periods regardless of the instrument used. However, though the difference was small, returns seemed to be higher when parental education was used as an IV relative to spouse education.

Aslam et al. (2012) used data from the Purposive Household Survey in Punjab and the North West Frontier Province (NWFP) of Pakistan 2006 -2007 and found that an extra year of schooling increased males' earnings by about 10% while education had no significant effect on female's earnings. For more details on recent studies based on IV strategy in developing countries see Appendix 3A.

Another strand of literature on returns to education in developing countries is the one that focuses on examining the possible heterogeneity in returns to education along the earning distribution and across groups of workers (such as gender, sector of employment and location). CHuang and Lai (2017) examined returns to education in Taiwan between 1978 and 2003 using data from Taiwan's 1978-2003 Manpower Utilization Survey. Quantile regression results showed that returns increased from 5.5% in 1978 to 8.2% in 2003 with an average of 6.5% The returns were high at the low end of the earnings distribution and vice versa. A similar study by Stefani and Biderman (2009) on Brazil used data from Brazil National Household Survey (BNHS) 1988 and 1996 to examine the evolution of returns to education. The study also employed a quantile regression approach to analyse the pattern of returns to education along the earnings distribution and found that returns were heterogeneous across race, gender and earning distribution, ranging from 6% to 32%. Furthermore, Girma and Kedir (2005) used Household panel data for Ethiopia's seven major cities, 1994, 1995, and 1997 to examine returns across time and earnings distribution. Using the same methodology, they found that returns differed across the earning distribution: highest (20%) at 25th quantile and lowest at 90th quantile (9%).

As far as heterogeneity across groups is concerned, the typical finding (like for developed countries) is that females have higher returns to education compared to males (Nikolov and Jimi, 2018; Peet et al., 2015; Salisbury, 2016; Schultz, 2004); public sector employees have higher returns than their private counterparts (Lassibille and Tan, 2005); rural workers have higher returns than urban workers and wage employees have higher returns than the self-employed and agricultural workers (Al-Samarrai and Reilly, 2008).

A rather unique study is by Serneels et al. (2017), which examined whether the type of questionnaire used in collecting individuals' labour market information matters in estimating returns to education in Tanzania. By using both short and detailed questionnaires, the study found that returns differed by the survey instrument: short module questionnaires led to biased estimates compared to detailed questionnaires. After controlling for endogeneity due to unobserved ability and selection by using a control function, Heckman and Heckman-Hotz methods, the estimated returns ranged 20-21% for men and 32-49% for women for a year of post-primary school if short modules were used. For the detailed modules, no effects of schooling on wage were found for men, while returns for women were between 29% and 50%.

Whilst much effort has been put in addressing issues like endogeneity of education in estimating returns to education, heterogeneity of returns across the earnings distribution and groups of workers, little to no attention has been focused on whether the pay period matters in estimating returns to education. What is evident in all the previous studies is the conventional method of aggregating earnings to a common period such as hourly, daily, monthly or annual earnings. However, what is not clear is the impact this has on their findings. In this essay, we demonstrate that the relationship between earnings and education may vary across workers reporting wage earnings over different periods. We argue that the precision of converting the reported wages to the universal unit may be plagued by errors and assumptions made by the researcher, leading to inaccurate/inefficient estimates of the returns. In fact, different common measures may give different estimates of returns to education. Unlike previous studies, this essay considers the implication of alternative ways of converting the reported wages to a common unit/measure and provides separate estimates for each pay period.

3.3 Empirical Strategy

3.3.1 Overview

Recent studies on Tanzania (as discussed in section 2) show increasing returns with levels of education (convex schooling-earning function). We adopt the Mincer equation with quadratic schooling from Söderbom et al. (2014) to ascertain the possible convexity in returns. Thus, our empirical model is specified as follows:

$$Y_{it} = \alpha_1 S_{it} + \alpha_2 S_{it}^2 + \beta X_{it} + \mu_{it} \tag{3.2}$$

Where Y is the log of earnings, S is individual's years of schooling, S^2 schooling squared, X is a vector (containing a constant) of individual characteristics (age in years and its square, gender, location, log of weeks worked and a dummy variable for individuals observed more than once), i and t index individual and time respectively and μ is a standard error term. The parameters of interest are α_1 and α_2. The sign of α_2 tells us about the shape of the earning function: positive implies convexity, negative implies concavity, and zero implies linearity.

Since the rates of return to schooling may differ by level of education, we also use an alternative specification that uses dummies for completed levels of education to estimate returns to each level of education. Three levels are used for this purpose: primary, secondary and higher (including tertiary non-university (post-secondary diploma) and university). Because we have very few observations with higher than secondary education, we merge the two post-secondary education levels into one group. The following specification is used for this purpose:

$$Y_{it} = \delta educ_{it} + \gamma X_{it} + \varepsilon_{it} \tag{3.3}$$

Where *educ* is a vector of dummies for the levels of education with "less than primary education (no education hereafter)" as the reference category, X is as defined earlier, and ε an error term. The returns associated with each level of education with respect to the reference category is given by the vector δ. The returns per additional year of level l with respect to level m can be obtained using the following equation:

$$r_l = \frac{\delta_l - \delta_m}{S_l - S_m} \tag{3.4}$$

Where r is the return per year, $\delta_l - \delta_m$ is the difference in returns between the two levels and $S_l - S_m$ the difference in years of schooling between the levels.

In our specifications, we use age and its square in place of experience and its square for two main reasons. Firstly, the surveys did not explicitly ask the years of experience the individual spent in the current job. Therefore, defining experience as age less years of schooling less school starting age as commonly defined in the literature might result in accumulation of errors, especially if there were measurement errors in age, years of schooling and/or school starting age. Furthermore, we would have missing values for those who did not report their school starting age, or otherwise choose an arbitrary starting age. Secondly, if schooling happens to be endogenous due to, among other reasons, unobserved ability, by construction experience would also be endogenous. To avoid these issues, we use age as a proxy for experience in our analysis.

While most workers in the sample are paid monthly, significant shares report earnings daily and weekly (also fortnightly, quarterly for Tanzania). The standard method is to convert/aggregate all wages into a common period such as monthly wage or annualised wage (then expressed in monthly) earnings and use their log as the dependent variable. Having constructed construct three common measures for wages as described in Chapter 2, we begin by examining whether these different conversions give different estimates of returns to education. In the next step, we use one of the common measures and estimate the returns to education for each of the three main pay periods (daily, weekly and monthly) separately to examine if the estimates vary by pay period.

3.3.2 OLS Estimation

As a baseline estimation, we estimate (3.2) and (3.2) using OLS. According to the literature it is well known, however, that OLS will give inconsistent estimates of α_1, α_2 and δ because of omitted variables, measurement errors or if there is sample selection bias. A typical example is when these variables are correlated with the residuals in (3.2) and (3.2) due to the presence of other factors that are associated with higher education and higher wages but are not included in the models, such as when more educated individuals possess other unobservable characteristics, such as high ability, which are associated with higher wages. Estimating (3.2) and (3.2) using OLS without controlling for ability will lead to inconsistent estimates (Cameron and Trivedi, 2009). Furthermore, without controlling for how individuals select into wage employment, OLS will also give biased estimates.

3.3.3 Gaussian Copula Estimation

One of the standard solutions to recovering consistent estimates for (3.2) and (3.2) is using instrumental variables. Several studies, as discussed in the literature section, have employed different instruments for education in estimating returns to schooling. Frequently used instruments include family background characteristics such as parental education, parents' occupation and spouse education; and school system features such as proximity to school, tuition fees, quality of the school, and (change in) compulsory schooling laws in minimum years of basic education (Card, 2001). To a large extent, the choice of instruments is dictated by the availability of data, and almost every instrument is subject to debate.

LSMS being a general household survey, only family background characteristics were available for us to use as instruments for education. We tried two instruments from the data, but they did not meet the requirements[1] for a good instrument (i.e., turned to be weak and failed to meet the overidentification restriction).

[1]Results for IV estimation not included but are available upon request.

Whereas parental education variables were weak, household average education (which combines parental education, siblings education and spouse education) failed overidentification test.

Another solution is to use instrument free methods (which do not require external instruments) reviewed in section 3.2 above. Heteroscedasticity based methods are good candidates, however, these methods are only suitable when there is one endogenous regressor whilst in our case there are three (potential) endogenous regressors (S, S^2 and the log of number of weeks worked in the last 12 months). We therefore employ the Gaussian Copula (GC) method as it can be easily extended to include more than one endogenous regressor. The GC approach models the correlation between the suspected endogenous variable and the error term by using copulas[2]. By including the copula term(s) of the endogenous regressor(s) as additional regressor(s) in the regression model, this method recovers the estimates of the endogenous regressor(s) which are free from endogeneity (Park and Gupta, 2012).

Significance of the copula terms in the GC regression implies critical endogeneity; otherwise, OLS results are consistent. Furthermore, the sign of the copula terms shows the direction of the correlation between the endogenous variables and the errors. However, although the model can recover the true effect of the endogenous regressors, it does not tell anything about the source of the endogeneity (Hult et al., 2018). It is noteworthy that the method is not suitable when the endogenous regressor is binary (Park and Gupta, 2012). Consequently, we cannot use GC with the more flexible Mincer specification (3.3) that uses dummies for education levels to estimate returns to the levels of education. Instead, we rely on quadratic schooling in (3.2) to infer whether higher levels of education have higher returns than lower levels or not. We do, however, report the results for the levels of education corrected for selection bias.

Following Park and Gupta (2012) and Rutz and Watson (2019), our model is

[2] Cherubini et al. (2004) defines Copulas as "functions that enable us to express a joint probability distribution as a function of the marginal ones"

derived as follows: recall (3.2) where both (S and S^2 are endogenous (we omit the individual and time indices (subscripts it) for mathematical convenience).

$$Y = \alpha_1 S + \alpha_2 S^2 + \beta X + \mu \tag{3.5}$$

The relationship between the endogenous variables and the error term is modelled as:

$$\begin{pmatrix} S^* \\ S^{2*} \\ \mu^* \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{12} & \rho_{\mu 1} \\ \rho_{12} & 1 & \rho_{\mu 2} \\ \rho_{\mu 1} & \rho_{\mu 2} & 1 \end{bmatrix} \right)$$

Where $S^* = \Phi^{-1}(F_1(S))$ and $S^{2*} = \Phi^{-1}(F_2(S^2))$ are GC functions; $F_1(.)$ and $F_2(.)$ are cumulative distribution functions for S and S^2 respectively; ρ_{12} is the correlation between S and S^2; $\rho_{\mu 1}$ the correlation between S and μ; and $\rho_{\mu 2}$ the correlation between S^2 and μ. The expression can then be written as:

$$\begin{pmatrix} S^* \\ S^{2*} \\ \mu^* \end{pmatrix} \sim N \begin{pmatrix} 1 & 0 & 0 \\ \rho_{12} & \sqrt{1-\rho_{12}^2} & 0 \\ \rho_{\mu 1} & \frac{\rho_{\mu 2} - \rho_{12}\rho_{\mu 1}}{\sqrt{1-\rho_{12}^2}} & \sqrt{1 - \rho_{\mu 1}^2 - \frac{(\rho_{\mu 2} - \rho_{12}\rho_{\mu 1})^2}{\sqrt{1-\rho_{\mu 1}^2}}} \end{pmatrix} \cdot \begin{pmatrix} \varpi_1 \\ \varpi_2 \\ \varpi_3 \end{pmatrix}$$

Where

$$\begin{pmatrix} \varpi_1 \\ \varpi_2 \\ \varpi_3 \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

It then follows that,

$$\begin{aligned} \mu = \sigma_\mu \mu^* = \sigma_\mu \frac{\rho_{\mu 1} - \rho_{12}\rho_{\mu 2}}{\sqrt{1-\rho_{12}^2}} S^* + \sigma_\mu \frac{\rho_{\mu 2} - \rho_{12}\rho_{\mu 1}}{\sqrt{1-\rho_{12}^2}} S^{2*} + \\ \sigma_\mu \sqrt{1 - \rho_{\mu 1}^2 - \frac{(\rho_{\mu 2} - \rho_{12}\rho_{\mu 1})^2}{\sqrt{1-\rho_{\mu 1}^2}}} \varpi_3 \end{aligned} \tag{3.6}$$

Where σ_μ^2 is the variance of the error term. Combining (3.5) and (3.6) we get

$$Y = \alpha_1 S + \alpha_2 S^2 + \beta X + \sigma_\mu \frac{\rho_{\mu 1} - \rho_{12}\rho_{\mu 2}}{\sqrt{1-\rho_{12}^2}} S^* + \sigma_\mu \frac{\rho_{\mu 2} - \rho_{12}\rho_{\mu 1}}{\sqrt{1-\rho_{12}^2}} S^{2*} + \sigma_\mu \sqrt{1 - \rho_{\mu 1}^2 - \frac{(\rho_{\mu 2} - \rho_{12}\rho_{\mu 1})^2}{\sqrt{1-\rho_{\mu 1}^2}}} \varpi_3 \tag{3.7}$$

Equation (3.7) is a linear regression model with the error term given by its last three component. The model disaggregates the endogenous regressors into two components; one is the part not correlated with the error term (S and S^2) and the other is the part which is correlated with the error term (S^* and S^{2*}). By including the copula functions as additional regressors, OLS on model (3.7) gives consistent estimates for (S and (S^2) (Park and Gupta, 2012). Let $\theta_1 = \sigma_\mu \frac{\rho_{\mu 1} - \rho_{12}\rho_{\mu 2}}{\sqrt{1-\rho_{12}^2}}$; $\theta_2 = \sigma_\mu \frac{\rho_{\mu 2} - \rho_{12}\rho_{\mu 1}}{\sqrt{1-\rho_{12}^2}}$ and $\xi = \sigma_\mu \sqrt{1 - \rho_{\mu 1}^2 - \frac{(\rho_{\mu 2} - \rho_{12}\rho_{\mu 1})^2}{\sqrt{1-\rho_{\mu 1}^2}}} \varpi_3$.
Equation (3.7) can be rewritten as:

$$Y = \alpha_1 S + \alpha_2 S^2 + \beta X + \theta_1 S^* + \theta_2 S^{2*} + \xi \tag{3.8}$$

Given the discrete nature of our endogenous regressors, the distribution functions $F_1(.)$ and $F_2(.)$ are step functions lying between two values, such that:

$$F(t-1) < U_t < F(t)$$

for any discrete endogenous regressor t; where U_t follows uniform distribution on [0,1]. It follows, therefore, that:

$\Phi^{-1}(F_1(S-1)) < S^* < \Phi^{-1}(F_1(S))$; and $\Phi^{-1}(F_2(S^2-1)) < S^{2*} < \Phi^{-1}(F_2(S^2))$. Since $F_1(.)$ and $F_2(.)$ are estimable from the data, model (3.8) can be estimated using OLS.

Equation (3.8) can also be extended to include more endogenous regressors. For example, because the number of weeks worked in the last 12 months may be endogenous due to a bidirectional relationship between total annual earnings and the number of weeks worked, we include the variable "W" (for log weeks) when using annualised wages. In this case, our empirical model then becomes

$$Y = \alpha_1 S + \alpha_2 S^2 + \alpha_3 W + \beta X + \theta_1 S^* + \theta_2 S^{2*} + \theta_3 W^* + \xi \tag{3.9}$$

3.3.4 Heckman Selection Model

We also control for possible bias from non-random missingness in earnings data and selection into periods of employment. Some individuals in our dataset do not have values of wage, either because they were unemployed, self-employed at the time of survey or did not respond. As we cannot guarantee that exclusion of these individuals from our sample and analysis is random, our OLS estimator is likely to give inconsistent estimates due to sample selection bias (Cameron and Trivedi, 2005; Verbeek, 2004). The study, therefore, employs Heckman (1979) selection correction method to deal with selection bias.

It is worth noting that pay period is not exogenous as it is an outcome which itself might be the result of education. Higher educated individuals are more likely to have monthly-paid jobs . This might create a selection problem when estimating separately for each pay period, addressed by estimating selection equations into each pay period. Note that we use a different Heckman selection equation for each period, each one representing selection into that pay period. An alternative approach is to use a multinomial probit model, but we leave that open for future research. Because we want to correct the selection bias after controlling for other sources of endogeneity, we include GC terms in the two-step Heckman selection model. In the first stage (equation (3.9)) we estimate the probit model for selection into periods of payment and paid employment, the regressors being the exogenous variables, GC terms and exclusion restrictions:

$$P = \vartheta_1 S + \vartheta_2 S^2 + \Psi_1 X + \Psi_2 \Sigma + \phi_1 S^* + \phi_2 S^{2*} + e \tag{3.10}$$

Where P is the probability of participating in paid employment, Σ is the vector of exclusion characteristics (dummy for the household head (*head*), marital status

($married$), the proportion of children under 5 ($kids5$), and proportion of children between 6 and 14 years ($kids14$) in the household). P is defined as follows:

$$P = \begin{cases} 1 & \text{if } Y \geq 0 \\ 0 & \text{if } Y = . \end{cases} \tag{3.11}$$

We obtain the inverse mills ratio (λ) from (3.10) and then include it as a regressor in the estimation of (3.8) and (3.9) (again omitting individual and time indices for convenience). That is, the selection corrected equation for (3.8) and (3.9) are respectively given by (3.12) and (3.13):

$$(Y|P = 1) = \alpha_{11}S + \alpha_{21}S^2 + \beta_1 X + \theta_{11}S^* + \theta_{21}S^{2*} + \pi_1\lambda_1 + \xi_1 \tag{3.12}$$

$$(Y|P = 1) = \alpha_{12}S+\alpha_{22}S^2+\alpha_{32}W+\beta_2 X+\theta_{12}S^*+\theta_{22}S^{2*}+\alpha_{42}W^*+\pi_2\lambda_2+\xi_2 \tag{3.13}$$

The obtained estimates of the returns to schooling from (3.12) and (3.13) using OLS are consistent and efficient if π_1 and π_2 are significantly different from 0; otherwise, there is no selection problem, and thus GC is more efficient.

3.4 Variables and Descriptive Statistics

The literature guides the variables used in this study. Table 3.1 shows the names and definitions of each variable as used in the study. Refer to Chapter 2 for a more detailed description of how the variables were constructed from the survey data. Table 3.2 shows the summary statistics for the variables included in the wage regression function. Wages in this table are aggregated to annualised wage as described in Chapter 2. "Pooled" refers to the total sample when all pay periods are combined, and it excludes the off-own-farm casual labour (*ganyu*) for Malawi.

Table 3.2 shows that workers in Malawi earn more than those in Tanzania and Uganda across the pay periods. Because ganyu labour is excluded, it may mean that there are very few unskilled workers in the regular employment in Malawi and this partly explains the high wages. In all three countries, workers reporting earnings monthly are the highest earners (most of them may be in the formal employment hence the high rate). Workers reporting earnings weekly are the lowest earners in Tanzania and Uganda, whereas those reporting earnings daily are the lowest earners in Malawi. Compared to the other countries, in Tanzania, the wage penalty associated with working in daily or weekly employment is vast. That is, workers reporting earnings daily and weekly earn at least three times less than their monthly counterparts.

As far as education is concerned, workers in Malawi have more years of schooling compared to Tanzania and Uganda. Malawi has a more significant proportion (14%) of the workers reporting daily earnings holding higher education (these may be professionals given that ganyu workers are excluded) compared to Uganda (3%) and Tanzania (0%). Monthly earners have more education than their daily and weekly counterparts in all three countries. In Tanzania, workers with higher education are only paid monthly.

Table 3.1: Definition of Variables Used in the Analysis

Variable	Description
Wage equation:	
log(wage)	the logarithm of the common wage measure, as explained in Chapter 2.
sch	individual's total number of years of schooling. Its square is included to test convexity/concavity of the earnings function.
noeduc	educational dummy, 1 if less than primary education and 0 otherwise.
primary	educational dummy, 1 if completed primary education and 0 otherwise.
secondary	educational dummy, 1 if completed ordinary/advanced secondary education and 0 otherwise.
higher	educational dummy, 1 if completed post-secondary (diploma/university) education and 0 otherwise
age	individual's age in years. Its square is included to capture the non-linear relationship between earning and age.
female	a gender dummy, 1 for females, included to capture the effects of gender on wages.
rural	location dummy, 1 for employment in rural areas, is used to control for rural-urban wage differentials.
panel	for Tanzania and Uganda, a dummy, 1 for individuals observed more than once since we are using imperfect panel surveys.
year	only for Malawi, a year dummy, 1 for 2016 and 0 for 2010.
weeks	number of weeks worked in the past 12 months.
Selection equation	
married	dummy for marital status, 1 if married or living together and 0 otherwise.
head	dummy equals 1 if head of the household and 0 otherwise.
kids5	proportion of children under 6 years of age in the household.
kids14	proportion of children aged 6 to 14 years of age in the household.

Table 3.2: Summary Statistics for the Main Variables Used in Analysis

Country & Sample	Obs.	Wage ($ month)		sch		age		weeks		primary	secondary	higher	female	rural	panel	year
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	%	%	%	%	%	%	
Malawi																
Daily	182	206.42	251.48	9.03	4.51	36.80	10.20	36.75	13.99	20	29	12	26	67	NA	74
Weekly	505	174.26	289.04	7.15	3.91	34.94	10.35	35.51	14.71	24	10	4	30	66	NA	53
Monthly	5,129	226.84	339.20	9.26	4.12	35.85	10.71	39.63	12.69	26	24	14	25	49	NA	50
Pooled	5,816	221.04	332.38	9.05	4.16	35.79	10.66	39.13	13.01	26	23	13	25	52	NA	51
Ganyu	16,528	33.76	53.52	4.77	3.52	33.42	11.89	15.64	12.66	14	2	0	51	92	NA	64
Tanzania																
Daily	3,738	38.91	87.35	5.26	3.21	33.14	11.92	15.32	15.39	57	4	0	40	78	25	NA
Weekly	1,929	32.54	85.08	5.28	3.24	33.64	12.08	13.62	15.16	56	5	0	36	82	19	NA
Monthly	4,830	123.90	161.28	8.16	3.69	33.34	11.89	34.95	16	51	27	7	38	53	41	NA
Pooled	11,215	69.10	126.88	6.35	3.67	33.37	11.93	22.10	18.38	55	13	3	38	70	45	NA
Uganda																
Daily	1,262	78.92	103.44	6.35	3.53	30.99	10.84	36.04	14.57	34	11	3	20	64	39	NA
Weekly	589	76.34	109.07	6.97	3.81	32.77	11.44	35.27	14.87	32	12	8	29	72	15	NA
Monthly	2,765	114.19	139.02	9.85	4.39	34.71	11.06	39.38	12.45	29	16	31	37	56	56	NA
Pooled	4,631	99.66	127.84	8.51	4.41	33.46	11.18	37.89	13.54	31	14	20	31	61	57	NA

Source: Author's computations from IHS, TNPS and UNPS.
Note: The last two columns show % observed multiple times for Tanzania and Uganda (panel) and % in 2016 for Malawi (year) respectively

3.5 Results and Discussion

3.5.1 Effects of Aggregating Earnings on Estimates of Returns to Schooling

Firstly, we present the estimates of returns to schooling obtained from pooling all workers together (as previous studies for Africa have done). As explained in Chapter 2, the reported wage earnings were converted to three different common periods: daily earnings (DailyC), monthly earnings (MonthlyC) and annualised earnings expressed per month (MonthlyA). Note that Malawi's off-own-farm labour (*Ganyu*) is excluded in this analysis and is analysed separately in the next section. Tables 3.3 - 3.5 compare the estimates of returns to schooling for the three countries when these three measures are used as the dependent variables in the regressions. The first three columns present the estimates from the baseline OLS regression (ignoring the possible endogeneity bias). The next three columns (columns $4-6$) present estimates corrected for endogeneity due to ability bias using GC model by Park and Gupta (2012). The last three columns (columns 7-9) present estimates corrected for both ability bias and selection into employment categories using Heckman sample selection model in combination with GC (HGC henceforth). To simplify comparison, the predicted average marginal effects of schooling (AME(sch)) are included in the tables since the quadratic component of years of schooling may complicate the interpretation.

As GC and HGC tend to produce unstable estimates when the endogenous regressors are discrete, bootstrap aggregating (bagging) is also employed to check the robustness of both the GC and HGC results, by estimating the regressions many times and averaging the coefficients. The results for bagging are presented in Appendix 3D and the results in Tables 3.3 - 3.5 appear robust.

Tables 3.3 - 3.5 show that for Malawi and Tanzania the coefficient of schooling (sch) is negative implying that there is a threshold in the years of schooling (about six years for Malawi and two years for Tanzania) below which the returns are negative. Irrespective of the measure of earnings or the estimation strategy used, Tables 3.3 - 3.5 show that in all countries the coefficient on schooling squared (*sch2*) is positive and highly statistically significant implying a strong convex relationship between earnings and years of schooling. This tells us that, while each additional year of education is associated with an increase in earnings, the rate of increase in earnings also increases with years of schooling. That is, the slope of the earnings function increases by some constant amount for each additional year of schooling.

In line with theory on ability bias but contrary to the consensus in the IV literature on returns to education, Tables 3.3 -3.5 show that OLS gives upward biased estimates. The predicted marginal effects of schooling in Table 3.3 -3.5 show that including the copula functions for education lowers the returns to education in Malawi by about 50% from 13.8% to 6.0% when the reported earnings are aggregated to daily earnings; from 14.7% to 6.9% when the reported earnings are aggregated to monthly earnings; and from 15.3% to 8.8% when the reported earnings are aggregated to annualised earnings. Correcting for selection to employment categories in addition to ability bias lowers the returns even further (daily to 4.1%, monthly to 4.8% and annualised to 7.2%). The coefficient of the inverse mills ratio (IMR) in Tables 3.3 and 3.4 is statistically significant, implying that ignoring selection leads to biased results. The negative (positive) sign of IMR implies that there are negative (positive) correlations between the errors in the wage equations and those from the labour force participation equations making OLS results inconsistent. That is, there are unobserved factors that increase (decrease) the likelihoods of both participation in wage employment and earning lower (higher).

Importantly, Tables 3.3 - 3.5 show the effects of using different earnings measures on the estimates of returns to schooling; estimates clearly differ depending on how earnings are measured. Table 3.3 shows that MonthlyA gives larger estimates of returns to schooling in Malawi compared to DailyC or MonthlyC. There is a small (negligible) difference between estimates from DailyC and to MonthlyC (mainly due to the small proportion of the daily earners relative to monthly earners in the sample). The pattern is irrespective of the estimation strategy used. The top panel of Figure 3.1 plots the HGC (preferred) estimates of returns to schooling from Table 3.3 for the selected grades. It shows how the estimates differ with the measure of earnings

Table 3.4 shows the corresponding results for Tanzania. For Tanzania, MonthlyC gives larger estimates of returns to schooling compared to DailyC or MonthlyA. In addition, the strength of correlation between years of schooling and the error terms in the regressions is significant and stronger for MonthlyC compared to DailyC and MonthlyA. The middle panel of Figure 3.1 plots the HGC estimates of returns to schooling in Tanzania. While the gap between the estimates from DailyC and MonthlyC is generally constant, that between estimates from MonthlyC and MonthlyA increases with education.

Table 3.5 shows the results for Uganda. Like the case for Malawi, MonthlyA gives larger estimates of returns to schooling compared to aggregating to DailyC or MonthlyC. Like Malawi and Tanzania, the correlation between years of schooling and the error terms in the regressions exists and is generally significant. The bottom panel of Figure 3.1 plots the HGC estimates of returns to schooling in Uganda. In Uganda, the gap between the estimates from DailyC and MonthlyC is also generally constant, while that between estimates from MonthlyC and DailyC as well as between MonthlyC and MonthlyA increases with education.

These results, therefore, raise a concern that the choice of the conversion of earnings matters in estimating returns to schooling in developing countries. The estimates will depend on whether the reported earnings are aggregated to daily, monthly or annualised earnings.

Table 3.3: Effects of Aggregating Earnings on Estimates of Returns to Schooling - Malawi

	OLS			GC			HGC		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	-0.077***	-0.073***	-0.066***	-0.109***	-0.106***	-0.093***	-0.110***	-0.107***	-0.094***
	(0.009)	(0.008)	(0.008)	(0.011)	(0.011)	(0.011)	(0.011)	(0.010)	(0.010)
sch2	0.011***	0.012***	0.012***	0.009***	0.009***	0.010***	0.008***	0.008***	0.009***
	(0.001)	0.000	0.000	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
age	0.063***	0.059***	0.055***	0.064***	0.060***	0.056***	0.054***	0.049***	0.048***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.006)
age2/100	-0.063***	-0.056***	-0.051***	-0.064***	-0.057***	-0.052***	-0.052***	-0.044***	-0.043***
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
female	-0.134***	-0.136***	-0.105***	-0.134***	-0.136***	-0.106***	-0.059*	-0.052	-0.044
	(0.023)	(0.022)	(0.021)	(0.022)	(0.021)	(0.021)	(0.035)	(0.033)	(0.032)
rural	-0.136***	-0.209***	-0.207***	-0.135***	-0.209***	-0.209***	-0.060*	-0.124***	-0.146***
	(0.021)	(0.020)	(0.019)	(0.021)	(0.019)	(0.019)	(0.032)	(0.032)	(0.031)
year	1.334***	1.239***	1.236***	1.334***	1.239***	1.235***	1.351***	1.258***	1.250***
	(0.020)	(0.019)	(0.018)	(0.020)	(0.019)	(0.019)	(0.021)	(0.020)	(0.019)
weeks			1.124***			1.149***			1.144***
			(0.018)			(0.029)			(0.027)
Copula(sch)				0.146***	0.149***	0.123***	0.151***	0.155***	0.127***
				(0.047)	(0.045)	(0.045)	(0.048)	(0.046)	(0.045)
Copula(sch2)				0.154***	0.154***	0.128***	0.157***	0.157***	0.131***
				(0.049)	(0.047)	(0.046)	(0.050)	(0.049)	(0.050)
Copula(weeks)						1.149***			-0.005
						(0.029)			(0.004)
IMR							-0.164***	-0.184***	-0.136***
							(0.054)	(0.051)	(0.050)
Constant	-0.438***	2.601***	-1.743***	0.105	3.150***	-1.362***	0.545***	3.645***	-0.981***
	(0.118)	(0.112)	(0.121)	(0.169)	(0.158)	(0.169)	(0.204)	(0.191)	(0.212)
AME(sch)	0.138***	0.147***	0.153***	0.060***	0.069***	0.088***	0.041**	0.048***	0.072***
	(0.003)	(0.003)	(0.003)	(0.015)	(0.014)	(0.014)	(0.016)	(0.015)	(0.015)
Obs.	5,816	5,816	5,816	5,816	5,816	5816	5,816	5,816	5,816
R^2	0.59	0.62	0.73	0.59	0.62	0.74			

Notes: Copula() are Gaussian Copula functions. The copula functions for schooling are positive and significant implying positive and significant correlation between schooling variables and the errors in the regression models. IMR is the inverse mills ratio. AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * p<0.10, * p<0.05, ** p<0.01.

Table 3.4: Effects of Aggregating Earnings on Estimates of Returns to Schooling - Tanzania

	OLS			GC			HGC		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	-0.029***	0.004	-0.017***	-0.033***	-0.031***	-0.033***	-0.028***	-0.010	-0.022**
	(0.006)	(0.007)	(0.006)	(0.010)	(0.012)	(0.010)	(0.009)	(0.012)	(0.010)
sch2	0.009***	0.010***	0.009***	0.009***	0.010***	0.009***	0.008***	0.007***	0.007***
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	(0.001)	0.000
age	0.075***	0.065***	0.067***	0.075***	0.065***	0.066***	0.064***	0.027***	0.046***
	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)	(0.005)
age2/100	-0.079***	-0.067***	-0.070***	-0.079***	-0.067***	-0.070***	-0.065***	-0.017**	-0.043***
	(0.005)	(0.007)	(0.006)	(0.005)	(0.007)	(0.006)	(0.006)	(0.008)	(0.007)
female	-0.445***	-0.641***	-0.554***	-0.445***	-0.642***	-0.549***	-0.349***	-0.303***	-0.370***
	(0.017)	(0.021)	(0.018)	(0.017)	(0.022)	(0.019)	(0.028)	(0.039)	(0.031)
rural	-0.171***	-0.664***	-0.339***	-0.171***	-0.662***	-0.330***	-0.136***	-0.538***	-0.270***
	(0.018)	(0.022)	(0.020)	(0.017)	(0.020)	(0.018)	(0.020)	(0.025)	(0.022)
panel	-0.103***	-0.001	-0.085***	-0.104***	-0.002	-0.087***	-0.095***	0.030	-0.068***
	(0.017)	(0.021)	(0.018)	(0.017)	(0.021)	(0.018)	(0.017)	(0.022)	(0.019)
weeks			1.123***			1.078***			1.069***
			(0.008)			(0.011)			(0.011)
Copula(sch)				-0.011	0.076**	0.039	-0.010	0.080**	0.041
				(0.027)	(0.033)	(0.028)	(0.027)	(0.032)	(0.028)
Copula(sch2)				0.032	0.084***	0.038	0.029	0.076**	0.035
				(0.026)	(0.033)	(0.028)	(0.026)	(0.033)	(0.029)
Copula(weeks)						0.029***			0.030***
						(0.005)			(0.005)
IMR							-0.237***	-0.837***	-0.446***
							(0.052)	(0.075)	(0.059)
Constant	-0.181**	2.584***	-1.340***	-0.147	2.841***	-1.105***	0.233*	4.182***	-0.368**
	(0.073)	(0.090)	(0.079)	(0.091)	(0.112)	(0.100)	(0.124)	(0.168)	(0.144)
AME(sch)	0.089***	0.142***	0.104***	0.084***	0.102***	0.083***	0.079***	0.086***	0.075***
	(0.002)	(0.003)	(0.002)	(0.009)	(0.011)	(0.009)	(0.009)	(0.011)	(0.010)
Obs.	11,215	11,215	11,215	11,215	11,215	11,215	11,215	11,215	11,215
pp R^2	0.27	0.37	0.78	0.27	0.37	0.78			

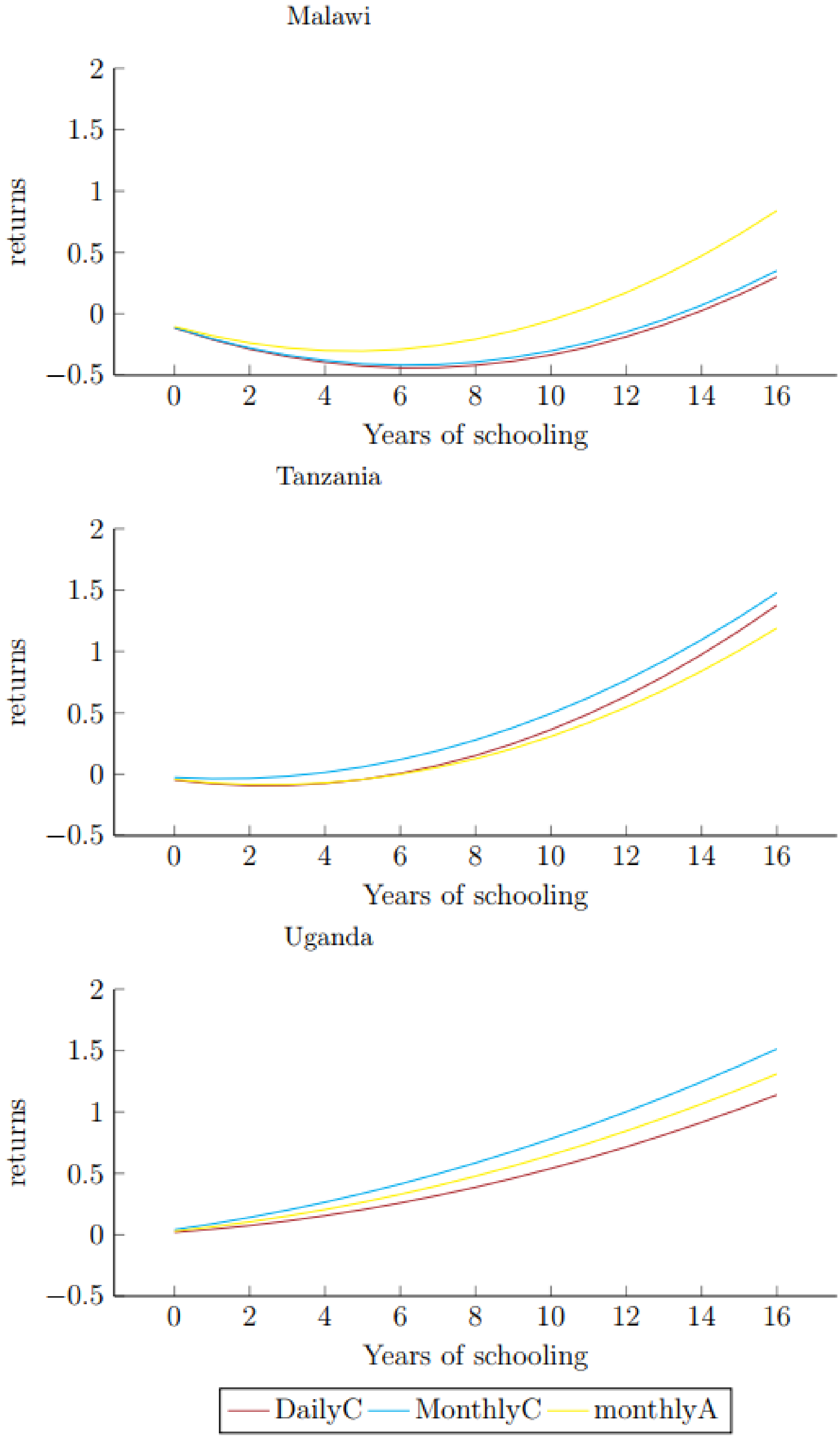
Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models. IMR is the inverse mills ratio. AME(sch) is the average marginal effects of schooling. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5: Effects of Aggregating Earnings on Estimates of Returns to Schooling - Uganda

	OLS			GC			HGC		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	0.061***	0.074***	0.069***	0.020	0.040*	0.030	0.019	0.041**	0.029
	(0.011)	(0.012)	(0.011)	(0.019)	(0.021)	(0.020)	(0.019)	(0.020)	(0.020)
sch2	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
age	0.080***	0.077***	0.068***	0.081***	0.077***	0.068***	0.082***	0.076***	0.070***
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)
age2/100	-0.090***	-0.088***	-0.077***	-0.090***	-0.088***	-0.077***	-0.092***	-0.086***	-0.079***
	(0.009)	(0.010)	(0.010)	(0.009)	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)
female	-0.443***	-0.451***	-0.438***	-0.439***	-0.448***	-0.435***	-0.458***	-0.422***	-0.459***
	(0.028)	(0.030)	(0.029)	(0.029)	(0.029)	(0.028)	(0.054)	(0.058)	(0.055)
rural	-0.224***	-0.293***	-0.239***	-0.221***	-0.290***	-0.239***	-0.237***	-0.267***	-0.259***
	(0.027)	(0.029)	(0.028)	(0.027)	(0.029)	(0.029)	(0.050)	(0.053)	(0.051)
panel	0.145***	0.212***	0.153***	0.146***	0.213***	0.155***	0.144***	0.215***	0.153***
	(0.028)	(0.030)	(0.029)	(0.030)	(0.032)	(0.030)	(0.031)	(0.032)	(0.031)
weeks			1.153***			1.172***			1.172***
			(0.022)			(0.038)			(0.036)
Copula(sch)				0.151**	0.129*	0.144*	0.150**	0.129*	0.144*
				(0.073)	(0.076)	(0.074)	(0.072)	(0.076)	(0.074)
Copula(sch2)				0.045	0.034	0.042	0.043	0.036	0.040
				(0.073)	(0.079)	(0.077)	(0.071)	(0.076)	(0.074)
Copula(weeks)						-0.005			-0.005
						(0.006)			(0.006)
IMR							0.045	-0.061	0.055
							(0.113)	(0.119)	(0.114)
constant	-1.025***	1.918***	-2.218***	-0.677***	2.208***	-1.933***	-0.745***	2.300***	-2.018***
	(0.128)	(0.135)	(0.143)	(0.180)	(0.196)	(0.214)	(0.255)	(0.259)	(0.270)
AME(sch)	0.113***	0.126***	0.122***	0.069***	0.090***	0.081***	0.072***	0.087***	0.084***
	(0.003)	(0.003)	(0.003)	(0.018)	(0.019)	(0.018)	(0.019)	(0.020)	(0.019)
Obs.	4,631	4,631	4,631	4,631	4,631	4,631	4,631	4,631	4,631
R^2	0.36	0.38	0.60	0.36	0.38	0.60			

Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models. IMR is the inverse mills ratio. AME(sch) is the average marginal effects of schooling. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3.1: Effects of Aggregating Earnings on Estimates of Returns to Schooling



Tables 3.6-3.8 present results when using the level of education attained instead of completed years of schooling. GC and HGC are not suitable here, but we present the results corrected for endogenous selection to the pay periods. The results are consistent with those from using years of schooling. Importantly, the different measures of earnings yield different estimates of the returns to the levels of education as observed when using years of schooling. MonthlyA gives larger estimates for Malawi and Uganda, while MonthlyC gives larger estimates for Tanzania. Higher levels of education are associated with higher returns, implying a convex relationship between earnings and education. Whether coefficients accounting for selection bias are higher or lower than OLS varies by pay period, level and country.

Table 3.9 shows how the coefficients of education from Tables 3.3-3.8 compare to other studies in the three countries. Despite the methodological differences, choice of measure of earnings and sample coverage, the results from this study are consistent with the previous studies in these countries, in the sense that returns increase with years of schooling and generally with levels of education (convex returns). The previous studies for Malawi reported in Table 3.9 found that having primary education increases earnings by 12% - 78%, secondary education increases earnings by 35% - 110%, and tertiary increases earnings by 150% - 192%. For Tanzania, previous studies estimated returns for primary education at 2.5% - 65%, secondary education at 41% - 169% and tertiary at 109% - 203%. For Uganda, returns for primary are estimated at 23% - 58%, for secondary at 74% - 119%, and tertiary at 119% - 210%. Although the coefficient estimates vary significantly across studies within each country, our results for all three countries fall well within the respective country's range.

In the next subsection, we estimate returns by pay period to show that the pooling employed by others doesn't capture returns for daily and weekly; and that pay period may help to capture labour market segmentation that cannot otherwise be identified in the data. The latter is a strength as data in developing countries are usually missing (or questions not asked) to identify segmented labour markets or even formal versus informal employment. We argue that we are providing a way to capture returns to different categories of workers (that may not correspond to occupations even if included in data).

Table 3.6: Effects of Aggregating Earnings (Levels of Education) - Malawi

	OLS			Heckman		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
primary	0.236***	0.274***	0.321***	0.180***	0.218***	0.283***
	(0.026)	(0.024)	(0.024)	(0.032)	(0.030)	(0.029)
secondary	0.693***	0.794***	0.851***	0.563***	0.663***	0.764***
	(0.027)	(0.025)	(0.025)	(0.051)	(0.048)	(0.047)
higher	1.443***	1.549***	1.601***	1.256***	1.359***	1.474***
	(0.032)	(0.030)	(0.030)	(0.070)	(0.066)	(0.065)
age	0.068***	0.063***	0.059***	0.057***	0.053***	0.052***
	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)
age2/100	-0.069***	-0.062***	-0.057***	-0.058***	-0.050***	-0.049***
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
female	-0.133***	-0.134***	-0.104***	-0.054	-0.055*	-0.051
	(0.023)	(0.022)	(0.021)	(0.035)	(0.033)	(0.032)
rural	-0.158***	-0.229***	-0.227***	-0.081**	-0.150***	-0.174***
	(0.021)	(0.019)	(0.019)	(0.033)	(0.031)	(0.031)
year	1.297***	1.201***	1.198***	1.316***	1.220***	1.210***
	(0.020)	(0.019)	(0.018)	(0.021)	(0.019)	(0.019)
weeks			1.118***			1.115***
			(0.018)			(0.018)
IMR				-0.168***	-0.170***	-0.115**
				(0.056)	(0.053)	(0.052)
Constant	-0.445***	2.625***	-1.672***	-0.042	3.033***	-1.389***
	(0.113)	(0.106)	(0.116)	(0.176)	(0.165)	(0.173)
Obs.	5,816	5,816	5,816	5,816	5,816	5,816
R^2	0.59	0.62	0.73	0.59	0.62	0.74

Standard errors in parentheses * p<0.10, * p<0.05, *** p<0.01.

Table 3.7: Effects of Aggregating Earnings (Levels of Education) - Tanzania

	OLS			Heckman		
	DailyC	MonthlyC	MonthlyA	Daily	MonthlyC	MonthlyA
primary	0.202***	0.402***	0.263***	0.200***	0.394***	0.261***
	(0.020)	(0.025)	(0.021)	(0.020)	(0.024)	(0.021)
secondary	0.797***	1.365***	0.962***	0.755***	1.217***	0.892***
	(0.027)	(0.033)	(0.029)	(0.028)	(0.035)	(0.031)
higher	1.613***	2.239***	1.762***	1.489***	1.798***	1.541***
	(0.048)	(0.059)	(0.051)	(0.055)	(0.068)	(0.059)
age	0.074***	0.063***	0.065***	0.063***	0.026***	0.046***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.006)	(0.005)
age2/100	-0.077***	-0.063***	-0.067***	-0.063***	-0.014*	-0.042***
	(0.005)	(0.007)	(0.006)	(0.006)	(0.008)	(0.007)
female	-0.460***	-0.675***	-0.572***	-0.365***	-0.341***	-0.401***
	(0.017)	(0.021)	(0.018)	(0.027)	(0.034)	(0.029)
rural	-0.179***	-0.677***	-0.341***	-0.145***	-0.558***	-0.285***
	(0.018)	(0.022)	(0.020)	(0.019)	(0.024)	(0.021)
panel	-0.106***	-0.005	-0.090***	-0.097***	0.027	-0.072***
	(0.017)	(0.021)	(0.018)	(0.017)	(0.021)	(0.018)
weeks			1.132***			1.124***
			(0.008)			(0.008)
IMR				-0.233***	-0.824***	-0.427***
				(0.052)	(0.065)	(0.056)
Constant	-0.117*	2.770***	-1.260***	0.260**	4.104***	-0.551***
	(0.071)	(0.089)	(0.078)	(0.111)	(0.137)	(0.122)
Obs.	11,215	11,215	11,215	11,215	11,215	11,215
R^2	0.26	0.36	0.78	0.26	0.37	0.78

Standard errors in parentheses * p<0.10, * p<0.05, *** p<0.01.

Table 3.8: Effects of Aggregating Earnings (Levels of Education) - Uganda

	OLS			Heckman		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
primary	0.515***	0.570***	0.557***	0.515***	0.551***	0.559***
	(0.033)	(0.035)	(0.034)	(0.036)	(0.038)	(0.037)
secondary	0.634***	0.764***	0.719***	0.635***	0.746***	0.721***
	(0.044)	(0.046)	(0.044)	(0.046)	(0.049)	(0.047)
higher	1.271***	1.415***	1.367***	1.273***	1.339***	1.376***
	(0.039)	(0.041)	(0.040)	(0.076)	(0.081)	(0.078)
age	0.084***	0.082***	0.072***	0.084***	0.079***	0.072***
	(0.007)	(0.008)	(0.007)	(0.008)	(0.008)	(0.008)
age2/100	-0.097***	-0.096***	-0.084***	-0.098***	-0.091***	-0.085***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)
female	-0.461***	-0.470***	-0.456***	-0.463***	-0.421***	-0.461***
	(0.029)	(0.031)	(0.030)	(0.051)	(0.054)	(0.052)
rural	-0.281***	-0.354***	-0.298***	-0.282***	-0.309***	-0.303***
	(0.028)	(0.029)	(0.028)	(0.047)	(0.050)	(0.048)
panel	0.158***	0.229***	0.167***	0.158***	0.233***	0.167***
	(0.029)	(0.031)	(0.030)	(0.029)	(0.031)	(0.030)
weeks			1.158***			1.158***
			(0.023)			(0.023)
IMR				0.003	-0.114	0.012
				(0.097)	(0.103)	(0.099)
Constant	-0.742***	2.246***	-1.920***	-0.747***	2.405***	-1.938***
	(0.124)	(0.132)	(0.141)	(0.185)	(0.195)	(0.201)
Obs.	4,631	4,631	4,631	4,631	4,631	4,631
R^2	0.34	0.36	0.58	0.34	0.36	0.58

Standard errors in parentheses * p<0.10, * p<0.05, *** p<0.01.

Table 3.9: Selected studies on Returns to Education in East Africa

Country & Study	Measure of Earnings	Estimator	Years of Education	Primary	Secondary	Higher
Malawi						
This study	DailyC	HGC/Heckman	0.041	0.18	0.563	1.256
	MonthlyC	HGC/Heckman	0.048	0.218	0.663	1.359
	MonthlyA	HGC/Heckman	0.072	0.283	0.764	1.474
Kim (2020)	Hourly earnings	Heckman	NA	0.183	0.348	1.894
	Monthly earnings	Heckman	NA	0.121	0.475	1.923
Peet et al. (2015)	Annual earnings	OLS	0.12	0.784	1.152	1.536
Tanzania						
This study	DailyC	HGC/Heckman	0.079	0.2	0.755	1.489
	MonthlyC	HGC/Heckman	0.086	0.394	1.217	1.798
	MonthlyA	HGC/Heckman	0.075	0.261	0.892	1.541
Nikolov and Jimi (2018)	Monthly earnings	OLS	0.12	0.44	1.685	NA
Serneels et al. (2017)	Daily earnings	OLS	0.08	NA	NA	NA
Baffour (2013)	Monthly earnings	OLS	NA	0.651	1.292	1.986
Bridges et al. (2017)	Monthly earnings	FE	NA	0.025	0.412	NA
Peet et al. (2015)	Annual earnings	OLS	0.11	0.053	0.578	1.087
Kahyarara and Teal (2008)	Monthly earnings	control function	NA	0.159	0.845	2.032
Uganda						
This study	DailyC	HGC/Heckman	0.072	0.515	0.635	1.273
	MonthlyC	HGC/Heckman	0.087	0.551	0.746	1.339
	MonthlyA	HGC/Heckman	0.084	0.559	0.721	1.376
Kavuma et al. (2015)	Monthly earnings	OLS	0.161	0.34	1.042	2.101
Peet et al. (2015)	Annual earnings	OLS	0.12	0.576	1.186	1.544
Cuaresma and Raggl (2016)	Hourly earnings	Heckman	0.069	0.231	0.736	1.188

3.5.2 Effects of Pay Period on Estimates of Returns to Schooling

The results presented in sections 3.5.1 show that if we convert the reported earnings to one common unit, MonthlyC leads to larger estimates of returns to education in Tanzania (compared to DailyC or MonthlyA) while for Malawi and Uganda converting to MonthlyA leads to larger estimates. In this section, we present the results for the samples for each of the pay periods to see whether the returns to education vary depending on the period in which workers are paid. Owing to its ability to allow for seasonal workers who only work some months in a year and some weeks in a month, we choose MonthlyA as our preferred common earnings measure and use it in examining returns by pay period. We focus our discussion of the results on the endogenous corrected (both for ability and selection) results, though we include OLS results for comparison. The corresponding GC results and results for the first stage HGC regressions are available in the Appendixes 3B and 3C while the corresponding bootstrap aggregating results are reported in Appendix 3D.

Table 3.10 shows results by pay period for Malawi. The last three columns of Table 3.10 shows that even after correcting for ability bias and selection, the coefficient of schooling is negative and significant across the pay periods which means there is a threshold below which the returns to education are negative. This threshold varies by pay period: four years for workers reporting earnings daily and monthly, and seven years for those reporting earnings weekly.

Comparing the returns to schooling from the different pay periods, workers reporting earnings daily are associated with higher returns to education than their weekly and monthly counterparts. More specifically, the average marginal effects indicate that an extra year of schooling raises earnings by 11.7% if they report earnings daily, 5.7% if report weekly, and 8.4% if report earnings monthly.

Table 3.10: Effects of Pay Period on Estimates of Returns to Schooling - Malawi

	OLS			HGC		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
sch	-0.002	-0.095***	-0.066***	-0.018	-0.133**	-0.091***
	(0.076)	(0.030)	(0.008)	(0.088)	(0.052)	(0.010)
sch2	0.006	0.013***	0.012***	0.007	0.012***	0.009***
	(0.005)	(0.002)	0.000	(0.007)	(0.002)	(0.001)
age	0.041	0.076***	0.053***	0.060	0.079***	0.045***
	(0.059)	(0.021)	(0.006)	(0.080)	(0.028)	(0.006)
age2/100	-0.035	-0.081***	-0.048***	-0.061	-0.083**	-0.039***
	(0.074)	(0.027)	(0.007)	(0.100)	(0.034)	(0.008)
female	0.014	-0.098	-0.116***	-0.107	-0.129	-0.041
	(0.210)	(0.078)	(0.022)	(0.289)	(0.105)	(0.033)
rural	-0.286	-0.322***	-0.185***	-0.297	-0.308***	-0.106***
	(0.204)	(0.079)	(0.019)	(0.227)	(0.091)	(0.033)
year	0.850***	1.020***	1.278***	0.971***	1.002***	1.301***
	(0.212)	(0.073)	(0.018)	(0.358)	(0.083)	(0.020)
weeks	1.259***	1.037***	1.138***	1.371***	1.226***	1.142***
	(0.166)	(0.052)	(0.019)	(0.221)	(0.081)	(0.029)
Copula(sch)				-0.497	-0.046	0.095**
				(0.463)	(0.235)	(0.048)
Copula(sch2)				0.534	0.277	0.134***
				(0.510)	(0.202)	(0.045)
Copula(weeks)				-0.025	-0.064***	-0.001
				(0.045)	(0.017)	(0.005)
IMR				0.327	0.020	-0.163***
				(0.843)	(0.271)	(0.052)
Constant	-2.01	-1.224***	-1.834***	-3.664	-1.467	-1.003***
	(1.256)	(0.419)	(0.121)	(3.887)	(1.144)	(0.223)
AME(sch)	0.101***	0.113***	0.161***	0.117	0.057	0.084***
	(0.026)	(0.009)	(0.003)	(0.124)	(0.057)	(0.015)
Obs.	182	505	5,129	182	505	5,129
R^2	0.44	0.66	0.77			

Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models. IMR is the inverse mills ratio. AME(sch) is the average marginal effects of schooling. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

After disaggregating the sample to pay periods, the coefficients of the copula functions and the inverse mills ratio in Table 3.10 are insignificant for daily and weekly suggesting that the correlation between earnings and the error terms observed earlier are associated with only the monthly sample.

The implied returns from Table 3.10 for the selected years of schooling are shown graphically in the top panel of Figure 3.2. The naïve estimates (pooled) from the last column of Table 3.3 are also included for comparison. Except for monthly, the pattern and slope of the curves for each pay period are different from that of the pooled curve implying that each period has different returns to education and ignoring this would lead to biased estimates. For monthly, it can be explained by the fact that it constitutes about 88% of the sample and thus pooling the periods together would very likely bias the returns in the direction of monthly.

Table 3.11 shows the results for Tanzania. Like Malawi, the coefficient of schooling for Tanzania is negative throughout which suggests that there is also a threshold below which the returns to education are negative (although this is only a few years of education); and the correlation correlation between schooling and the error terms is significant only for the monthly sample. When ability and selection biases are accounted for, there are mixed results: the estimates of returns for the monthly decrease while for daily and weekly increase (see the AME(sch) in Table 3.11. This suggests that the way endogeneity affects OLS results is not homogenous across the pay periods. For instance, unlike OLS, HGC results show that returns for monthly are lower than for daily earners and the difference increases with education (consistent with a particular level of education needed to secure a job paid monthly but does not then affect earnings). This indicates that selection was biasing the returns to schooling downwards for the daily, while for monthly selection was biasing the returns upwards. As it is essential to take into account the *sch2* term, plotting returns over the range of years of education reveals the pattern. The middle panel of Figure 3.2 plots the implied returns from Table 3.11 (marginal effects of schooling on earnings, pooled estimates derived from Table 3.7). Returns for the weekly earners are not only higher but also increase at a higher rate than the other periods (reflecting the higher coefficient on *sch2*).

Table 3.12 shows the effects of the pay period on the estimates of returns to schooling in Uganda. While the coefficients on *sch* are positive across the pay periods, the coefficient of on *sch2* for those reporting daily earnings is negative, implying concave returns to education, that is, the returns to an extra year of education decreases as one acquires more schooling. The concavity persists even after accounting for ability bias and selection. The bottom panel of Figure 3.2 plots the implied returns from Table 3.12. The patterns of the curves for each pay period are very different from that of the pooled curve, implying that each pay period has different returns to education.

Table 3.11: Effects of Pay Period on Estimates of Returns to Schooling - Tanzania

	OLS			HGC		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
sch	-0.023	-0.026	0.033***	-0.045**	-0.033	-0.015
	(0.015)	(0.017)	(0.009)	(0.018)	(0.020)	(0.016)
sch2	0.006***	0.009***	0.007***	0.010***	0.010***	0.005***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
age	0.054***	0.046***	0.098***	0.027***	0.040***	0.085***
	(0.008)	(0.009)	(0.006)	(0.010)	(0.014)	(0.006)
age2/100	-0.066***	-0.053***	-0.093***	-0.027*	-0.044**	-0.077***
	(0.011)	(0.012)	(0.008)	(0.015)	(0.019)	(0.008)
female	-0.751***	-0.555***	-0.331***	-0.494***	-0.502***	-0.207***
	(0.036)	(0.040)	(0.024)	(0.063)	(0.089)	(0.031)
rural	-0.537***	-0.228***	-0.258***	-0.547***	-0.211***	-0.088**
	(0.041)	(0.049)	(0.023)	(0.043)	(0.062)	(0.034)
panel	-0.219***	-0.127***	-0.067***	-0.172***	-0.127**	-0.074***
	(0.039)	(0.048)	(0.024)	(0.039)	(0.051)	(0.023)
weeks	1.165***	1.096***	1.073***	1.068***	0.997***	1.036***
	(0.015)	(0.015)	(0.017)	(0.021)	(0.023)	(0.024)
Copula(sch)				-0.016	-0.027	0.038**
				(0.044)	(0.047)	(0.042)
Copula(sch2)				0.036	0.026	0.101**
				(0.045)	(0.046)	(0.042)
Copula(weeks)				0.073***	0.096***	0.009
				(0.013)	(0.018)	(0.006)
IMR				-0.707***	-0.159	-0.517***
				(0.153)	(0.247)	(0.070)
constant	-0.722***	-0.785***	-2.345***	0.894**	-0.250	-0.796***
	(0.148)	(0.165)	(0.113)	(0.364)	(0.648)	(0.230)
AME(sch)	0.050***	0.073***	0.148***	0.069***	0.077**	0.063**
	(0.006)	(0.007)	(0.003)	(0.020)	(0.021)	(0.015)
Obs.	3,738	1,929	4,830	3,738	1,929	4,830
R^2	0.73	0.79	0.71			

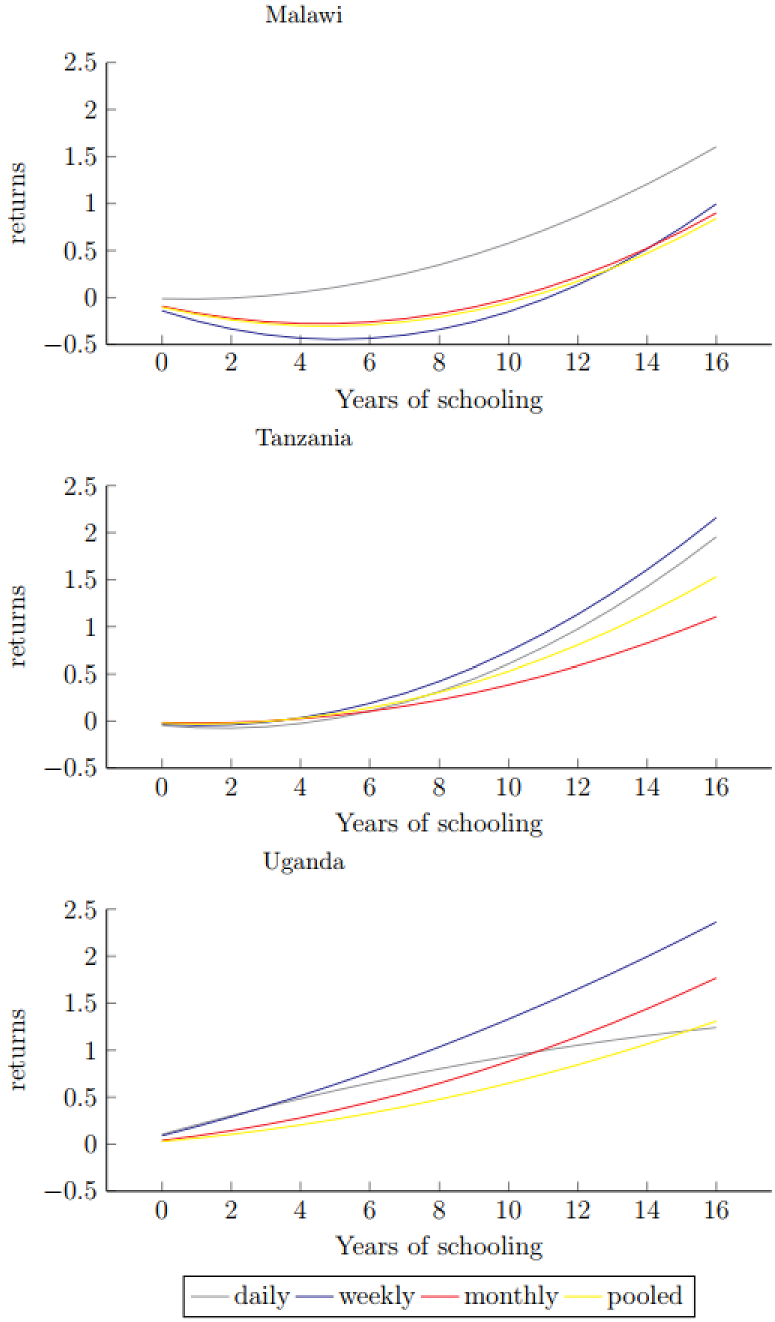
Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models. IMR is the inverse mills ratio. AME(sch) is the average marginal effects of schooling. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.12: Effects of Pay Period on Estimates of Returns to Schooling - Uganda

	OLS			HGC		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
sch	0.120***	0.058*	0.064***	0.105*	0.091	0.040*
	(0.023)	(0.033)	(0.015)	(0.061)	(0.079)	(0.022)
sch2	-0.002	0.003	0.004***	-0.002	0.003	0.004***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
age	0.063***	0.077***	0.080***	0.065***	0.075***	0.082***
	(0.013)	(0.019)	(0.010)	(0.014)	(0.020)	(0.012)
age2/100	-0.082***	-0.090***	-0.084***	-0.088***	-0.084***	-0.086***
	(0.018)	(0.025)	(0.013)	(0.020)	(0.027)	(0.015)
female	-0.596***	-0.535***	-0.312***	-0.693***	-0.415**	-0.328***
	(0.066)	(0.087)	(0.035)	(0.138)	(0.181)	(0.053)
rural	-0.248***	-0.358***	-0.232***	-0.303***	-0.317***	-0.251***
	(0.055)	(0.087)	(0.034)	(0.101)	(0.101)	(0.059)
panel	0.015	-0.063	0.176***	0.013	-0.067	0.171***
	(0.056)	(0.102)	(0.037)	(0.054)	(0.080)	(0.042)
weeks	1.237***	1.093***	1.133***	1.166***	1.153***	1.164***
	(0.039)	(0.054)	(0.031)	(0.067)	(0.090)	(0.051)
Copula(sch)				0.136	-0.072	0.052
				(0.162)	(0.257)	(0.066)
Copula(sch2)				-0.082	-0.079	0.107
				(0.164)	(0.209)	(0.081)
Copula(weeks)				0.024*	-0.020	-0.007
				(0.013)	(0.020)	(0.008)
IMR				0.173	-0.348	0.069
				(0.203)	(0.416)	(0.181)
constant	-2.106***	-1.790***	-2.655***	-2.084***	-1.448	-2.592***
	(0.260)	(0.396)	(0.191)	(0.569)	(1.110)	(0.482)
AME(sch)	0.096***	0.097***	0.145***	0.076	0.137*	0.116***
	(0.008)	(0.011)	(0.004)	(0.011)	(0.074)	(0.026)
Obs.	1,262	589	2,765	1,262	589	2,765
R2	0.57	0.57	0.63			

Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models. IMR is the inverse mills ratio. AME(sch) is the average marginal effects of schooling. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 3.2: Effects of Pay Period on Estimates of Returns to Schooling



Tables 3.13 – 3.15 show the corresponding results for levels of education. Table 3.13 shows the results for Malawi. Here the pattern of the returns is mixed: the returns to primary education are highest if reporting earnings monthly; returns to secondary education are highest if reporting earnings daily; and returns to higher education are highest if reporting earnings weekly. Like for the years of schooling, we do not find evidence of significant selection problems for daily and

weekly, although it might mean that the sample sizes are too small to detect it (and include professionals with relatively high earnings and education). Recall, however, that *ganyu* workers (the majority by far) are excluded so Malawi is not fully comparable to Tanzania and Uganda.

Table 3.14 presents the results for levels of education for Tanzania. As can be seen, generally, we find a pattern of results similar to those in Table 3.11. The returns to the levels of education differ by pay period and weekly have higher returns than their daily and monthly counterparts. Compared to those for daily and weekly, the results for monthly are closer to the results for the pooled sample reported earlier in Table 3.7. This may suggest that the larger monthly sample biases the pooled results into its direction.

Table 3.15 shows results for the levels of education by pay period for Uganda. The results for daily are inconsistent with those obtained when using years of schooling in Table 3.13. While results from Table 3.13 shows a concave relationship between earnings and education, results from Table 3.15 shows a convex relationship, i.e. returns to education increase with the levels of education. A possibility is few observations at more years of education so estimates are imprecise, exacerbated by the (negative) *sch2* effect and perhaps some of those with more education are 'waiting' to get into monthly paid work. Our data do not provide enough information to investigate this issue further, but future research could explore if agricultural wage employment plays a role here.

Table 3.13: Effects of Pay Period on Estimates of Returns to Levels of Education - Malawi

	OLS			HGC		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
primary	0.26	0.135	0.355***	0.315	0.139	0.313***
	(0.233)	(0.083)	(0.024)	(0.298)	(0.089)	(0.030)
secondary	0.684***	0.826***	0.885***	0.795*	0.830***	0.788***
	(0.249)	(0.107)	(0.025)	(0.452)	(0.115)	(0.049)
higher	1.067***	1.643***	1.648***	1.224**	1.651***	1.510***
	(0.299)	(0.139)	(0.029)	(0.612)	(0.158)	(0.067)
age	0.036	0.082***	0.057***	0.049	0.084***	0.050***
	(0.059)	(0.020)	(0.006)	(0.073)	(0.024)	(0.006)
age2/100	-0.032	-0.093***	-0.053***	-0.047	-0.094***	-0.045***
	(0.074)	(0.026)	(0.007)	(0.090)	(0.031)	(0.008)
female	0.006	-0.104	-0.114***	-0.057	-0.112	-0.056*
	(0.210)	(0.077)	(0.021)	(0.301)	(0.109)	(0.033)
rural	-0.307	-0.375***	-0.205***	-0.307	-0.380***	-0.145***
	(0.204)	(0.078)	(0.019)	(0.204)	(0.090)	(0.032)
year	0.866***	1.014***	1.233***	0.938***	1.017***	1.253***
	(0.212)	(0.072)	(0.018)	(0.325)	(0.078)	(0.020)
weeks	1.265***	1.020***	1.136***	1.263***	1.020***	1.132***
	(0.167)	(0.051)	(0.019)	(0.167)	(0.052)	(0.019)
IMR				0.244	0.029	-0.125**
				(0.831)	(0.268)	(0.054)
Constant	-1.708	-1.176***	-1.765***	-2.71	-1.273	-1.456***
	(1.215)	(0.402)	(0.117)	(3.616)	(0.970)	(0.178)
Obs.	182	505	5,129	182	505	5,129
R^2	0.44	0.68	0.77	0.45	0.64	0.76

Standard errors in parentheses * p<0.10, * p<0.05, *** p<0.01.

Table 3.14: Effects of Pay Period on Estimates of Returns to Levels of Education - Tanzania

	OLS			Heckman		
	Daily	Weekly	Monthly	Daily	Weekly	Monthly
primary	0.143***	0.295***	0.354***	0.225***	0.311***	0.226***
	(0.036)	(0.040)	(0.036)	(0.040)	(0.047)	(0.040)
secondary	0.425***	0.809***	1.159***	0.753***	0.864***	0.819***
	(0.077)	(0.085)	(0.039)	(0.102)	(0.122)	(0.064)
higher			1.892***			1.337***
			(0.052)			(0.098)
age	0.053***	0.043***	0.095***	0.027***	0.037***	0.085***
	(0.008)	(0.009)	(0.006)	(0.010)	(0.012)	(0.006)
age2/100	-0.065***	-0.049***	-0.087***	-0.028**	-0.041**	-0.076***
	(0.011)	(0.012)	(0.008)	(0.013)	(0.017)	(0.008)
female	-0.765***	-0.569***	-0.352***	-0.534***	-0.525***	-0.240***
	(0.036)	(0.039)	(0.024)	(0.059)	(0.079)	(0.029)
rural	-0.541***	-0.233***	-0.265***	-0.570***	-0.251***	-0.118***
	(0.041)	(0.049)	(0.023)	(0.042)	(0.057)	(0.032)
panel	-0.220***	-0.137***	-0.086***	-0.179***	-0.136***	-0.087***
	(0.040)	(0.049)	(0.024)	(0.040)	(0.049)	(0.024)
weeks	1.170***	1.099***	1.083***	1.162***	1.099***	1.076***
	(0.015)	(0.015)	(0.017)	(0.015)	(0.015)	(0.017)
IMR				-0.681***	-0.135	-0.443***
				(0.139)	(0.214)	(0.066)
Constant	-0.677***	-0.702***	-2.117***	0.733**	-0.274	-1.208***
	(0.146)	(0.162)	(0.111)	(0.322)	(0.574)	(0.175)
Obs.	3,738	1,929	4,830	3,738	1,929	4,830
R^2	0.73	0.79	0.71	0.73	0.79	0.71

Standard errors in parentheses * p<0.10, * p<0.05, *** p<0.01.

Table 3.15: Effects of Pay Period on Estimates of Returns to Levels of Education - Uganda

	OLS			Heckman		
	Daily	Weekly	Monthly	Daily	Weekly	Monthly
primary	0.466***	0.439***	0.650***	0.462***	0.425***	0.666***
	(0.060)	(0.088)	(0.047)	(0.061)	(0.089)	(0.063)
secondary	0.628***	0.476***	0.866***	0.609***	0.491***	0.888***
	(0.092)	(0.133)	(0.057)	(0.103)	(0.134)	(0.080)
higher	1.009***	1.230***	1.481***	0.964***	1.251***	1.539***
	(0.161)	(0.157)	(0.049)	(0.198)	(0.159)	(0.157)
age	0.067***	0.075***	0.086***	0.068***	0.072***	0.087***
	(0.014)	(0.019)	(0.010)	(0.014)	(0.020)	(0.011)
age2/100	-0.091***	-0.091***	-0.093***	-0.094***	-0.083***	-0.095***
	(0.019)	(0.025)	(0.013)	(0.019)	(0.027)	(0.014)
female	-0.678***	-0.586***	-0.318***	-0.721***	-0.464***	-0.332***
	(0.065)	(0.087)	(0.036)	(0.126)	(0.158)	(0.051)
rural	-0.297***	-0.460***	-0.292***	-0.327***	-0.413***	-0.310***
	(0.056)	(0.088)	(0.035)	(0.093)	(0.102)	(0.058)
panel	0.001	-0.042	0.209***	0.003	-0.046	0.204***
	(0.057)	(0.103)	(0.038)	(0.057)	(0.103)	(0.041)
weeks	1.238***	1.083***	1.140***	1.239***	1.081***	1.140***
	(0.039)	(0.055)	(0.032)	(0.039)	(0.055)	(0.032)
IMR				0.075	-0.326	0.055
				(0.189)	(0.354)	(0.143)
Constant	-1.678***	-1.326***	-2.414***	-1.796***	-0.600	-2.531***
	(0.254)	(0.379)	(0.191)	(0.392)	(0.875)	(0.358)
Obs.	1,262	589	2,765	1,262	589	2,765
R^2	0.55	0.55	0.6	0.55	0.56	0.6

Standard errors in parentheses * p<0.10, * p<0.05, *** p<0.01.

The results for the probit estimates (the first stage results for the selection model) for the probabilities of participating in each of the periods of payment are reported in Appendix 3C. Importantly for Tanzania (Table 3C.2), the coefficient on *sch2* for daily and weekly is negative and statistically significant, implying that an extra year of schooling reduces participation in daily and weekly employment. For monthly, the coefficient is positive and significant, indicating that an extra year of schooling increases participation in this employment category. Thus, a minimum level of education is essential for gaining employment in monthly paid jobs but, conditional on securing such jobs, the marginal effect of schooling on wages is lower than for weekly or daily paid work (which has lower education entry requirements).

Like Tanzania, Table 3C.3 in the Appendix 3C (the first stage regression for HGC) shows that an extra year of schooling reduces participation in daily and weekly employment (as indicated by the negative sign of the coefficient on *sch2*) but increases participation into monthly employment in Uganda. Therefore, in Uganda, a minimum level of education is essential for gaining employment in monthly paid jobs, conditional on securing such jobs, the marginal effect of schooling on wages is lower than for weekly but higher than for daily paid work.

3.5.3 Measuring Returns for Casual Employment: A Case of *Ganyu* in Malawi

This section presents the results for returns to education for *ganyu* labour. In line with the objectives of the study, we explored how using different measures of earnings affects the estimates of returns to education for *ganyu* labour. Table 3.16 shows returns to years of education by the measure of earnings while Figure 3.3 plots the estimates for the selected years of education. Generally, MonthlyC yields larger estimates of returns to education than DailyC or MonthlyA. Furthermore, the difference between estimates from DailyC and MonthlyA is small and the standard errors using these measures of earnings are smaller, suggesting that converting to MonthlyC gives larger and less efficient estimates. There does not seem to be a critical endogeneity of education, as indicated by the insignificance of the copula function for the education variables, but the results should be interpreted with cautious given the close to normal distribution of *ganyu* earnings.

Table 3.17 shows the corresponding results for levels of education. Because there is a very small proportion of workers with higher education doing *ganyu* labour, we will reserve the discussion on returns to higher education. As can be seen, the results in Table 3.17 are consistent with those in Table 3.16 in the sense that the three earnings measures yield different returns to education; and MonthlyC results in higher estimates. In addition, the signs, pattern and significance of the inverse mills ratios are similar to those in Table 3.17, suggesting that the estimates are precise thanks to the large sample.

Figure 3.3: Earnings Measures and Returns to Years of Education for *Ganyu* Labour

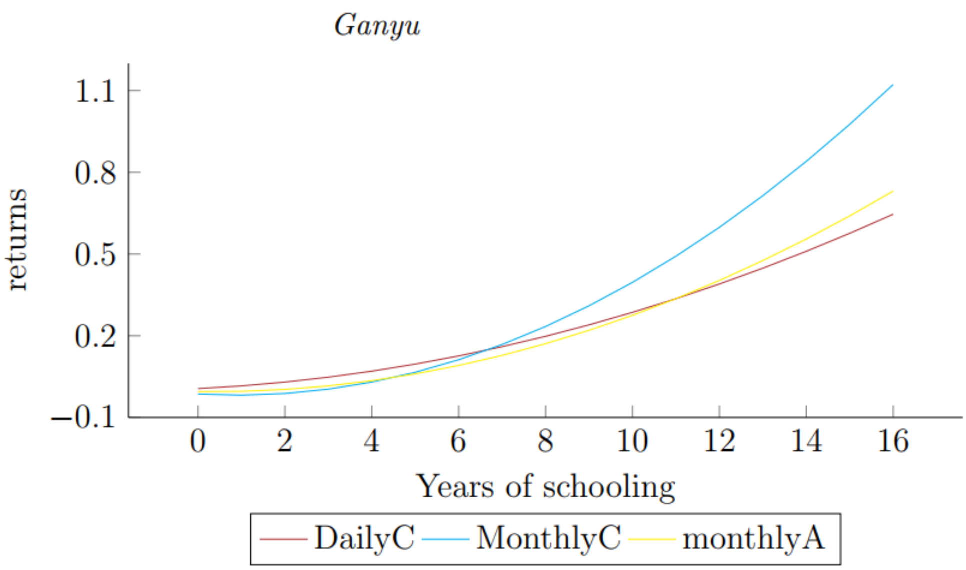


Table 3.16: Earnings Measures and Returns to years of Schooling for *Ganyu* Labour

	OLS			HGC		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	0.004	-0.018***	0.000	0.006	-0.014	-0.005
	(0.004)	(0.006)	(0.005)	(0.008)	(0.011)	(0.010)
sch2	0.002***	0.003***	0.002***	0.002***	0.005***	0.003***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
age	0.034***	0.047***	0.040***	0.032***	0.039***	0.036***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.004)	(0.003)
age2/100	-0.040***	-0.057***	-0.048***	-0.037***	-0.043***	-0.040***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.005)	(0.004)
female	-0.240***	-0.444***	-0.341***	-0.223***	-0.367***	-0.299***
	(0.010)	(0.014)	(0.012)	(0.012)	(0.017)	(0.014)
rural	-0.284***	-0.420***	-0.331***	-0.330***	-0.624***	-0.440***
	(0.018)	(0.025)	(0.021)	(0.029)	(0.039)	(0.034)
year	1.326***	1.223***	1.218***	1.298***	1.097***	1.150***
	(0.010)	(0.013)	(0.012)	(0.015)	(0.022)	(0.018)
weeks			1.006***			0.981***
			(0.006)			(0.011)
Copula(sch)				-0.011	0.008	0.012
				(0.020)	(0.026)	(0.023)
Copula(sch2)				-0.003	-0.041	-0.003
				(0.019)	(0.027)	(0.023)
Copula(weeks)						0.019**
						(0.008)
IMR				-0.096**	-0.430***	-0.237***
				(0.042)	(0.059)	(0.050)
Constant	0.192***	2.642***	-0.912***	0.317***	3.237***	-0.495***
	(0.047)	(0.065)	(0.056)	(0.091)	(0.130)	(0.111)
AME(sch)	0.02	0.009***	0.018***	0.028***	0.035***	0.025***
	(0.001)	(0.002)	(0.002)	(0.007)	(0.010)	(0.009)
Obs.	16,528	16,528	16,528	16,528	16,528	16,528
R^2	0.60	0.38	0.77			

Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models. IMR is the inverse mills ratio. AME(sch) is the average marginal effects of schooling. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.17: Earnings Measures and Returns to Levels of Education for *Ganyu* Labour

	OLS			Heckman		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
primary	0.122***	0.095***	0.122***	0.149***	0.207***	0.187***
	(0.014)	(0.019)	(0.016)	(0.017)	(0.023)	(0.020)
secondary	0.148***	0.109**	0.133***	0.219***	0.402***	0.303***
	(0.031)	(0.043)	(0.036)	(0.041)	(0.057)	(0.048)
higher	1.150***	1.282***	1.292***	1.282***	1.834***	1.611***
	(0.107)	(0.148)	(0.124)	(0.118)	(0.163)	(0.138)
age	0.034***	0.047***	0.040***	0.032***	0.039***	0.036***
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
age2	-0.041***	-0.057***	-0.049***	-0.037***	-0.044***	-0.041***
	(0.003)	(0.004)	(0.004)	(0.003)	(0.005)	(0.004)
female	-0.256***	-0.447***	-0.355***	-0.238***	-0.373***	-0.313***
	(0.010)	(0.013)	(0.011)	(0.012)	(0.016)	(0.014)
rural	-0.289***	-0.415***	-0.333***	-0.341***	-0.629***	-0.458***
	(0.018)	(0.025)	(0.021)	(0.027)	(0.037)	(0.031)
year	1.336***	1.223***	1.227***	1.306***	1.097***	1.155***
	(0.010)	(0.013)	(0.012)	(0.015)	(0.021)	(0.018)
weeks			1.004***			1.001***
			(0.006)			(0.006)
IMR				-0.106***	-0.443***	-0.257***
				(0.041)	(0.056)	(0.047)
constant	0.277***	2.629***	-0.846***	0.448***	3.339***	-0.429***
	(0.045)	(0.062)	(0.054)	(0.079)	(0.109)	(0.094)
Obs.	16,528	16,528	16,528	16,528	16,528	16,528
R^2	0.56	0.39	0.77			

Note: IMR is the inverse mills ratio. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

3.6 Conclusion

In this essay, we estimated returns to schooling in Malawi, Tanzania and Uganda using nationally representative and comparable data from the Living Standard Measurement Study. Of interest was whether the relationship between earnings and education varies across workers reporting earnings over different pay periods, and our results suggest that this is the case. This is the first study on Africa to examine this issue. After controlling for endogeneity due to unobserved ability and selection, by using Gaussian Copula to account for endogeneity and Heckman method to account for selection, we showed that returns to education differ by pay period and that pooling the periods together leads to imprecise estimates. Returns to education do vary according to the period of payment and how the vary differs across the three countries. Specifically, in Malawi the returns for non-*ganyu* workers reporting earnings daily are the highest, followed by monthly and then weekly; in Tanzania, the returns for weekly are not only higher but also increase at a higher rate than for the other periods; and in Uganda, returns are highest for weekly followed by monthly and then daily. Our results also show that pooling/aggregating earnings to different common pay period measures yield different estimates of returns to education and that estimates are generally leaning toward the direction of the pay period that constitutes the largest proportion of the sample. In this regard, our analysis suggests that estimating returns separately for workers paid over different periods is more reliable than pooling.

The findings regarding the three common measures of earning yielding different estimates of returns to education are sound. Given the seasonality of casual work, earnings measures that allow for workers who do not work all weeks in the month and for seasonal workers who only work some months in a year are more reliable than measures that do not. Another explanation for the observed pattern of results is that while workers who are paid monthly are more likely those in formal jobs, their daily and weekly counterparts are more likely to be in informal jobs. In a developing country context, a minimum level of education is essential for gaining employment in the formal sector, but conditional on securing such jobs the marginal effect of schooling on wages may be lower than for the informal sector.

It is worth pointing out that since in Malawi *ganyu* labour is treated separately, regular employment is likely to constitute only those from the formal sector who are the better educated and hence better paid. Many of the workers reporting earnings hourly, daily and weekly in Tanzania and Uganda may have been in *ganyu* labour had they been residing in Malawi, and vice versa. Given this characteristic of Malawi's labour market, comparing the results with those for Tanzania and Uganda need to be done with caution. This clearly deserves a further and independent investigation but is beyond the scope of this essay.

Appendices

Appendix 3A: Selected Studies on Returns to Schooling in Developing Countries

Table 3A.1: Selected Studies on Returns to Schooling in Developing Countries

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Nikolov and Jimi (2018)	Log monthly earnings	No	Both linear and non-linear schooling using, three dummies for education levels: primary, O-level and A-level secondary (reference category 'no schooling')	ILFS 2014 and Dar es Salaam Perceived Returns Survey (DPRS) 2014. ILFS 16,817 observations and DPRS 1,211	OLS for non-linear and IV(2SLS) for linear schooling	Convex returns, national returns are 12% and 7% while Dar es Salaam returns are 11% and 9% respectively for OLS and IV. Estimates are insignificant when sample is split into the three levels of education.
Bridges et al. (2017)	Log monthly earnings	No	Non-linear schooling: dummies for completed levels of education: primary, O-level secondary, A-level secondary, vocational/technical.	All three rounds of Tanzania Household Urban Panel Survey (THUPS), a subsample of youth aged 20 to 35 inclusive. A sample of 365 individuals	Fixed Effects	All dummies for education levels are insignificant after controlling for family fixed effects.

Table 3A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Serneels et al. (2017)	Log daily wages	Yes	Non-linear schooling: primary vs post primary school levels	Survey of Household Welfare and Labor in Tanzania (SHWALITA), 291 sample of wage workers	IV (Control function)	Returns differ by survey instrument but not by type of respondent. Short module questionnaires lead to biased estimates compared to detailed questionnaires. After controlling for endogeneity and selection using Heckman method, returns are about 20% and 49% for a year of post primary school respectively for men and women if short modules are used. Using Heckman-Hotz method, the returns are respectively 21% and 32%. While generally schooling is insignificant for men when detailed modules are used, post primary returns are 50% and 29% for women using Heckman and Heckman-Hotz method respectively.

Table 3A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Peet et al. (2015)	Log annual earnings	No	Linear schooling	LSMS 2004, 2008, 2010. Sample 985; 1,807 and 2,716 respectively	OLS	The returns are convex. Returns are 12.1% 9% and 12.2% for the survey years respectively with the period average of 11.1%. Returns are higher for female and urban employees
Barouni and Broecke (2014)		No	Non-linear schooling: dummies for different completed levels of education		OLS	Returns are 5%, 100% and 51% for basic, Alevel and tertiary education respectively
Kahyarara and Teal (2008)	Log monthly earnings	No	Non-linear schooling: dummies for completed levels of education: primary, O-level secondary, A-level secondary, vocational, technical, professional and university	Fourth and fifth rounds of the Tanzanian Manufacturing Enterprise surveys. Total sample of 2527 employees	IV (control function) with firm fixed effects: parental education and main occupation as instruments	Returns are convex: higher levels of education(academic) have higher returns. Returns to vocation and technical education depend on the level of education(academic) with which one enters vocational/technical college. The higher the entry level the lower the returns.

Table 3A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Al-Samarrai and Reilly (2008)	Log monthly earnings	No	Non-linear post primary school levels	Tracer survey 2001. A sample of 965 respondents		The returns are convex. The rate of returns for a year of Alevel of education for the wage employees is 8.8% while the rates for a year of university education is 17.1%. No significant effect of these levels of education on the Self-employed.
Soderbom et al. (2006)	Log monthly earnings	No	Linear schooling	Surveys of employees in the manufacturing sector 1993, 1994, 1999 and 2001. Total sample of 2,738 workers	IV (control function): parental education, main occupation, distance to primary school at age 6 and to secondary at age 12 as instruments	The returns are convex. There has been an increase in returns from early 1990s to 2000. The earning profiles for young and old people are significantly different. After controlling for endogeneity, youth returns are 10.6% and is insignificant for the old

Table 3A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Agrawal and Agrawal (2018)	Log hourly earnings	No	Both linear and non-linear schooling	India Human Development Survey (IHDS), 2011–2012	OLS corrected for employment selection bias	Returns are higher for females (5.7% compared to 5% for males). Wage employees have highest returns followed by self-employed and agriculture. Returns are convex ranging from 2.2% for primary education to 18.9% for university education.
Chuang and Lai (2017)	Log hourly earnings	No	Linear schooling	Taiwan's 1978-2003 Manpower Utilization Survey	Quantile regression	Returns increased from 5.5% in 1978 to 8.2 % in 2003 with an average of 6.5% The returns are higher for lower quantiles and vice versa.
Salisbury (2016)	Log monthly earnings	No	Both linear and non-linear schooling	National Income Dynamics Study 2008 (south Africa)	OLS	Returns are 18.7%, lowest for Africans (16%) and highest for Asians/Indians (25%). The returns are also higher for females. When allowing for non-linearities in schooling, returns are

						convex: 7%, 13% and 29% respectively for primary, secondary and tertiary education.
Kuepié and Nordman (2016)	Log hourly earnings	No	Non-linear schooling: dummies for different completed levels of education	Employment and Informal Sector Survey (EESIC) 2009 (Republic of Congo)	IV (control function): father's education and job professional status as instruments	Convex returns. Primary education no effect on earning, returns for lower secondary, upper secondary and higher education are respectively 9%, 5% and 12% for Brazzaville and 9%, 14% and 13% for Pointe-Noire.
Peet et al. (2015)	Log annual earnings	No	Linear schooling	LSMS data from 25 developing countries of which 9 countries from Africa: Cote d'Ivoire, Ethiopia, Ghana, Malawi, Niger, Nigeria, SA, Tanzania and Uganda	OLS	Returns differ significantly by countries and within countries by survey years. But generally, they range from 3.2% to 12.5%. The pattern of returns across the levels of education also differs by countries and by survey years. Returns are generally higher for women though the difference is small

Table 3A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Barouni and Broecke (2014)	Not specified	No	Non-linear schooling: dummies for completed levels of education	Post 2005 household and labour force surveys in Burundi, Egypt, Ghana, Mali, Nigeria, Rwanda, Sudan, South Africa, Tanzania, Togo, Tunisia, and Uganda.	OLS	The average Mincer returns for the 12 countries are 7%, 26% and 26% for basic, upper secondary and tertiary education respectively. Returns are higher for women except for tertiary education where they are equal. The pattern of returns across the levels of education differs by countries.
Wang (2013)	Log annual wages	No	Linear schooling	urban sample of the China Household Income Project (CHIP) 1995 and 2002	IV (2SLS): parental education vs spouse education	The returns increased over the two survey periods regardless of the instrument used. Returns are higher using parental education as IV relative to spouse education, but the difference is not statistically significant.

Table 3A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Aslam et al. (2012)	Log daily/regular wages	No	Linear and quadratic schooling	Purposive household survey in Punjab and the North West Frontier Province (NWFP) Pakistan 2006 -2007	OLS with ability proxy, IV (2SLS) and Fixed effects	Males have returns of 10% using IV method, schooling not significant for females. Using Fixed effects, returns are 5% similar to OLS
Stefani and Biderman (2009)	Log hourly wage	No	Linear schooling	Brazil National Household Survey 1988 and 1996	IV Quantile regression: parental education and family size as instruments	Returns are heterogeneous across colour gender and earning distribution, ranging from 6% to 32%
Pietro (2008)	Log hourly wage	No	Linear schooling	The Argentine Permanent Household Survey 1995 - 2003	OLS with selection correction and IV (2SLS): spouse education	Decrease in returns between 1996 and 1999 and increase in returns 1999 to 2002. Returns from OLS corrected for selection average at 8.5% while IV estimates are averaged at 11.5%
Soderbom et al. (2005)	Log monthly earnings	No	Linear schooling	Surveys of manufacturing firms in Ghana and Kenya.	OLS	Returns are 8.3% in Ghana and 10.4% in Kenya

Table 3A.1 (*continued*)

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
Lassibille and Tan (2005)	Log hourly wage	No	Both linear and non-linear schooling	Household Living Conditions Survey 1999-2001(Rwanda)	OLS corrected for employment selection bias	Returns are 17.5% and convex: primary - 19%, secondary 29% and tertiary 33%. Generally public sector has higher returns compared to private sector.
Girma and Kedir (2005)	Log hourly wage	No	Linear schooling	Household panel data for Ethiopian seven major cities 1994, 1995, and 1997	IV Quantile regression: parental education as instrument	Returns are 14%. The returns differ across the earning distribution: highest at 25th (20%) quantile and lowest at 90th quantile (%). Lower returns for public sector (12%) relative to private sector (16%).
Schultz (2004)	Log hourly wage	No	Non-linear schooling: dummies for completed levels of education	Various national representative household surveys from 6 African Countries: Burkina Faso, Cote d'Ivoire, Ghana, Kenya, Nigeria and SA in the period 1985 - 1999	OLS	Returns differ significantly by countries and by levels of education. Generally, an extra year is associated with 5 to 20% increase in earnings. Primary school returns range between 3 - 10% while

Study	Dependent Variable	Pay Period Considered?	Measure/Specification of schooling	Data and Sample	Estimator(s)	Main Results
						tertiary education returns range between 10 - 15%.
Moock et al. (2003)	Log monthly earnings	No	Both linear and non-linear schooling	Vietnam Living Standards Survey (VLSS) 1992-1993	OLS	Using linear schooling returns are 8% while using education dummies returns are highest at primary school(13%), followed by university (11%). Returns for secondary and vocational education are respectively 5% and 4%.

Appendix 3B: GC Estimates of Returns to Schooling by Period

Table 3B.1: GC Estimates of Returns to Schooling by Period - Malawi

Period	Daily	Weekly	Monthly	Pooled
sch	-0.015	-0.140**	-0.091***	-0.101***
	(0.086)	(0.057)	(0.011)	(0.011)
sch2	0.005	0.012***	0.010***	0.009***
	(0.007)	(0.002)	(0.001)	(0.001)
age	0.039	0.078***	0.054***	0.055***
	(0.058)	(0.026)	(0.006)	(0.006)
age200	-0.034	-0.081**	-0.049***	-0.051***
	(0.074)	(0.032)	(0.007)	(0.008)
female	-0.031	-0.118*	-0.119***	-0.106***
	(0.188)	(0.071)	(0.022)	(0.021)
rural	-0.243	-0.305***	-0.188***	-0.211***
	(0.222)	(0.081)	(0.019)	(0.019)
year	0.851***	0.990***	1.277***	1.234***
	(0.214)	(0.078)	(0.017)	(0.019)
weeks	1.366***	1.231***	1.146***	0.168***
	(0.207)	(0.076)	(0.029)	(0.050)
copula(sch)	0.644	0.189	0.129***	0.168***
	(0.400)	(0.261)	(0.046)	(0.049)
copula(sch2)	-0.533	0.074	0.120**	1.151***
	(0.511)	(0.170)	(0.048)	(0.029)
copula(weeks)	-0.038	-0.063***	-0.001	-0.006
	(0.043)	(0.018)	(0.004)	(0.005)
constant	-2.052	-1.360*	-1.401***	-1.199***
	(1.427)	(0.707)	(0.171)	(0.166)
AME(sch)	0.075	0.050	0.096***	0.066***
	(0.113)	(0.643)	(0.015)	(0.014)
Obs.	182	505	5,129	5,816
R^2	0.45	0.67	0.77	0.74

Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3B.2: GC Estimates of Returns to Schooling by Period - Tanzania

Period	Daily	Weekly	Monthly	Pooled
sch	-0.022	-0.028	-0.006	-0.033***
	(0.017)	(0.019)	(0.016)	(0.010)
sch2	0.005***	0.009***	0.006***	0.009***
	(0.002)	(0.002)	0.000	0.000
age	0.053***	0.046***	0.098***	0.066***
	(0.008)	(0.009)	(0.006)	(0.004)
age2/100	-0.065***	-0.053***	-0.093***	-0.070***
	(0.011)	(0.012)	(0.008)	(0.006)
female	-0.738***	-0.553***	-0.329***	-0.549***
	(0.036)	(0.036)	(0.024)	(0.019)
rural	-0.521***	-0.191***	-0.256***	-0.330***
	(0.041)	(0.050)	(0.023)	(0.018)
panel	-0.216***	-0.129***	-0.069***	-0.087***
	(0.039)	(0.049)	(0.022)	(0.018)
weeks	1.075***	0.998***	1.044***	0.039
	(0.023)	(0.024)	(0.026)	(0.028)
copula(sch)	-0.018	-0.027	0.058	0.038
	(0.046)	(0.045)	(0.042)	(0.028)
copula(sch2)	0.037	0.026	0.112***	1.078***
	(0.045)	(0.046)	(0.042)	(0.011)
copula(weeks)	0.074***	0.096***	0.009*	0.029***
	(0.013)	(0.018)	(0.006)	(0.005)
constant	-0.542***	-0.652***	-1.886***	-1.105***
	(0.174)	(0.170)	(0.184)	(0.100)
AME(sch)	0.038*	0.071***	0.103***	0.083***
	(0.019)	(0.020)	(0.014)	(0.009)
Obs.	3,738	1,929	4,830	11,215
R^2	0.73	0.80	0.72	0.78

Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3B.3: GC Estimates of Returns to Schooling by Period - Uganda

Period	Daily	Weekly	Monthly	Pooled
sch	0.105*	0.091	0.040*	0.030
	(0.059)	(0.078)	(0.022)	(0.020)
sch2	-0.002	0.003	0.003***	0.003***
	(0.001)	(0.002)	(0.001)	(0.001)
age	0.063***	0.079***	0.080***	0.068***
	(0.014)	(0.020)	(0.010)	(0.007)
age2/100	-0.082***	-0.092***	-0.083***	-0.077***
	(0.019)	(0.026)	(0.013)	(0.009)
female	-0.592***	-0.544***	-0.313***	-0.435***
	(0.063)	(0.084)	(0.033)	(0.028)
rural	-0.233***	-0.369***	-0.231***	-0.239***
	(0.056)	(0.083)	(0.036)	(0.029)
panel	0.009	-0.062	0.177***	0.155***
	(0.055)	(0.082)	(0.038)	(0.030)
weeks	1.167***	1.158***	1.164***	0.144*
	(0.069)	(0.090)	(0.048)	(0.074)
Copula(sch)	0.137	-0.056	0.055	0.042
	(0.161)	(0.269)	(0.065)	(0.077)
Copula(sch2)	-0.085	-0.076	0.108	1.172***
	(0.169)	(0.200)	(0.081)	(0.038)
Copula(weeks)	0.023*	-0.020	-0.008	-0.005
	(0.014)	(0.020)	(0.007)	(0.006)
constant	-1.830***	-2.204***	-2.426***	-1.933***
	(0.503)	(0.686)	(0.258)	(0.214)
AME(sch)	0.082	0.131*	0.108***	0.081***
	(0.057)	(0.074)	(0.021)	(0.018)
Obs.	1,262	589	2,765	4,631
R^2	0.57	0.57	0.63	0.60

Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3B.4: GC Estimates of Returns to Schooling - *Ganyu*

	DailyC	MonthlyC	MonthlyA
sch	0.008	-0.007	-0.002
	(0.009)	(0.011)	(0.010)
sch2	0.002***	0.003***	0.002***
	0.000	(0.001)	0.000
age	0.034***	0.047***	0.040***
	(0.002)	(0.003)	(0.003)
age2/100	-0.040***	-0.057***	-0.048***
	(0.003)	(0.004)	(0.004)
female	-0.240***	-0.444***	-0.341***
	(0.010)	(0.013)	(0.012)
rural	-0.284***	-0.420***	-0.326***
	(0.020)	(0.028)	(0.023)
year	1.326***	1.223***	1.217***
	(0.010)	(0.014)	(0.012)
weeks			0.982***
			(0.011)
Copula(sch)	-0.012	0.004	0.010
	(0.020)	(0.027)	(0.023)
Copula(sch2)	-0.003	-0.043	-0.004
	(0.020)	(0.027)	(0.023)
Copula(weeks)			0.020**
			(0.008)
Constant	0.172***	2.591***	-0.855***
	(0.061)	(0.082)	(0.078)
AME(sch)	0.024***	0.020*	0.016*
	(0.007)	(0.010)	(0.009)
Obs.	16,528	16,528	16,528
R^2	0.56	0.38	0.77

Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models AME(sch) is the average marginal effects of schooling. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 3C: Determinants of Selection to Employment

Table 3C.1: Determinants of Selection to Employment - Malawi

Period	Daily	Weekly	Monthly	Pooled	*Ganyu*
sch	-0.025	0.031	-0.096***	-0.105***	0.020*
	(0.043)	(0.030)	(0.016)	(0.015)	(0.011)
sch2	0.004***	0.000	0.009***	0.010***	-0.007***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.000)
age	0.042**	0.041***	0.070***	0.072***	0.016***
	(0.017)	(0.011)	(0.006)	(0.005)	(0.004)
age2/100	-0.050**	-0.055***	-0.083***	-0.087***	-0.037***
	(0.021)	(0.014)	(0.007)	(0.007)	(0.005)
female	-0.142*	-0.169***	-0.394***	-0.388***	-0.149***
	(0.073)	(0.046)	(0.024)	(0.023)	(0.018)
rural	0.027	-0.191***	-0.677***	-0.659***	0.627***
	(0.075)	(0.044)	(0.021)	(0.020)	(0.019)
year	0.347***	0.101***	-0.216***	-0.152***	0.444***
	(0.059)	(0.035)	(0.019)	(0.018)	(0.012)
kids5	0.113	-0.203	-0.552***	-0.527***	0.423***
	(0.209)	(0.136)	(0.062)	(0.061)	(0.045)
kids14	-0.024	-0.016	-0.408***	-0.367***	0.128***
	(0.177)	(0.100)	(0.049)	(0.046)	(0.037)
married	-0.075	-0.069	-0.040*	-0.049**	-0.190***
	(0.074)	(0.047)	(0.023)	(0.022)	(0.018)
head	0.343***	0.393***	0.573***	0.613***	0.323***
	(0.076)	(0.052)	(0.027)	(0.025)	(0.018)
Copula(sch)	0.183	-0.110	0.137***	0.138***	-0.045
	(0.153)	(0.086)	(0.048)	(0.046)	(0.030)
Copula(sch2)	-0.105	0.098	0.144***	0.158***	-0.013
	(0.132)	(0.093)	(0.048)	(0.046)	(0.030)
Constant	-3.964***	-3.171***	-1.942***	-1.922***	-0.982***
	(0.421)	(0.278)	(0.140)	(0.136)	(0.084)
Obs.	45,494	45,494	45,494	45,494	45,494

Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models. Standard errors in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3C.2: Determinants of Selection to Employment - Tanzania

Period	Daily	Weekly	Monthly	Pooled
sch	0.036***	0.045***	-0.024**	-0.037***
	(0.009)	(0.011)	(0.010)	(0.008)
sch2	-0.009***	-0.007***	0.007***	0.005***
	(0.001)	(0.001)	(0.001)	(0.000)
age	0.039***	0.039***	0.025***	0.050***
	(0.005)	(0.006)	(0.005)	(0.004)
age2/100	-0.062***	-0.057***	-0.035***	-0.075***
	(0.006)	(0.007)	(0.006)	(0.005)
female	-0.349***	-0.306***	-0.179***	-0.388***
	(0.021)	(0.027)	(0.021)	(0.017)
rural	0.046**	0.150***	-0.410***	-0.171***
	(0.021)	(0.028)	(0.018)	(0.015)
kids5	-0.191***	-0.035	-0.022	-0.064
	(0.020)	(0.025)	(0.020)	(0.049)
kids14	0.182***	0.203***	-0.419***	-0.173***
	(0.066)	(0.078)	(0.065)	(0.043)
panel	0.055	0.079	-0.419***	-0.116***
	(0.057)	(0.069)	(0.053)	(0.015)
married	-0.130***	-0.103***	-0.173***	-0.168***
	(0.022)	(0.027)	(0.021)	(0.017)
head	0.260***	0.217***	0.482***	0.510***
	(0.025)	(0.032)	(0.026)	(0.019)
Copula(sch)	0.030	-0.022	0.035	-0.001
	(0.026)	(0.033)	(0.029)	(0.020)
Copula(sch2)	0.013	-0.054	0.098***	-0.010
	(0.027)	(0.033)	(0.029)	(0.021)
Constant	-1.428***	-2.217***	-1.412***	-0.966***
	(0.097)	(0.121)	(0.100)	(0.074)
Obs.	38,857	38,857	38,857	38,857

Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models. Standard errors in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3C.3: Determinants of Selection to Employment - Uganda

Period	Daily	Weekly	Monthly	Pooled
sch	0.037	0.024	0.037	-0.010
	(0.037)	(0.045)	(0.024)	(0.023)
sch2	-0.004***	-0.001	0.007***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
age	0.031***	0.021**	0.067***	0.058***
	(0.008)	(0.010)	(0.006)	(0.005)
age2/100	-0.058***	-0.038***	-0.085***	-0.083***
	(0.010)	(0.013)	(0.008)	(0.007)
female	-0.667***	-0.390***	-0.306***	-0.554***
	(0.032)	(0.042)	(0.027)	(0.022)
rural	-0.470***	-0.160***	-0.330***	-0.448***
	(0.033)	(0.039)	(0.026)	(0.023)
kids5	0.024	-0.024	-0.276***	-0.492***
	(0.043)	(0.055)	(0.033)	(0.138)
kids14	-0.236	-0.478*	-0.445**	-0.385***
	(0.201)	(0.279)	(0.185)	(0.046)
panel	-0.413***	-0.208**	-0.254***	-0.160***
	(0.074)	(0.088)	(0.052)	(0.030)
married	-0.299***	-0.192***	-0.208***	-0.297***
	(0.033)	(0.041)	(0.027)	(0.023)
head	0.182***	0.220***	0.090***	0.176***
	(0.039)	(0.048)	(0.030)	(0.026)
Pstar(sch)	-0.118	-0.057	-0.202***	-0.152***
	(0.084)	(0.101)	(0.059)	(0.052)
Pstar(sch2)	0.001	0.018	0.020	0.049
	(0.102)	(0.102)	(0.067)	(0.062)
Constant	-1.280***	-1.945***	-2.457***	-1.280***
	(0.224)	(0.297)	(0.173)	(0.153)
Obs.	29,188	29,188	29,188	29,188

Notes: Copula() are Gaussian Copula functions. Significance of copula functions for a variable implies a significant correlation between variable and the errors in the regression models. Standard errors in parentheses.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 3D: Bootstrap Aggregating (Bagging) Results

Table 3D.1: Bootstrap Aggregation Results (Converted Earnings) -Malawi

	GC			HGC		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	-0.110***	-0.105***	-0.095***	-0.111***	-0.106***	-0.096***
	(0.005)	(0.010)	(0.010)	(0.005)	(0.005)	(0.005)
sch2	0.009***	0.009***	0.009***	0.008***	0.008***	0.009***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
age	0.064***	0.060***	0.056***	0.053***	0.048***	0.047***
	(0.000)	(0.006)	(0.006)	(0.001)	(0.001)	(0.001)
age2/100	-0.064***	-0.057***	-0.052***	-0.052***	-0.044***	-0.042***
	(0.000)	(0.007)	(0.007)	(0.001)	(0.001)	(0.001)
female	-0.135***	-0.136***	-0.106***	-0.055***	-0.049***	-0.040***
	(0.001)	(0.022)	(0.021)	(0.004)	(0.004)	(0.003)
rural	-0.137***	-0.210***	-0.210***	-0.057***	-0.122***	-0.144***
	(0.001)	(0.020)	(0.019)	(0.004)	(0.003)	(0.003)
year	1.333***	1.239***	1.235***	1.352***	1.259***	1.250***
	(0.001)	(0.019)	(0.018)	(0.001)	(0.001)	(0.001)
weeks			1.149***			1.144***
			(0.026)			(0.001)
Copula(sch)	0.158***	0.149***	0.137***	0.164***	0.156***	0.142***
	(0.046)	(0.047)	(0.046)	(0.045)	(0.043)	(0.042)
Copula(sch2)	0.158***	0.149***	0.138***	0.164***	0.156***	0.142***
	(0.045)	(0.046)	(0.046)	(0.047)	(0.043)	(0.043)
Copula(weeks)			-0.006			-0.005***
			(0.004)			(0.000)
IMR				-0.173***	-0.191***	-0.144***
				(0.008)	(0.007)	(0.007)
Constant	0.134	3.141***	-1.320***	0.607***	3.663***	-0.914***
	(0.082)	(0.152)	(0.166)	(0.096)	(0.089)	(0.086)
Obs.	5,816	5,816	5,816	5,816	5,816	5,816
R^2	0.59	0.62	0.74			

Notes: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Copula() are Gaussian Copula functions. Significant copula functions imply significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3D.2: Bootstrap Aggregation Results (Converted Earnings) -Tanzania

	GC			HGC		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	-0.046***	-0.038***	-0.043***	-0.040***	-0.017**	-0.032***
	(0.006)	(0.008)	(0.007)	(0.007)	(0.008)	(0.007)
sch2	0.008***	0.010***	0.008***	0.008***	0.007***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
age	0.075***	0.065***	0.066***	0.064***	0.027***	0.046***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
age2/100	-0.079***	-0.067***	-0.070***	-0.065***	-0.017***	-0.043***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
female	-0.445***	-0.642***	-0.549***	-0.349***	-0.303***	-0.370***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.003)	(0.002)
rural	-0.170***	-0.662***	-0.330***	-0.135***	-0.539***	-0.271***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
panel	-0.104***	-0.002**	-0.087***	-0.095***	0.030***	-0.068***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
weeks			1.078***			1.069***
			(0.000)			(0.000)
Copula(sch)	0.039	0.097***	0.061**	0.039	0.093***	0.060**
	(0.025)	(0.032)	(0.027)	(0.025)	(0.032)	(0.027)
Copula(sch2)	0.039	0.097***	0.061**	0.038	0.094***	0.060**
	(0.025)	(0.031)	(0.027)	(0.025)	(0.030)	(0.027)
Copula(weeks)			0.029***			0.030***
			(0.000)			(0.000)
IMR				-0.237***	-0.835***	-0.446***
				(0.003)	(0.007)	(0.005)
Constant	-0.054	2.902***	-1.029***	0.327***	4.238***	-0.292***
	(0.048)	(0.060)	(0.052)	(0.048)	(0.060)	(0.051)
Obs.	11,215	11,215	11,215	11,215	11,215	11,215
R2	0.27	0.37	0.78			

Notes: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Copula() are Gaussian Copula functions. Significant copula functions imply significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3D.3: Bootstrap Aggregation Results (Converted Earnings) -Uganda

	GC			HGC		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	0.008	0.027**	0.022*	0.008	0.027**	0.021*
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
sch2	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
age	0.081***	0.077***	0.068***	0.082***	0.076***	0.069***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
age2/100	-0.091***	-0.088***	-0.077***	-0.092***	-0.085***	-0.079***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
female	-0.439***	-0.448***	-0.436***	-0.454***	-0.416***	-0.454***
	(0.001)	(0.001)	(0.001)	(0.008)	(0.009)	(0.008)
rural	-0.220***	-0.289***	-0.239***	-0.234***	-0.262***	-0.255***
	(0.001)	(0.001)	(0.001)	(0.007)	(0.008)	(0.007)
panel	0.147***	0.214***	0.156***	0.145***	0.216***	0.154***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
weeks			1.171***			1.171***
			(0.001)			(0.001)
Copula(sch)	0.125**	0.112*	0.111*	0.122**	0.113*	0.109*
	(0.061)	(0.064)	(0.062)	(0.061)	(0.065)	(0.062)
Copula(sch2)	0.124**	0.111*	0.111*	0.124**	0.115*	0.110*
	(0.062)	(0.065)	(0.063)	(0.061)	(0.065)	(0.062)
Copula(weeks)			-0.005***			-0.005***
			(0.000)			(0.000)
IMR				0.036*	-0.074***	0.044**
				(0.019)	(0.020)	(0.019)
Constant	-0.579***	2.318***	-1.864***	-0.637***	2.436***	-1.933***
	(0.098)	(0.104)	(0.100)	(0.112)	(0.120)	(0.116)
Obs.	4,631	4,631	4,631	4,631	4,631	4,631
R^2	0.36	0.39	0.60			

Notes: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Copula() are Gaussian Copula functions. Significant copula functions imply significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3D.4: Bootstrap Aggregation Results (Pay Periods) -Malawi

	GC			HGC		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
sch	-0.0002	-0.144***	-0.093***	0.002	-0.143***	-0.094***
	(0.087)	(0.047)	(0.010)	(0.037)	(0.034)	(0.004)
sch2	0.006	0.012***	0.010***	0.006	0.012***	0.009***
	(0.006)	(0.002)	(0.001)	(0.004)	(0.001)	(0.000)
age	0.039	0.079***	0.054***	0.036***	0.079***	0.044***
	(0.059)	(0.020)	(0.006)	(0.013)	(0.003)	(0.001)
age2	-0.033	-0.082***	-0.049***	-0.029*	-0.083***	-0.038***
	(0.073)	(0.026)	(0.007)	(0.016)	(0.003)	(0.001)
female	-0.006	-0.118	-0.117***	0.011	-0.122***	-0.041***
	(0.212)	(0.078)	(0.021)	(0.068)	(0.014)	(0.004)
rural	-0.288	-0.304***	-0.187***	-0.288***	-0.306***	-0.106***
	(0.202)	(0.078)	(0.019)	(0.020)	(0.009)	(0.004)
year	0.837***	0.990***	1.278***	0.818***	0.991***	1.303***
	(0.211)	(0.072)	(0.018)	(0.076)	(0.008)	(0.002)
weeks	1.386***	1.228***	1.148***	1.385***	1.228***	1.141***
	(0.241)	(0.074)	(0.027)	(0.021)	(0.005)	(0.002)
Copula(sch)	-0.02	0.143	0.137***	-0.008	0.139	0.141***
	(0.402)	(0.178)	(0.046)	(0.402)	(0.198)	(0.043)
Copula(sch2)	-0.013	0.139	0.139***	-0.027	0.140	0.145***
	(0.311)	(0.198)	(0.047)	(0.385)	(0.188)	(0.042)
Copula(weeks)	-0.034	-0.063***	-0.002	-0.034***	-0.063***	-0.001***
	(0.046)	(0.017)	(0.004)	(0.004)	(0.001)	(0.000)
IMR				-0.063	0.015	-0.166***
				(0.238)	(0.046)	(0.008)
Constant	-2.392	-1.324**	-1.359***	-2.134*	-1.380***	-0.890***
	(1.518)	(0.605)	(0.168)	(1.280)	(0.397)	(0.084)
Obs.	182	505	5,129	182	505	5,129
R^2	0.44	0.66	0.77			

Notes: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Copula() are Gaussian Copula functions. Significant copula functions implies significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3D.5: Bootstrap Aggregation Results (Pay Periods) -Tanzania

	GC			HGC		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
sch	-0.028***	-0.035***	-0.010	-0.050***	-0.040***	-0.019*
	(0.007)	(0.008)	(0.010)	(0.007)	(0.008)	(0.010)
sch2	0.005***	0.009***	0.006***	0.010***	0.010***	0.005***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
age	0.053***	0.046***	0.099***	0.028***	0.039***	0.085***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)
age2/100	-0.065***	-0.053***	-0.095***	-0.028***	-0.043***	-0.078***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)	(0.000)
female	-0.736***	-0.552***	-0.336***	-0.498***	-0.491***	-0.211***
	(0.001)	(0.001)	(0.001)	(0.005)	(0.011)	(0.003)
rural	-0.520***	-0.191***	-0.257***	-0.547***	-0.215***	-0.093***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.005)	(0.003)
panel	-0.216***	-0.129***	-0.089***	-0.173***	-0.126***	-0.074***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)
weeks	1.074***	0.999***	1.044***	1.068***	0.998***	1.036***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Copula(sch)	0.032	0.022	0.094**	0.032	0.023	0.082**
	(0.043)	(0.045)	(0.039)	(0.043)	(0.046)	(0.039)
Copula(sch2)	0.031	0.023	0.094**	0.031	0.020	0.080**
	(0.043)	(0.044)	(0.039)	(0.044)	(0.045)	(0.039)
Copula(weeks)	0.074***	0.096***	0.009***	0.073***	0.096***	0.009***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
IMR				-0.689***	-0.190***	-0.503***
				(0.016)	(0.035)	(0.008)
constant	-0.481***	-0.597***	-1.846***	0.919***	-0.121	-0.782***
	(0.075)	(0.072)	(0.098)	(0.078)	(0.107)	(0.091)
Obs.	3,738	1,929	4,830	3,738	1,929	4,830
R^2	0.73	0.79	0.71			

Notes: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Copula() are Gaussian Copula functions. Significant copula functions implies significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3D.6: Bootstrap Aggregation Results (Pay Periods) - Uganda

	GC			HGC		
Period	Daily	Weekly	Monthly	Daily	Weekly	Monthly
sch	0.103**	0.106*	0.040***	0.106**	0.100*	0.040***
	(0.051)	(0.058)	(0.009)	(0.051)	(0.059)	(0.009)
sch2	-0.002***	0.003***	0.004***	-0.002***	0.003***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
age	0.063***	0.078***	0.078***	0.065***	0.074***	0.081***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
age2	-0.082***	-0.092***	-0.081***	-0.087***	-0.082***	-0.086***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.002)	(0.002)
female	-0.592***	-0.542***	-0.308***	-0.680***	-0.405***	-0.328***
	(0.003)	(0.005)	(0.001)	(0.019)	(0.025)	(0.010)
rural	-0.233***	-0.368***	-0.229***	-0.294***	-0.313***	-0.250***
	(0.002)	(0.004)	(0.002)	(0.013)	(0.011)	(0.013)
panel	0.009***	-0.062***	0.187***	0.013***	-0.066***	0.172***
	(0.002)	(0.004)	(0.001)	(0.002)	(0.005)	(0.004)
weeks	1.164***	1.158***	1.157***	1.163***	1.154***	1.162***
	(0.002)	(0.004)	(0.001)	(0.002)	(0.004)	(0.002)
Copula(sch)	0.032	-0.094	0.08	0.028	-0.087	0.081
	(0.148)	(0.195)	(0.066)	(0.145)	(0.198)	(0.068)
Copula(sch2)	0.029	-0.092	0.08	0.027	-0.093	0.077
	(0.150)	(0.193)	(0.066)	(0.147)	(0.203)	(0.066)
Copula(weeks)	0.024***	-0.020***	-0.007***	0.024***	-0.020***	-0.008***
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)
IMR				0.151***	-0.366***	0.063
				(0.031)	(0.068)	(0.044)
constant	-1.804***	-2.297***	-2.406***	-2.040***	-1.457***	-2.573***
	(0.311)	(0.375)	(0.106)	(0.319)	(0.413)	(0.188)
Obs.	1,262	589	2,765	1,262	589	2,765
R^2	0.57	0.57	0.63			

Notes: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Copula() are Gaussian Copula functions. Significant copula functions implies significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3D.7: Bootstrap Aggregation Results - *Ganyu* Labour

	GC			HGC		
	DailyC	MonthlyC	MonthlyA	DailyC	MonthlyC	MonthlyA
sch	-0.001	-0.018**	-0.007	-0.003	-0.024***	-0.010
	(0.006)	(0.008)	(0.007)	(0.006)	(0.008)	(0.007)
sch2	0.002***	0.003***	0.002***	0.002***	0.005***	0.003***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
age	0.034***	0.047***	0.040***	0.032***	0.039***	0.036***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
age2/100	-0.040***	-0.057***	-0.048***	-0.037***	-0.043***	-0.040***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
female	-0.240***	-0.444***	-0.341***	-0.223***	-0.367***	-0.299***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
rural	-0.284***	-0.420***	-0.326***	-0.329***	-0.623***	-0.439***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.002)	(0.002)
year	1.326***	1.223***	1.217***	1.298***	1.097***	1.150***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
weeks			0.982***			0.981***
			(0.000)			(0.000)
Copula(sch)	0.009	0.001	0.012	0.010	0.003	0.013
	(0.018)	(0.025)	(0.021)	(0.018)	(0.025)	(0.021)
Copula(sch2)	0.009	0.0002	0.012	0.010	0.003	0.013
	(0.018)	(0.025)	(0.021)	(0.018)	(0.025)	(0.021)
Copula(weeks)			0.020***			0.019***
			(0.000)			(0.000)
IMR				-0.095***	-0.429***	-0.236***
				(0.003)	(0.004)	(0.004)
Constant	0.217***	2.643***	-0.832***	0.361***	3.289***	-0.473***
	(0.031)	(0.042)	(0.034)	(0.032)	(0.042)	(0.036)
Obs.	16,528	16,528	16,528	16,528	16,528	16,528
R^2	0.60	0.38	0.77			

Notes: GC and HGC results computed by averaging coefficients from 10,000 and 1,000 replications, respectively. Copula() are Gaussian Copula functions. Significant copula functions implies significant correlation between the variable and the errors in the regression models. IMR is the inverse mills ratio. Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 4

Pay Period and the Distributional Effect of Education on Earnings: Evidence from Recentered Influence Function Regressions

4.1 Introduction

Earning gaps and inequality between different groups and their determinants have been extensively explored over the last three decades (Fortin et al., 2011). Studies on earning gaps and inequalities have been approached from several angles including gender (men vs women), location (rural vs urban), sector of employment (public vs private or formal vs informal), and age group (youth vs adults). Research has also analysed the trend in earnings inequality over time, whereby inequality between any two periods is compared (Firpo et al., 2018; Rios-Avila, 2020b).

There is a wide range of measures in economics that have been employed to measure earnings inequality, ranging from simple mean comparison to more advanced measures that go beyond the mean. Popular measures include the Gini, the Theil index, variance of the logarithm of earnings, interquantile share ratios (such as the Palma Ratio[1]) and interquantile range. Table 4.1 below shows the trend of income[2] inequality for Malawi, Tanzania and Uganda, computed from nationally representative household survey. On average, Malawi and Uganda recorded about the same level of inequality while Tanzania had the lowest

[1]Proposed in 2013 by Alex Cobham and Andy Sumner and named after Jose G. Palma, Palma Ratio is a measure of income inequality defined as the ratio of income earned by the richest 10% to that of the poorest 40% (Cobham et al., 2016).

[2]Income proxied by per adult equivalent household consumption.

inequality.

Table 4.1: Income Inequality Trends in Malawi, Tanzania and Malawi

Country &Year	Data	Measure of inequality	Inequality
Malawi			
2004	IHS2	Gini	40
2010	IHS3	Gini	45
2016	IHS4	Gini	45
Tanzania			
2000	HBS 2001	Gini	37
2007	HBS 2007	Gini	40
2011	HBS 2012	Gini	38
2017	HBS 2018	Gini	40
Uganda			
2002	UNHS 2002	Gini	45
2005	UNHS 2005	Gini	43
2009	UNHS 2009	Gini	44
2012	UNHS 2012	Gini	41
2016	UNHS 2016	Gini	43

Source: World Bank Povcal. Note: IHS stands for the Integrated Household Survey, HBS for the Household Budget Survey and UNHS for Uganda National Household Survey.

Empirical estimation of the determinants of earning gaps and inequality tends to aggregate reported earnings to a common unit. Typically, earnings are recorded by day, week, month, and year. While that could work well for developed countries with well-developed labour markets, the situation might be different for developing countries, including those in Sub Saharan Africa (SSA) where pay periods shorter than monthly (such as daily or weekly) are most common in the informal sector. In the first essay, we found that, in East Africa, the relationship between earnings and education varies across workers reporting wage earnings daily, weekly, and monthly. Since pooling the periods can lead to inaccurate estimates of returns to education, it is also likely to affect estimates of the distributional effect of education on earnings, essentially because the pay periods have different earnings distributions.

This essay adds to the literature by analysing both the distributional effect of education within each of the three pay periods; and examining how gender differences in educational attainment explain earning gaps and inequality between women and men in Malawi, Tanzania and Uganda. Specifically, the essay seeks to answer three questions. Firstly, for each pay period, how does a change in education distribution affect the unconditional distribution of earnings? Secondly, does education's role in explaining the unconditional distribution of earnings differ along the earnings distribution? Lastly, does the difference in education by gender significantly explain the gender earnings gap within the pay periods?

As an extension, this essay seeks to answer the above questions while exploring a particular type of casual employment specific to Malawi, namely *ganyu*. To the best of our knowledge, this essay is the first study in SSA to analyse the link between education and earnings distribution while considering the pay period's effect. The aim is to better understand the effects of education on the earnings distribution and the gender earnings gap in East Africa, which is essential to devise education policies and programs to curb earnings inequality.

Using comparable nationally representative data from the three countries (i.e., Malawi, Tanzania, and Uganda), we employ Recentered Influence Function (RIF) Regressions to examine how the distributional effect of education on wage earnings vary by pay period. Precisely, we begin with employing unconditional quantile regressions (Firpo et al., 2009) to examine the possible heterogeneous effects of education on earnings and how these effects vary when workers report wages over different periods. We then employ the reweighted RIF Oaxaca-Blinder decomposition (Firpo et al., 2018; Rios-Avila, 2020b) to assess how education explains gender earnings inequalities within each pay period.

Our results show that there is heterogeneity in the effects of a change in the distribution of education on the distribution of earnings across pay periods. Generally, the effect is stronger for workers reporting monthly earnings compared to their daily and weekly counterparts. The results of how education contributes to earnings inequality within the pay periods suggest that if the average education of the population were to increase by a year, it would reduce earnings inequality for workers paid daily and monthly but increase inequality for workers paid weekly. We also show that gender differences in educational attainment is a significant factor in explaining earnings differences between female and male workers in Tanzania and Uganda.

Finally, we examine earnings inequality in casual employment using *ganyu* labour in Malawi as our case study. We find that increase in the population's average education by a year increases the mean earnings of *ganyu* workers by 7 – 16% depending on the quantile of earnings distribution. About seven percent (7%) of the gender earnings gap associated with gender differences in endowments can be attributed to gender differences in educational attainment. This suggests that policies to raise females' education endowments are a good solution to curb gender earnings inequality in the *ganyu* labour market.

It is worth noting that pooling the periods together gives an incomplete picture of the distribution effects of education on earnings. For example, in the case of Malawi (excluding *ganyu*) and Uganda where most of the workers report earnings monthly, the results show that pooling yields estimates of the effects of education on earnings which are leaning towards those from the monthly sample. This reiterates the need to estimate for each of the periods separately for a more informative inference

The rest of the chapter is organised as follows: Section 2 provides an overview of the related literature. Section 3 describes the empirical methodology used, followed by Section 4 on data and description. Section 5 presents the results and discussions, and Section 6 concludes.

4.2 Literature Review

Labour earnings account for a sizable proportion of individual incomes and thus are important in explaining income inequality (Peichl and Pestel, 2015). The determinants of earnings inequality between different groups or its trend over time have consequently attracted much research over the last three decades. Ever since the seminal works of Oaxaca (1973) and Blinder (1973), economists have developed several methodological frameworks to analyse earnings inequality. Recent research has focused on formulation and application of methods that go beyond mean decomposition to other inequality measures such as variance, quantiles (conditional and unconditional), inter-quantile range and the Gini. We do not discuss the different methods here; a review of these are provided in Fortin et al. (2011).

While most research on earnings inequality has focused on developed countries, little research on this matter has been undertaken in SSA. However, with increasing availability of data over the past decade, there has been a growing body of empirical research across the region. Given that in labour market surveys workers report wages over different pay periods (such as hourly, daily, weekly etc.), these studies usually have aggregated these wages into a common earnings measure such as monthly or annual earnings. Peichl and Pestel (2015) argue that the distribution of such a common measure of earnings is affected by the components that are used in its construction. However, since this has been the practice in the literature, in this part, we summarise some of these studies on SSA with a particular focus on education as one of the key determinants of earnings inequality.

One of the dimensions of inequality that has received considerable attention is the gender earnings inequality/gap. The consensus is that, like many other regions of the world, males earn considerably more than females (Agesa et al., 2013; Joseph and Leyaro, 2019; Nix et al., 2016; Nordman et al., 2011). Many papers have explored what factors determine the earnings differentials across gender. According to Nix et al. (2016), the gender wage gap is predominantly explained by differences in returns to the observable characteristics, although differences in endowments do matter. They also found that the coefficients of the determinants of the earnings gap between gender vary across the earnings distribution and sectors of employment. Because males generally have more education than females, increasing education endowments of women would raise women's wages and thus reduce the gender gap in earnings (Agesa et al., 2013; Joseph and Leyaro, 2019).

In another gender dimension, Kilic et al. (2015) examined gender differences in agricultural productivity (as measured by plot gross value of output) using nationally representative data from Malawi. They applied both the Oaxaca-Blinder type decomposition and RIF regressions to decomposing the productivity differences both for the mean and at various quantiles of the productivity distribution. They found that gender differences in education endowments were only significant in explaining the productivity inequality at the mean and centre of the productivity distribution. Women had higher returns to education than men. Thus, assigning their coefficients to men would increase men's productivity and widen the gender gap further.

Numerous studies have also considered earnings differentials between public and private sector workers, with a definite gap in favour of the public sector (Kwenda and Ntuli, 2018; Nielsen and Rosholm, 2001). Nielsen and Rosholm (2001), applied quantile regressions on three waves of household data from Zambia and found that the returns to education varied along the earnings distribution, however, there existed a larger gap for those in the bottom end of the distribution across all education levels. Kwenda and Ntuli (2018), on the other hand, employed the Oaxaca-Blinder decomposition method on an extensive cross-sectional dataset[3] from South Africa. They found that the distribution effects of education on wages differed within and across sectors of employment. Across sectors, the effect was higher in the private than the public sector. Within the public sector, the distributional effect of education declined by quantile while within the private sector, the effect of education was non-monotonic: initially declined and then increased by quantile.

Earnings differences between periods has also been widely explored to analyse trends in earnings inequality. Essama-Nssah et al. (2013) used expenditure data for 2001 and 2007 in Cameroon to analyse income inequality between the two periods. They employed the RIF decomposition method and found that returns to

[3]The Post-Apartheid Labour Market Series (PALMS)

education varied slightly along the earnings distribution (not very heterogenous), were positive and statistically significant across quantiles, and were on average higher for 2001 than 2007.

Although research on earnings inequality in SSA has increased in recent years, what is noticed in all the previous studies (regardless of the kind or measure of inequality) is the conventional method of aggregating earnings to a common unit. So far, researchers have disregarded the importance of pay periods on inequality decomposition. In the first essay, we found that the relationship between earnings and education may vary across workers reporting wage earnings over different periods including day, week, and month. We believe that pooling all the workers together and aggregating their earnings to a common unit might also lead to inaccurate estimates of the distributional effect of education on earnings, especially if different pay periods have different wage distributions. This study aims to fill this gap by estimating the distributional effects of education on earnings as well as in decomposing the earnings gap between gender by pay period.

4.3 Empirical Strategy

The following extended Mincer equation (with education assumed to be exogenous[4]) is used to model the relationship between education and earnings.

$$Y_{it} = \alpha_1 S_{it} + \alpha_2 S_{it}^2 + \delta Z_{it} + \mu_{it} \tag{4.1}$$

Where Y is the log of wage earnings, S and S^2 are individual's years of schooling and its quadratic, Z is a vector (containing a constant) of individual characteristics (age in years and its square, logarithm of number of weeks worked over the last twelve months, and dummies for gender, rural residence, individuals observed more than once, and survey year), i and t index individual and time respectively and μ is a standard error term. The parameters[5] of interest are α_1 and α_2.

[4]Note that in RIF regression and decomposition, we ignore endogeneity due to unobserved ability and selection into employment categories because the methods to address the same are not yet explicit or available (Rios-Avila, 2020a; Kwenda and Ntuli, 2018)

[5]On a few occasions, we assume $\alpha_2 = 0$ to simplify interpretation.

4.3.1 RIF Regression

For simplicity and mathematical convenience, we rewrite the linear regression (4.1) in the following form:

$$Y = X'\beta + \varepsilon \tag{4.2}$$

Where $E(\varepsilon)$ =0, $Y = [y_1, y_1, ...y_n]$ is a vector of the observed values of Y, and X is a vector (containing the constant) of all explanatory variables. The influence function of the observed value y of the distribution statistic of interest $v(F_Y)$ is defined as $IF(y; v)$. The RIF is then defined as

$$RIF(y; v) = v(F_Y) + IF(y; v) \tag{4.3}$$

Such that the statistic of interest can be obtained by integration. That is,

$$v(F_Y) = \int RIF(y; v)dF(y) \tag{4.4}$$

The expectation of the RIF conditional on X (the explanatory variables) is modelled as a linear function of X as:

$$E[RIF(Y; v)|X] = v(F_Y) = X'\beta \tag{4.5}$$

Where β is a vector of parameters which can be estimated using OLS. For quantiles, RIF is given by

$$RIF(Y; q_\tau) = q_\tau + \tau - \frac{1\{Y \leq q_\tau\}}{f_Y(q_\tau)} \tag{4.6}$$

q_τ can be estimated from the data by sample quantile $\hat{q_\tau}$ whereas $f_Y(\hat{q_\tau})$ can be estimated using Kernel density. The RIF for quantile of Y is an indicator variable (that is, $1(Y \leq q_\tau)$) which takes the value of 1 if the outcome variable is less than or equal to the quantile q_τ. It can, therefore, be modelled using a linear probability model (LPM), probit or a non-parametric binomial model (Firpo et al., 2009). Empirically, estimation of the RIF regressions for quantiles of log wages (or any other statistic such as interquantile share ratio in our case) can be performed in Stata using user-written command rifhdreg (Rios-Avila, 2020b). As explained earlier, interquantile share ratio (iqsr) is a measure of inequality calculated as the ratio of total income received/earned by a certain percent of the population with the highest income to the total received/earned by some percent of the population with the lowest income. Iqsr is among the most flexible, intuitive and easy to interpret measures of inequality, and can easily be customised to fit the purpose. For instance, can be calculated as the income ratio of the income earned by the richest 10% to the poorest 10%, the ratio earned by the richest 20% to the poorest 20%, the ratio earned by the richest 10% the poorest 40% etc. In this essay we use the latter which is also known as the Palma Ratio. While it is expected that estimates of the determinants of inequality to differ depending on the choice of the measure of inequality (see for example Rios-Avila (2020a)) the reasons and extent to which the results differ are beyond the scope of this essay.

4.3.2 RIF Decomposition

To examine gender wage gap and gender differences in earnings inequality within the pay periods, we employ RIF based decomposition, an extension of Oaxaca-Blinder (OB hereafter) methodology proposed by Firpo et al. (2009) and further extended by Firpo et al. (2018). RIF decomposition uses RIF regression in combination with reweighting to decompose any statistic of interest into two parts: the difference due to endowments (characteristics or composition effect) and the difference due to wage structure effects (coefficient effect). Using this decomposition, the contribution of each explanatory variable on the two parts can be examined. In addition, the decomposition shows the size of the specification and reweighting errors which are essential in assessing the accuracy of the model.

Following Rios-Avila (2020b), the derivation of RIF decomposition is as follows: Recall (4.2), i.e., $Y = X'\beta + \varepsilon$. Suppose there is some categorical variable T such that the joint distribution function of Y, X and T is given by $f_{Y,X,T}(y_i, x_i, T_i)$. For only two groups (T=0 and T=1) the joint distribution function is given as:

$$f^k_{Y,X}(y,x) = f^k_{Y|X}(Y|X)f^k_X(X) \tag{4.7}$$

Where $T = k \in [0,1]$; and its cumulative distribution function conditional on T as:

$$F^k_Y(y) = \int f^k_{Y|X}(Y|X)dF^k_X(X) \tag{4.8}$$

In our case T is an indicator variable for gender defined by

$$T = \begin{cases} 1, & \text{if } female \\ 0, & \text{if } male \end{cases}$$

The cumulative distribution of Y conditional on T can then be used to decompose the difference in the distribution of statistic v between the two groups. Accordingly,

$$\Delta v = v_1 - v_0 = v(f^1_Y) - v(f^0_Y) \tag{4.9}$$

Which implies

$$\Delta v = v(f^1_{Y|X}(Y|X)dF(X)) - v(f^0_{Y|X}(Y|X)dF(X))$$

We can rewrite (4.9) as

$$\Delta v = v_1 - v_c + v_c - v_0$$

Alternatively, in a reduced form

$$\Delta v = \Delta v_S + \Delta v_X$$

Where v_c is some counterfactual statistic defined as

$$v_c = v(f^c_Y) = v(f^0_{Y|X}(Y|X)dF^1_X(X)) \tag{4.10}$$

$\Delta v_S = v_1 - v_c$ is the difference attributed to the relationship between Y and

X; and $\Delta v_X = v_c - v_0$ the difference arising due to differences in characteristics, the Xs.

From (4.5), $v(F_Y) = X'\beta$, It follows therefore that

$$v_1 = E(RIF(y_i; v(f_Y^1))) = \bar{X}^{1'}\hat{\beta}^1$$

$$v_0 = E(RIF(y_i; v(f_Y^0))) = \bar{X}^{0'}\hat{\beta}^0$$

and

$$v_c = \bar{X}^{1'}\hat{\beta}^0$$

Since the counterfactual distribution is not observed, it is approximated as follows

$$F_Y^c = \int f_{Y|X}^0(Y|X)dF_X^1(X) \cong \int f_{Y|X}^0(Y|X)dF_X^0(X)\omega(X) \tag{4.11}$$

Where $\omega(X)$ is a reweighting factor defined as

$$\omega(X) = \frac{1-p}{p}\frac{P(T=1|X)}{1-P(T=1|X)} \tag{4.12}$$

where p is the proportion of people in group 1 and $P(T=1|X)$ the probability that an individual belongs to group 1 given that she has characteristics X. The reweighting factor can be obtained after the conditional probability is estimated using a probit or logit model. Plugging the reweighting factor into (4.10) yields

$$v_c = E(RIF(y_i; v(f_Y^c))) = \bar{X}^{c'}\hat{\beta}^c \tag{4.13}$$

The decomposition can then be rewritten as

$$\Delta v = \bar{X}^{1'}(\hat{\beta}^1 - \hat{\beta}^c) + (\bar{X}^1 - \bar{X}^c)'\hat{\beta}^c + (\bar{X}^c - \bar{X}^0)'\hat{\beta}^0 + \bar{X}^{c'}(\hat{\beta}^c - \hat{\beta}^0)$$

Define $\Delta v_S^p = \bar{X}^{1'}(\hat{\beta}^1 - \hat{\beta}^c)$; $\Delta v_S^e = (\bar{X}^1 - \bar{X}^c)'\hat{\beta}^c$; $\Delta v_X^p = \bar{X}^c - \bar{X}^0)'\hat{\beta}^0$ and

$\Delta v_X^e = \bar{X}^{c'}(\hat{\beta}^c - \hat{\beta}^0)$

Then,

$$\Delta v = \Delta v_S^p + \Delta v_S^e + \Delta v_X^p + \Delta v_X^e \tag{4.14}$$

The component $\Delta v_S^p + \Delta v_S^e$ is called the coefficient effect which constitutes of the pure coefficient effect (Δv_S^p) and the reweighting error (Δv_S^e). The component $\Delta v_X^p + \Delta v_X^e$ is called the aggregate composition effect and constitutes the pure composition effect (Δv_X^p) and specification error (Δv_X^e). The error components help assess the quality of the reweighting and specification of the regression function (Rios-Avila, 2020b). For more robust results, the coefficients of these components should be smaller and insignificant. The empirical estimation of the RIF decomposition for the mean of log consumption to poverty line ratio is performed in Stata using user-written command *oaxaca_rif* (Rios-Avila, 2020b).

4.4 Data and Descriptive Statistics

The sources of data for analysis are described in detail in Chapter 2 of this thesis. The literature guides the variables used in this essay. Table 4.2 shows the names and definitions of each variable used in the analysis. Chapter 2 of the thesis provides a detailed description of the data sources and how the variables were constructed from the data.

Table 4.3 shows, for each country and pay period, the means and standard deviations for the continuous variables and the percentage composition of the categorical variables in the sample. Earnings are annualised and expressed monthly and thus comparable across countries and pay periods. Workers in Malawi earn more than those in Tanzania and Uganda across the pay periods. In all three countries, workers reporting earnings monthly are the highest wage earners (a possible reason is monthly may have relatively larger proportion of workers in formal employment). Workers reporting earnings by week are the lowest wage earners in Tanzania and Uganda. Compared to the other countries, in Tanzania the wage penalty associated with working in daily or weekly employment is enormous.

Table 4.2: Definition of Variables Used in the Analysis

Variable	Description
Log(earnings)	Logarithm of annualised earnings (expressed monthly).
sch	Individual's total number of years of schooling.
age	Individual's age in years. Its square is included to capture the non-linear relationship between earning and age.
noeduc	educational dummy, 1 if less than primary education and 0 otherwise.
primary	educational dummy, 1 if completed primary education and 0 otherwise.
secondary	educational dummy, 1 if completed ordinary/advanced secondary education and 0 otherwise.
higher	educational dummy, 1 if completed post-secondary (diploma/university) education and 0 otherwise.
female	a gender dummy, 1 for females, included to capture the effects of gender on wages.
rural	location dummy, 1 for employment in rural areas, is used to control for rural-urban wage differentials.
panel	a dummy (for Tanzania and Uganda), 1 for individuals observed more than once since we are using an imperfect panel survey.
year	year dummy (for Malawi), 1 for 2016 and 0 for 2010 since we are using pooled cross-section data.
weeks	logarithm of the number of weeks worked in the past 12 months.

That is, workers reporting earnings by day and by week earn no more than a third of their counterparts who report earnings by month.

As far as education is concerned, there are a few issues that could potentially affect our results. While workers in Malawi have more years of schooling compared to Tanzania and Uganda, monthly earners have more education than their daily and weekly counterparts in all three countries. In Tanzania, there are no workers with higher education reporting earnings by day or week. In Malawi, only 12% of the workers reporting earnings by day and 4% of those reporting earnings by week have higher education, while in Uganda 3% and 8% of the workers reporting earnings by day and week respectively have higher education.

Table 4.3: Summary Statistics for the Main Variables Used in Analysis

Country & Sample	Obs.	Wage ($ month)		sch		age		weeks		primary	secondary	higher	female	rural	panel	year
		Mean	SD	Mean	SD	Mean	SD	Mean	SD	%	%	%	%	%	%	
Malawi																
Daily	182	206.42	251.48	9.03	4.51	36.80	10.20	36.75	13.99	20	29	12	26	67	NA	74
Weekly	505	174.26	289.04	7.15	3.91	34.94	10.35	35.51	14.71	24	10	4	30	66	NA	53
Monthly	5,129	226.84	339.20	9.26	4.12	35.85	10.71	39.63	12.69	26	24	14	25	49	NA	50
Pooled	5,816	221.04	332.38	9.05	4.16	35.79	10.66	39.13	13.01	26	23	13	25	52	NA	51
Ganyu	16,528	33.76	53.52	4.77	3.52	33.42	11.89	15.64	12.66	14	2	0	51	92	NA	64
Tanzania																
Daily	3,738	38.91	87.35	5.26	3.21	33.14	11.92	15.32	15.39	57	4	0	40	78	25	NA
Weekly	1,929	32.54	85.08	5.28	3.24	33.64	12.08	13.62	15.16	56	5	0	36	82	19	NA
Monthly	4,830	123.90	161.28	8.16	3.69	33.34	11.89	34.95	16.00	51	27	7	38	53	41	NA
Pooled	11,215	69.10	126.88	6.35	3.67	33.37	11.93	22.10	18.38	55	13	3	38	70	45	NA
Uganda																
Daily	1,262	78.92	103.44	6.35	3.53	30.99	10.84	36.04	14.57	34	11	3	20	64	39	NA
Weekly	589	76.34	109.07	6.97	3.81	32.77	11.44	35.27	14.87	32	12	8	29	72	15	NA
Monthly	2,765	114.19	139.02	9.85	4.39	34.71	11.06	39.38	12.45	29	16	31	37	56	56	NA
Pooled	4,631	99.66	127.84	8.51	4.41	33.46	11.18	37.89	13.54	31	14	20	31	61	57	NA

Source: Author's computations from IHS, TNPS and UNPS.
Note: The last two columns show % observed multiple times for Tanzania and Uganda (panel) and % in 2016 for Malawi (year) respectively. The variable 'weeks' is in its original scale (prior to taking logs) The figures are adjusted by survey weights.

Figures 4.1 - 4.3 compare the distribution of earnings by pay period for each country. Figure 4.1 shows the distribution for Malawi. Because 88% of the workers report earnings by month, the distribution of the pooled periods looks very similar to that of workers who report earnings by month. This is also true for Uganda, where 60% of the workers report their earnings by month (Figure 4.3). For Tanzania, however, the distribution for the pooled periods is very different from those of the disaggregated pay periods (Figure 4.2). While the distribution for those reporting earnings by month and pooled for Tanzania are left-skewed, we observe more complicated shapes for the distribution curves for those reporting earnings by day and week.

Figure 4.1: Distribution of Monthly Earnings by Pay Period in Malawi

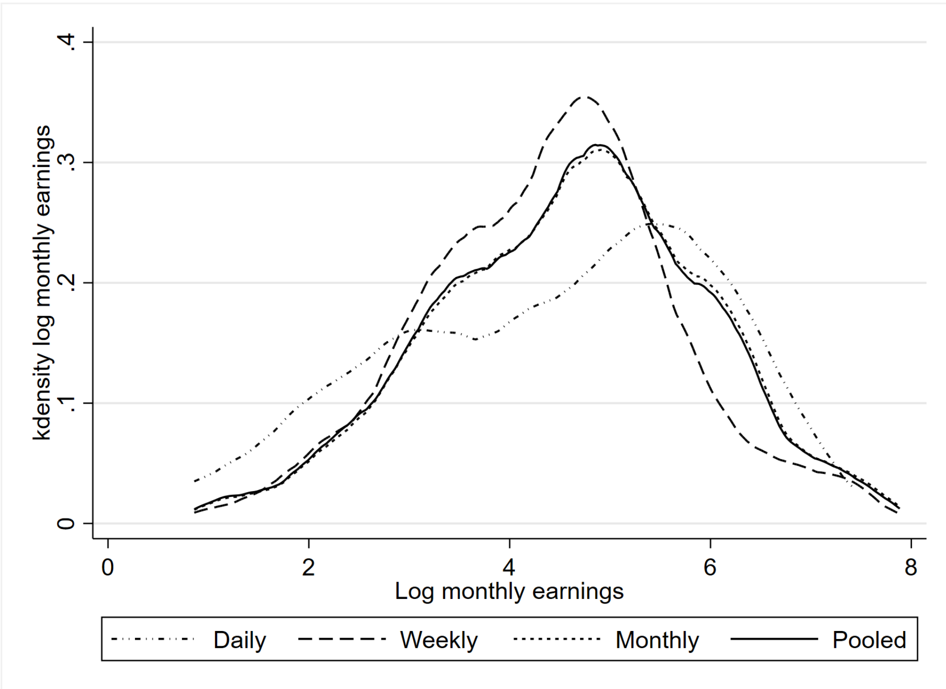


Figure 4.2: Distribution of Monthly Earnings by Pay Period in Tanzania

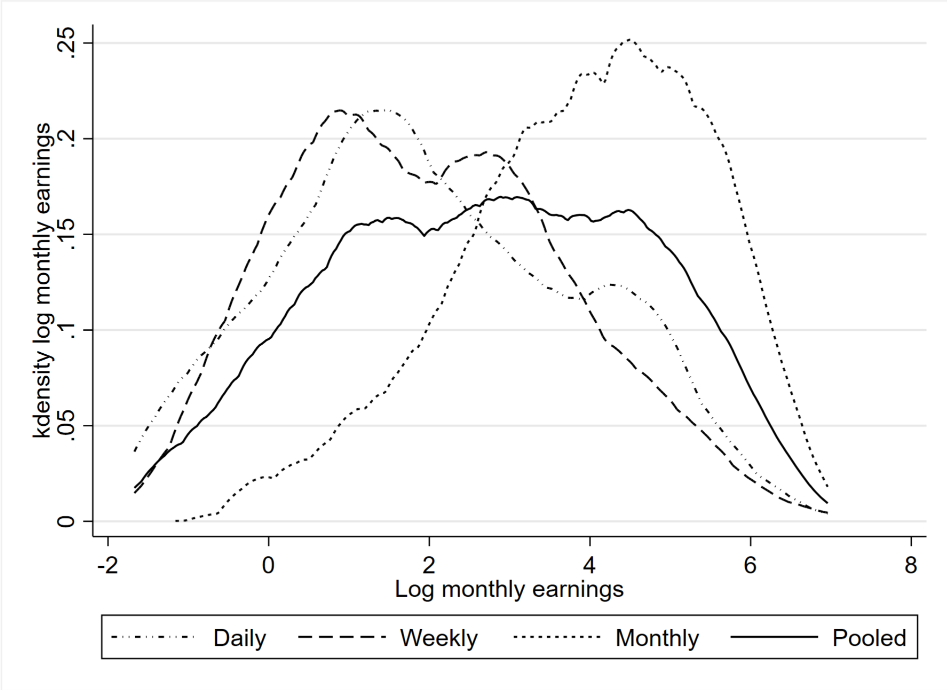
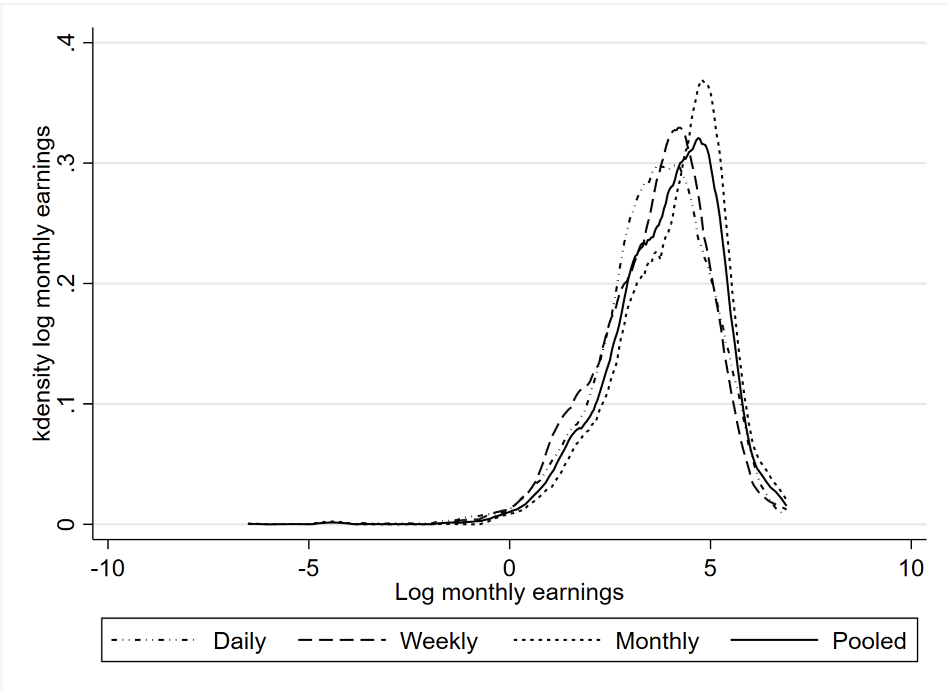


Figure 4.3: Distribution of Monthly Earnings by Pay Period in Uganda



Figures 4.4 – 4.6 show the cross-sectional relationship between earnings and education by pay period. The bars show the mean earnings by years of education. As expected, on average more years of education are associated with higher earnings in all three countries. The relationship is especially vivid for workers reporting earnings by month.

Figure 4.4: Distribution of Earnings by Education and Pay Period in Malawi

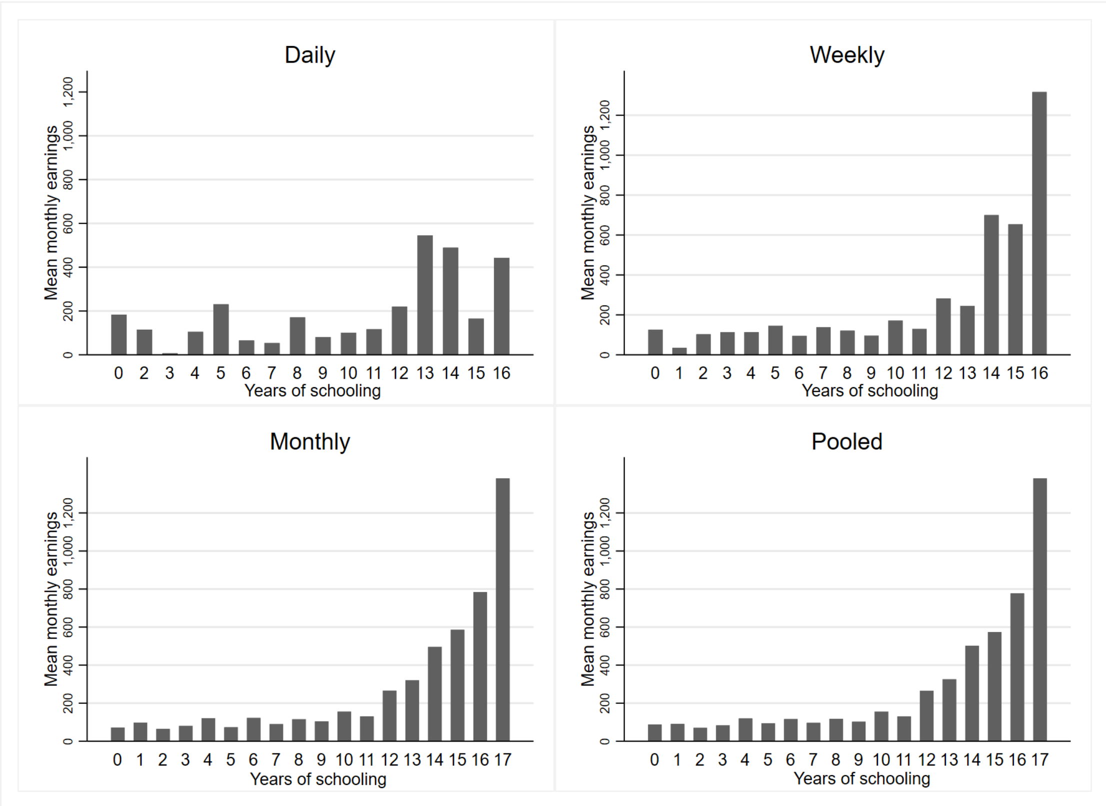


Figure 4.5: Distribution of Earnings by Education and Pay Period in Tanzania

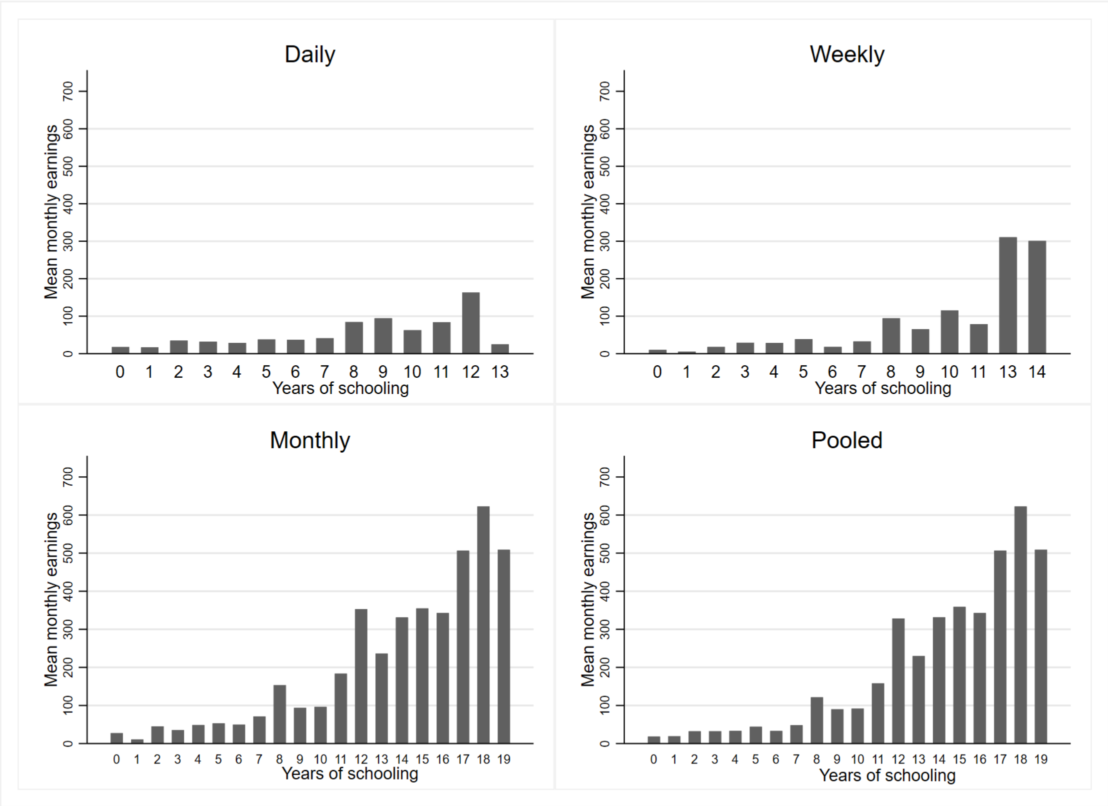
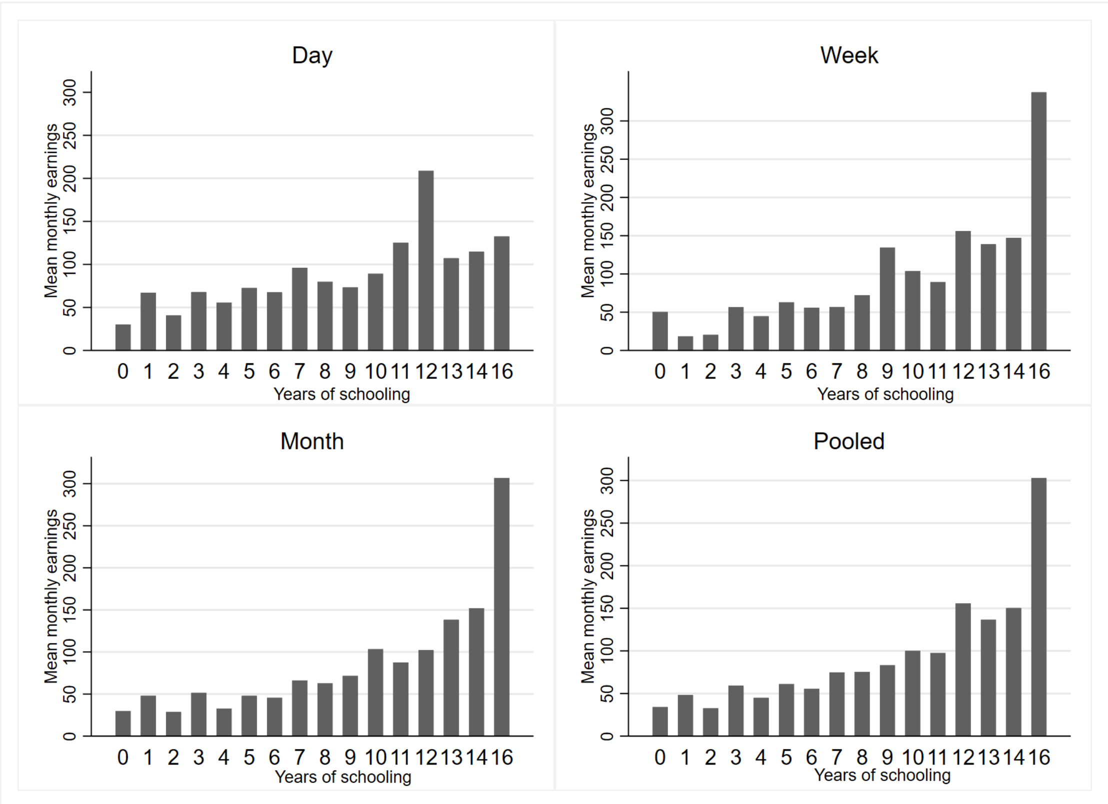


Figure 4.6: Distribution of Earnings by Education and Pay Period in Uganda



4.5 Results and Discussion

4.5.1 Pooled RIF Regression

In this section, we present estimates based on the pooled[6] RIF regressions. For each country, we analyse how education affects the distribution of earnings of the workers. That is, using unconditional quantile regressions, we examine the possible heterogeneous effects of education on earnings along the earnings distribution. The coefficients of education measure the expected change in the unconditional distribution of the earnings (as measured by quantiles of log earnings) when there is a small change in the distribution of education (Rios-Avila, 2020b).

Owing to its easy interpretability, we begin by presenting the results for linear specification (assuming $\alpha_2 = 0$ in equation (4.1). Table 4.4 present the results of the distribution effects of education in each of the three countries. The top panel of Table 4.4 shows the results for Malawi, the middle panel for Tanzania, and the bottom panel for Uganda, for the selected unconditional quantiles of earnings. The detailed results for all nine deciles are shown in Appendix 4A.

Table 4.4 shows that, in each country, education affects individuals at different points of the earnings distribution differently. For instance the results for Malawi suggest that an increase in the population's education by one year increases the mean of wages by approximately 7%, 13% and 12% at the bottom[7], centre and top of the distribution, respectively. The corresponding figures for Tanzania are -0.3%, 13% and 17%, respectively; while for Uganda are 11%, 15% and 9%, respectively. Figure 4.7 plots the coefficients of education from Table 4.4 but for all nine deciles. Figure 4.7 shows that for Malawi and Tanzania the effect of education generally increases (although declines sharply after 80th quantile for Malawi) with the quantiles of earnings while for Uganda it initially increases up to the 40th quantile then declines monotonically by quantile of earnings.

[6] Like the usual practice in the literature.

[7] Throughout the chapter bottom end of the distribution refers to the first deciles (10th percentile), middle/centre to the 50th percentile and top end to the nineth decile (90th percentile) of the earnings distribution.

Table 4.4: Unconditional Quantile Regression by Country

	(1) q(10)	(2) q(25)	(3) q(50)	(4) q(75)	(5) q(90)
Malawi					
sch	0.070***	0.125***	0.128***	0.161***	0.123***
	(0.007)	(0.006)	(0.005)	(0.007)	(0.006)
covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	2.841	3.672	4.692	5.558	6.298
R^2	0.39	0.45	0.48	0.37	0.19
Obs.	5,816	5,816	5,816	5,816	5,816
Tanzania					
sch	-0.003	0.023***	0.125***	0.174***	0.171***
	(0.007)	(0.006)	(0.008)	(0.006)	(0.007)
covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	0.307	1.545	3.199	4.615	5.461
R^2	0.30	0.55	0.58	0.37	0.20
Obs.	11,215	11,215	11,215	11,215	11,215
Uganda					
sch	0.105***	0.118***	0.153***	0.110***	0.092***
	(0.010)	(0.007)	(0.007)	(0.007)	(0.006)
covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	1.931	3.052	4.055	4.846	5.446
R^2	0.41	0.42	0.38	0.26	0.14
Obs.	4,631	4,631	4,631	4,631	4,631

Notes: Bootstrap standard errors computed by 500 replications in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Full results with all variables available in Appendix 4A.

Table 4.4 shows that, in all three countries, there is considerable difference in the RIF mean between the top and bottom deciles implying high degree of earnings inequality. Using RIF regression, we examine whether education is a significant determinant of earnings inequality, measured by the interquantile share ratio (*iqsr*) defined as the ratio of the share earned by workers in the top decile of earnings relative to that earned by those in the bottom four deciles. The results are presented in Table 4.5.

Table 4.5 shows that there is a substantial earnings inequality between the workers in the top decile of earnings and those in the bottom four deciles in all three countries. Inequality is highest among workers in Tanzania and lowest among workers in Uganda. Workers in the top decile in Tanzania earn

Figure 4.7: RIF returns - Malawi

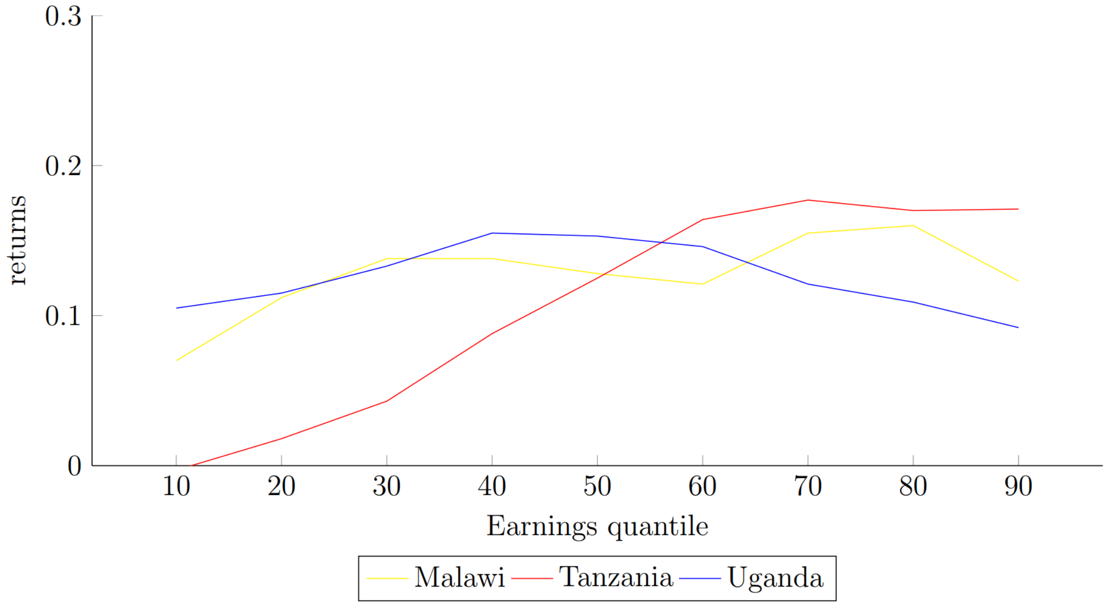


Table 4.5: Education and Wage Inequality (*iqsr*) by Country

Country	(1) Malawi	(2) Tanzania	(3) Uganda
sch	-0.072	2.321***	-0.347***
	(0.060)	(0.216)	(0.056)
covariates	Yes	Yes	Yes
iqsr	7.495	24.819	6.051
R2	0.11	0.38	0.17
Obs.	5,816	11,215	4,631

Notes: iqsr is the interquantile share ratio defined as the ratio of the share earned by the top decile relative to bottom four deciles. Bootstrap standard errors computed by 500 replications in parentheses; * p < 0.10, ** p < 0.05, *** p < 0.01. Full results with all variables available upon request.

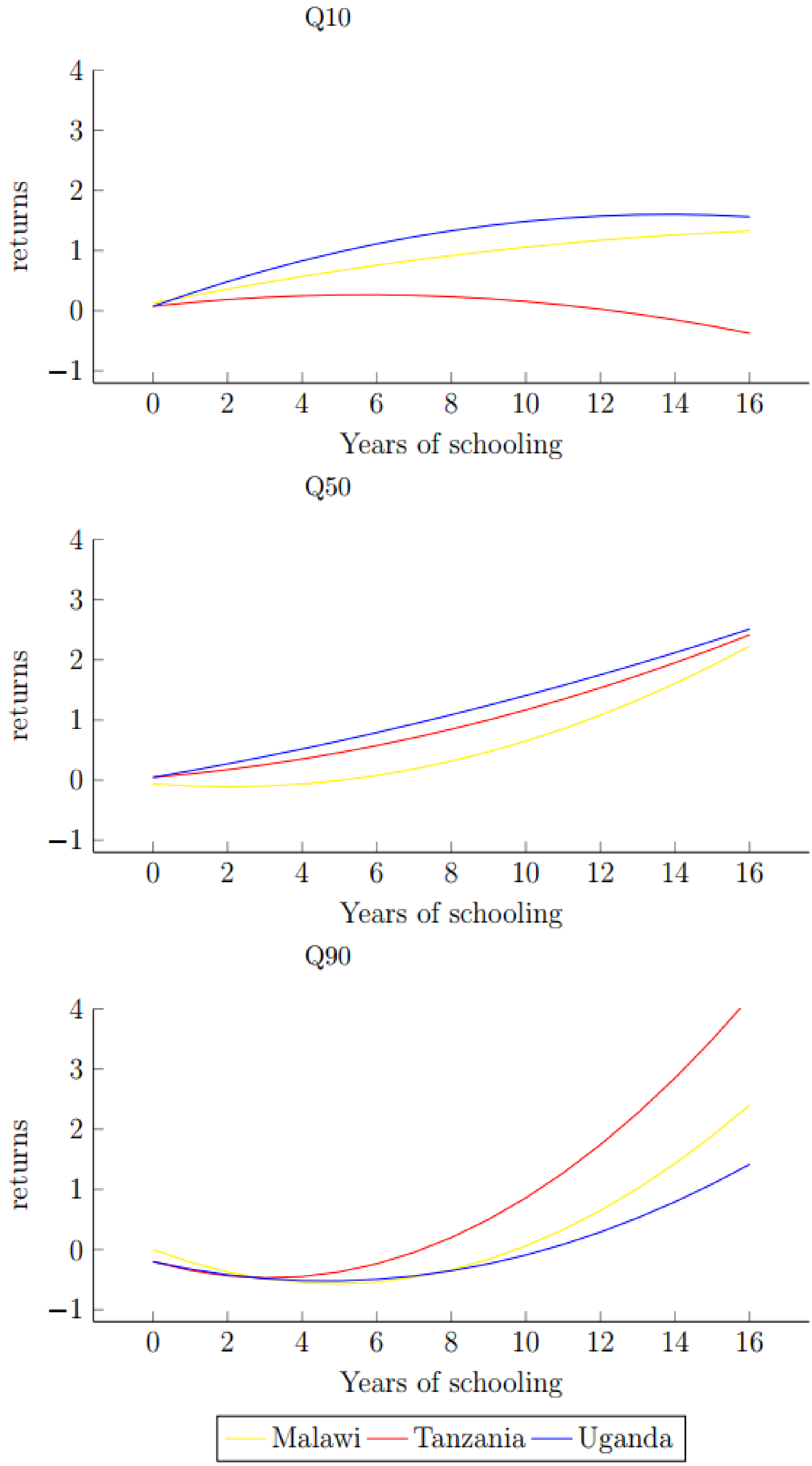
approximately twenty five (25) times as much as those in the bottom four deciles. The corresponding figures for Malawi and Uganda are 7.5 and 6, respectively. However, despite the high inequality within workers in Malawi, education does not seem to significantly drive the inequality. For Tanzania, an increase in education in the population by one year would result in an increase in earnings inequality by 9.4%,[8] other things equal. This implies that education is likely to benefit more those in higher than in low paying jobs. For Uganda, an increase in

[8]That is (2.321/24.819)*100.

average education in the population by a year will reduce inequality by 5.7%.

Figures 4.8 plots the marginal effects for the quadratic specification of returns against years of schooling for selected quantiles (10th, 50th and 90th); detailed results are in Appendix 4B. Figure 4.8 shows a concave relationship for workers in the bottom decile and strong convex relationship for workers in the top decile of earnings in all three countries. This suggest that, in all three countries, an increase in education in the population is more likely to benefit the higher than the lower earnings workers and hence likely to increase inequality. For workers in the top decile, the effects of education are very small (even negative) for early years of schooling but increase rapidly after about the 6th year.

Figure 4.8: RIF Coefficients of Education (quadratic) by Country



4.5.2 Pooled RIF Decomposition

In this subsection, we assess the role of education in explaining inequality in gender earnings. We begin by comparing the differences in means of earnings, educational attainment, and returns to education between male and female workers. Table 4.6 shows the raw gender differences in the three variables. Female workers have lower wage earnings than their male counterparts across the countries, but the difference is not statistically significant for Malawi. Females in Malawi and Uganda have more education than males but again the difference is not statistically different for Malawi. In line with the previous literature on returns to education in SSA, the right column of Table 4.6 shows that female workers have higher returns than their male counterparts in all three countries.

Following Rios-Avila (2020b) and Firpo et al. (2018) we then decompose the gender differences in the mean of log earnings as well as the difference in wage inequality (iqsr) by country (see Appendix 4C for decomposition for quantiles of log earnings). Since, as discussed earlier in Table 4.6, there is no statistical gender difference in the mean of wages for Malawi, we present the decomposition results only for Tanzania and Uganda. Table 4.7 shows the results for the reweighted RIF OB gender decomposition for the two countries. The first and the third columns of Table 4.7 show the results for mean decomposition and the second and the fourth columns the results for interquantile share ratio decomposition. Counterfactual is the estimated distribution of earnings, showing what would female mean wages (or inequality) be if they had the coefficients of male. Explained refers to the part of the gap due to gender differences in characteristics/endowments. Unexplained refers to the part of the gap due to gender differences in returns to those characteristics. The pure components are the differences net of specification and reweight errors.

The results in Table 4.7 suggest that education is among the significant factors explaining the gender earnings gap in both countries. Of the pure explained gender

Table 4.6: Gender Differences in Earnings (US$ per month) by Period and Country

	Earnings(US$ per month)	Education(years)	Returns to Education
Malawi			
Male	98.49	8.99	0.142
Female	97.04	9.23	0.182
Difference	1.45	-0.23	-0.040***
Obs. Male	4,358	4,358	4,358
Obs. Female	1,458	1,458	1,458
Obs. Total	5,816	5,816	5,816
Tanzania			
Male	28.21	6.68	0.093
Female	10.98	5.81	0.123
Difference	17.23***	0.86***	-0.030***
Obs. Male	7,142	7,142	7,142
Obs. Female	4,073	4,073	4,073
Obs. Total	11,215	11,215	11,215
Uganda			
Male	54.48	8.36	0.139
Female	31.43	8.83	0.158
Difference	23.05***	-0.46**	-0.019***
Obs. Male	3,156	3,156	3,156
Obs. Female	1,475	1,475	1,475
Obs. Total	4,631	4,631	4,631

Notes: Difference for earnings is defined as geometric mean for males minus geometric mean for females. Difference for education is defined as arithmetic mean for males minus arithmetic mean for females. Difference for returns to education defined as the returns (AME(sch)) for males minus the corresponding value for females. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

gap in earnings for Tanzania, differences in education explains approximately 7%. This suggests that if females had the same level of education endowments as males, their wage earnings would have been respectively 7% higher. The coefficient on education for Uganda is negative implying that, while education has a positive effect on earnings, females are better endowed with education and thus the gender difference in earnings would have been larger if on average females had the same (low) education endowments as males. In addition, the coefficients on education for the unexplained component are negative for both contries consistent with the fact that women have higher returns to education than men. On the other hand, the results from the iqsr decomposition in Table 4.7 suggest that gender differences

Table 4.7: Reweighted RIF Oaxaca-Blinder Gender Decomposition by Country

Country	Tanzania		Uganda	
RIF	Mean	*iqsr*	Mean	*iqsr*
Overall				
Male	3.340***	17.167***	3.998***	5.176***
Counterfactual	2.951***	27.568***	3.901***	6.821***
Female	2.396***	38.566***	3.448***	7.510***
Difference	0.944***	-21.400***	0.550***	-2.334***
Explained	0.389***	-10.401***	0.097**	-1.645***
Unexplained	0.555***	-10.998***	0.454***	-0.689
Pure explained	0.384***	-9.848***	0.096**	-1.538***
education	0.028***	-0.974***	-0.043**	-0.474***
covariates	0.356***	-8.874***	0.140***	-1.063***
Pure unexplained	0.549***	-11.369***	0.443***	-0.686
education	-0.115***	-17.536***	-0.256***	3.034*
covariates	0.236	37.625**	-0.299	10.289
constant	0.428**	-31.458*	0.998***	-14.010*
Specification error	0.005**	-0.553	0.001	-0.107
Reweight error	0.006	0.371	0.011	-0.003
Obs. Male	7,142	7,142	3,156	3,156
Obs. Female	4,073	4,073	1,475	1,475
Obs. Total	11,215	11,215	4,631	4,631

Notes: *iqsr* is the interquantile share ratio defined as the ratio of the share earned by the top decile relative to the bottom four deciles within each sex. P-values calculated from bootstrap (500 replications for Uganda and 2000 replications for Tanzania) standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean. Some significant specification and reweight errors detected, warranting a cautious inference.

in educational attainment and returns to education, do play a significant role in explaining the inequality differences across gender in both countries.

4.5.3 RIF Regression by Pay Period

In this subsection, we present estimates based on RIF regressions by pay period.. For each country and pay period, we examine the possible heterogeneous effects of education on earnings and how these effects vary when workers are paid over different periods. Like for the pooled analysis, we begin by presenting the results for linear specification (assuming $\alpha_2 = 0$ in equation (4.1). Tables 4.8, 4.9 and 4.10 present the results of the distribution effects of education in Malawi, Tanzania, and Uganda, respectively, for the selected unconditional quantiles of earnings when the samples are disaggregated by pay periods—daily, weekly, and monthly. The detailed results for all nine deciles by pay period are shown in Appendix 4A.

Table 4.8 shows the RIF regressions results for Malawi. It shows two important things—in Malawi education affects individuals at different points of the earnings distribution differently, and within the corresponding earnings distributions, education affects individuals in different pay periods differently. Since the latter is the focus of our study, it deserves more interpretation. The effect of a change in the distribution of education for the daily sample decreases with the quantiles of earnings for the quantiles above the centre of the distribution. The results in Table 4.8 suggest that an increase of the population education by one year increases the mean of wages by approximately 12% at the centre of the distribution compared to only approximately 4% increase at the top of the distribution. For the weekly sample, the effect increases monotonically from about 0.7% at the bottom end to 21% at the top end of distribution, while no specific pattern is observed for the monthly.

Figure 4.9 plots the coefficients of education from Table 4.8 and appendix Table 4A.. The curve for the pooled sample is included for comparison. The trend observed in curve for pooled is consistent with the monthly sample but not with the daily or weekly samples. This suggests that for countries where workers are employed primarily by the month converting all wages to a monthly figure will not distort the findings. However, in a country where workers are more likely to be paid by day or week this could lead to imprecise estimates.

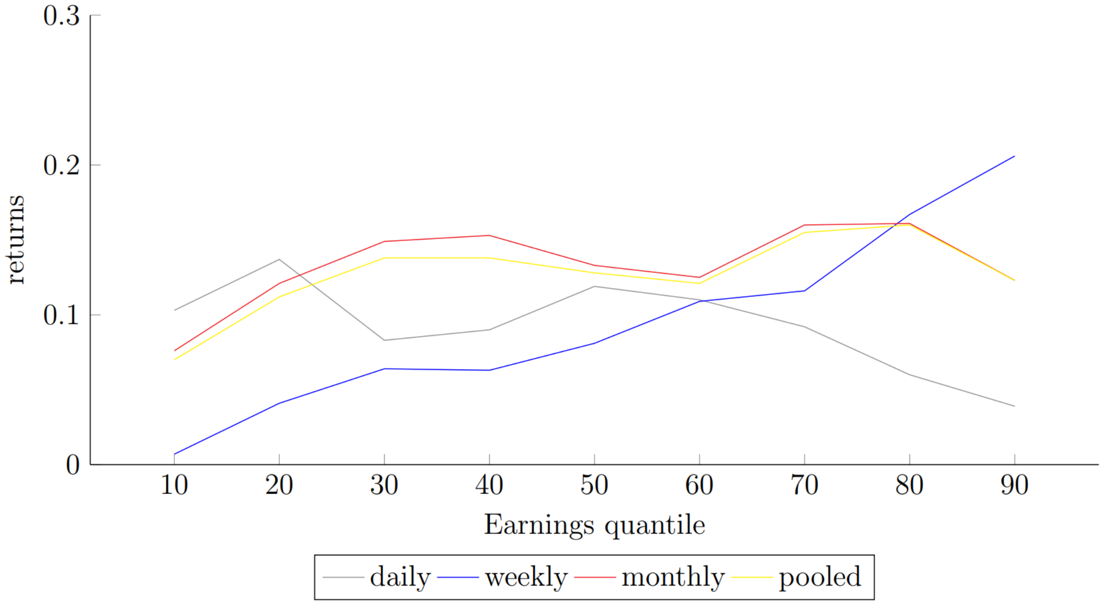
Table 4.8: Unconditional Quantile Regression by Pay Period - Malawi

quantile	(1) q(10)	(2) q(25)	(3) q(50)	(4) q(75)	(5) q(90)
A. Daily					
sch	0.103**	0.091**	0.119***	0.072**	0.039
	(0.043)	(0.043)	(0.042)	(0.037)	(0.042)
covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	2.19	3.205	4.583	5.618	6.298
R^2	0.25	0.33	0.35	0.16	0.11
Obs.	182	182	182	182	182
B. Weekly					
sch	0.007	0.058***	0.081***	0.151***	0.206***
	(0.021)	(0.018)	(0.016)	(0.023)	(0.040)
covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	2.946	3.602	4.592	5.365	6.243
R^2	0.50	0.44	0.38	0.30	0.23
Obs.	505	505	505	505	505
C. Monthly					
sch	0.076***	0.139***	0.133***	0.168***	0.123***
	(0.008)	(0.007)	(0.006)	(0.007)	(0.007)
covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	2.893	3.694	4.702	5.544	6.304
R^2	0.39	0.46	0.50	0.38	0.20
Obs.	5,129	5,129	5,129	5,129	5,129

Notes: Bootstrap standard errors computed by 500 replications in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Full results with all variables available upon request in appendix 4A.

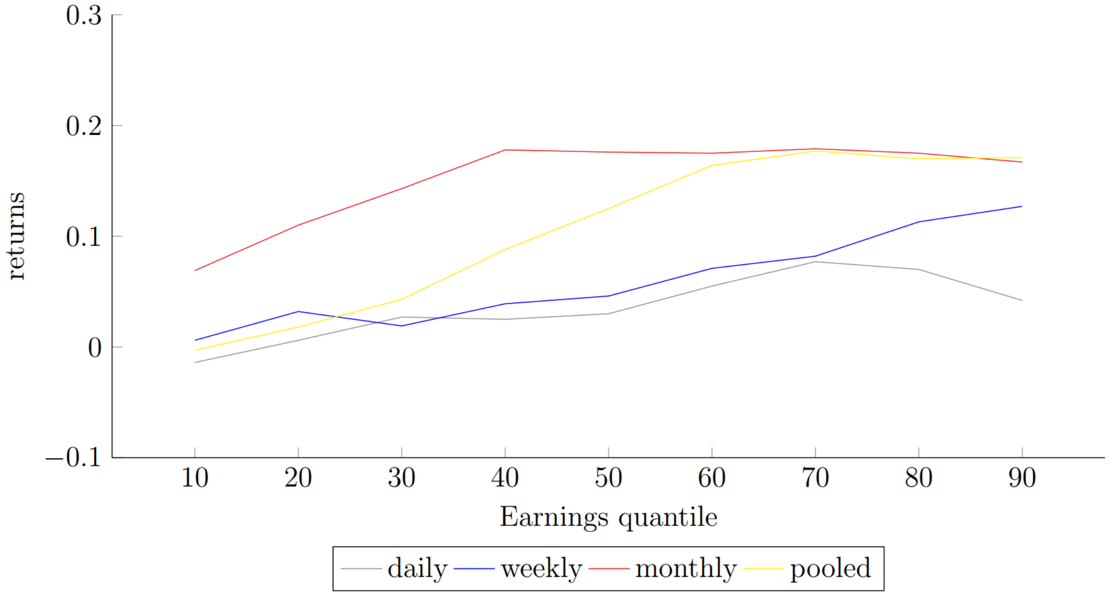
Table 4.9 presents the RIF regression results for Tanzania, illustrated in Figure 4.10 (with detailed results for all nine deciles in Appendix 4A). As for Malawi, Table 4.9 and Figure 4.10 show that in Tanzania, there is heterogeneity in the distributional effects of education across pay periods. Returns to monthly are much higher, increase up the 40^{th} quantile and are then flat. Generally, returns increase gently but with no significant effect of increase in education on wage for workers at the bottom of the distribution in the daily (peaking at the 70^{th} quantile) and weekly (returns increase sharply after 70^{th} quantile) samples. Comparing the patterns of the results from the pay periods to those from the pooled, it indicates

Figure 4.9: RIF Returns by Pay Period - Malawi



that the common practice in the literature of pooling all the pay periods together leads to imprecise estimates of the distributional effects of education.

Figure 4.10: RIF Returns by Pay Period - Tanzania



Finally, Table 4.10 reports the distribution effects of education in Uganda, illustrated in Figure 4.11 (with detailed results for all nine deciles in Appendix 4A). As for Malawi and Tanzania, in Uganda, there is also heterogeneity in the effects

Table 4.9: Unconditional Quantile Regression for Tanzania

quantile	(1) q(10)	(2) q(25)	(3) q(50)	(4) q(75)	(5) q(90)
A. Daily					
sch	-0.014	0.026**	0.030***	0.068***	0.042***
	(0.012)	(0.011)	(0.010)	(0.013)	(0.012)
covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	-0.274	0.789	2.091	3.854	4.922
R^2	0.19	0.41	0.53	0.44	0.20
Obs.	3,738	3,738	3,738	3,738	3,738
B. Weekly					
sch	0.006	0.030**	0.046***	0.091***	0.127***
	(0.016)	(0.012)	(0.011)	(0.016)	(0.019)
covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	-0.232	0.752	2.048	3.342	4.509
R^2	0.23	0.44	0.60	0.48	0.27
Obs.	1,929	1,929	1,929	1,929	1,929
C. Monthly					
sch	0.069***	0.119***	0.176***	0.177***	0.167***
	(0.010)	(0.008)	(0.007)	(0.008)	(0.009)
covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	2.137	3.178	4.352	5.203	5.838
R^2	0.45	0.51	0.44	0.29	0.19
Obs.	4,830	4,830	4,830	4,830	4,830

Notes: Bootstrap standard errors computed by 500 replications in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Full results with all variables available in appendix 4A.

of education across pay periods and throughout the earnings distribution. Table 4.10 and Figure 4.11 show that, along the earnings distribution, while education significantly increases earnings across the quantiles, the effect is higher at lower quantiles for monthly (peaking at 40^{th} quantile), generally declining for daily and gradually increasing for the weekly sample. This suggests that education is vital in reducing earnings inequality for workers paid daily and monthly but will increase inequality for workers paid weekly. On the other hand, comparing the pay periods, an additional year of education in the population increases the mean of earnings by a higher proportion for workers paid monthly relative to their daily and weekly counterparts.

Table 4.10: Unconditional Quantile Regression by Pay Period - Uganda

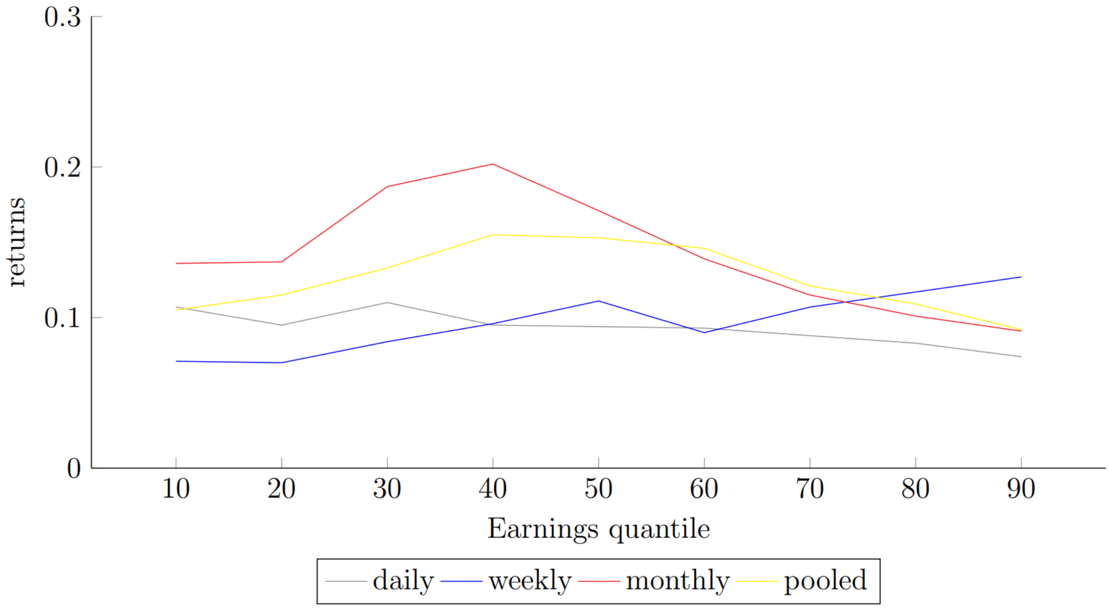
quantile	(1) q(10)	(2) q(25)	(3) q(50)	(4) q(75)	(5) q(90)
A. Daily					
sch	0.107***	0.114***	0.094***	0.090***	0.074***
	(0.022)	(0.015)	(0.012)	(0.013)	(0.015)
covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	1.786	2.774	3.719	4.58	5.27
R^2	0.36	0.39	0.31	0.21	0.11
Obs.	1,262	1,262	1,262	1,262	1,262
B. Weekly					
sch	0.071**	0.082***	0.111***	0.116***	0.127***
	(0.028)	(0.019)	(0.017)	(0.017)	(0.027)
covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	1.568	2.756	3.793	4.424	5.179
R^2	0.28	0.44	0.38	0.23	0.14
Obs.	589	589	589	589	589
C. Monthly					
sch	0.136***	0.156***	0.171***	0.109***	0.091***
	(0.016)	(0.010)	(0.009)	(0.006)	(0.007)
covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	2.16	3.242	4.35	5.059	5.544
R^2	0.45	0.45	0.44	0.27	0.15
Obs.	2,765	2,765	2,765	2,765	2,765

Notes: Bootstrap standard errors computed by 500 replications in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Full results with all variables available in appendix 4A.

Tables 4.8 – 4.10 show, in all pay periods, there is considerable difference in the RIF mean between the top and bottom deciles implying high degree of earnings inequality. Using RIF regression, we examine whether education is a significant determinant of earnings inequality, measured by the interquantile share ratio (iqsr) defined earlier. The results are presented in Table 4.11.

Table 4.11 shows that there is significant earnings inequality between the workers in the top decile of earnings and those in the bottom four deciles. Panel A shows that in Malawi the inequality is highest among workers paid daily and lowest among those paid monthly. Workers in the top decile earn approximately ten (10) times as much as those in the bottom four deciles. However, despite the

Figure 4.11: RIF Returns by Pay Period - Uganda



high inequality within workers paid daily, education does not seem to significantly drive the inequality. For workers paid weekly, an increase in education in the population by one year would result in an increase in earnings inequality by 9%,[9] other things equal. This implies that conditional on working and paid weekly, education is likely to benefit those in higher paying jobs. For workers paid monthly, an increase in average education in the population by a year will reduce inequality by 1.7%.

Panel B of Table 4.11 presents the results for Tanzania. Inequality is very high among workers in daily and weekly where those in the top decile earn at least 35 times as much as those in the bottom 4 deciles. Increase in education would worsen the inequality although not significantly for workers paid daily. In contrast, for workers paid monthly, education reduces inequality, but the coefficient is not statistically significant.

The bottom panel (Panel C) of Table 4.11 shows the results for Uganda. For workers paid daily and monthly, an increase in education in the population by one year would result in a reduction in wage inequality by 5% and 9%, respectively.

[9] That is (0.744/8.138)*100.

For those paid daily, increase in education would worsen the inequality but the coefficient is not statistically significant.

Table 4.11: Education and Wage Inequality (*iqsr*) by Country and Pay Period

	(1) Daily	(2) Weekly	(3) Monthly
A. Malawi			
sch	-0.615	0.744***	-0.126*
	(0.479)	(0.248)	(0.065)
covariates	Yes	Yes	Yes
iqsr	10.021	8.138	7.373
R^2	0.23	0.13	0.10
Obs.	182	505	5,129
B. Tanzania			
sch	0.295	2.669***	-0.011
	(0.514)	(0.835)	(0.083)
covariates	Yes	Yes	Yes
iqsr	36.546	35.055	6.799
R^2	0.13	0.04	0.32
Obs.	3,738	1,929	4,830
C. Uganda			
sch	-0.322***	0.259	-0.505***
	(0.112)	(0.264)	(0.074)
covariates	Yes	Yes	Yes
iqsr	6.037	7.101	5.389
R^2	0.15	0.16	0.20
Obs.	1,262	589	2,765

Notes: *iqsr* is the interquantile share ratio defined as the ratio of the share earned by the top 10% relative to bottom 40% within each pay period. Bootstrap standard errors computed by 500 replications in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Full results with all variables available upon request.

Figures 4.12 - 4.14 plot the marginal effects for the quadratic specification of returns against years of schooling for selected quantiles (10th, 50th and 90th); detailed results are Appendix 4B. Figures 4.12 - 4.14 show a concave relationship for workers in the bottom decile and strong convex relationship for workers in the top decile of earnings. This suggest that, in all three countries, an increase in education in the population is more likely to benefit the higher than the lower earnings workers and hence likely to increase inequality. For workers in the top decile, the effects of education are very small (even negative for Malawi) for early years of schooling but increase rapidly after about the 6th year, regardless of the pay period. Figures 4.12 - 4.14 also show that an increase in education in the population generates different earning outcomes depending on the pay period (and on the level of education).

4.5.4 RIF Decomposition by Pay Period

In this subsection, we assess the role of education in explaining gender earnings inequality by pay period. We begin by comparing the differences in means of earnings and educational attainment between male and female workers as well as gender differences in returns[10] to education. Table 4.12 shows the raw differences in the three variables by pay period and country. The second - fourth columns of Table 4.12 reveal that while there is a slight gender gap in earnings in favour of females paid daily and monthly in Malawi, females paid weekly earn significantly less than their male counterparts. In Tanzania and Uganda female workers have substantial lower wage earnings than their male counterparts across the pay periods. Nonetheless, the mean difference for Tanzania is more substantial for workers reporting daily earnings relative to their weekly and monthly counterparts.

Columns 5 - 7 of Table 4.12 shows that in all three countries there are significant differences in educational attainment between female and male workers. As expected, in most pay periods, female workers have lower educational attainment than their male counterparts. Columns 8 - 9 of Table 4.12, on the other hand, show that returns are higher for females except for workers paid weekly in Malawi and Tanzania, and those paid daily in Tanzania.

We then decompose the gender differences in the mean of log of earnings by country and pay period (see Appendix 4C for decomposition for quantiles of log earnings). Table 4.13 presents the results for the reweighted RIF OB gender decomposition by pay period for each of the three countries. Counterfactual is the estimated distribution of earnings, showing what would female mean wages

[10]Returns in 4.12 are based on AME(sch) coefficients from Tables 4C.1-4C.3 in Appendix 4C

Figure 4.12: RIF coefficients of education (quadratic) by Pay Period - Malawi

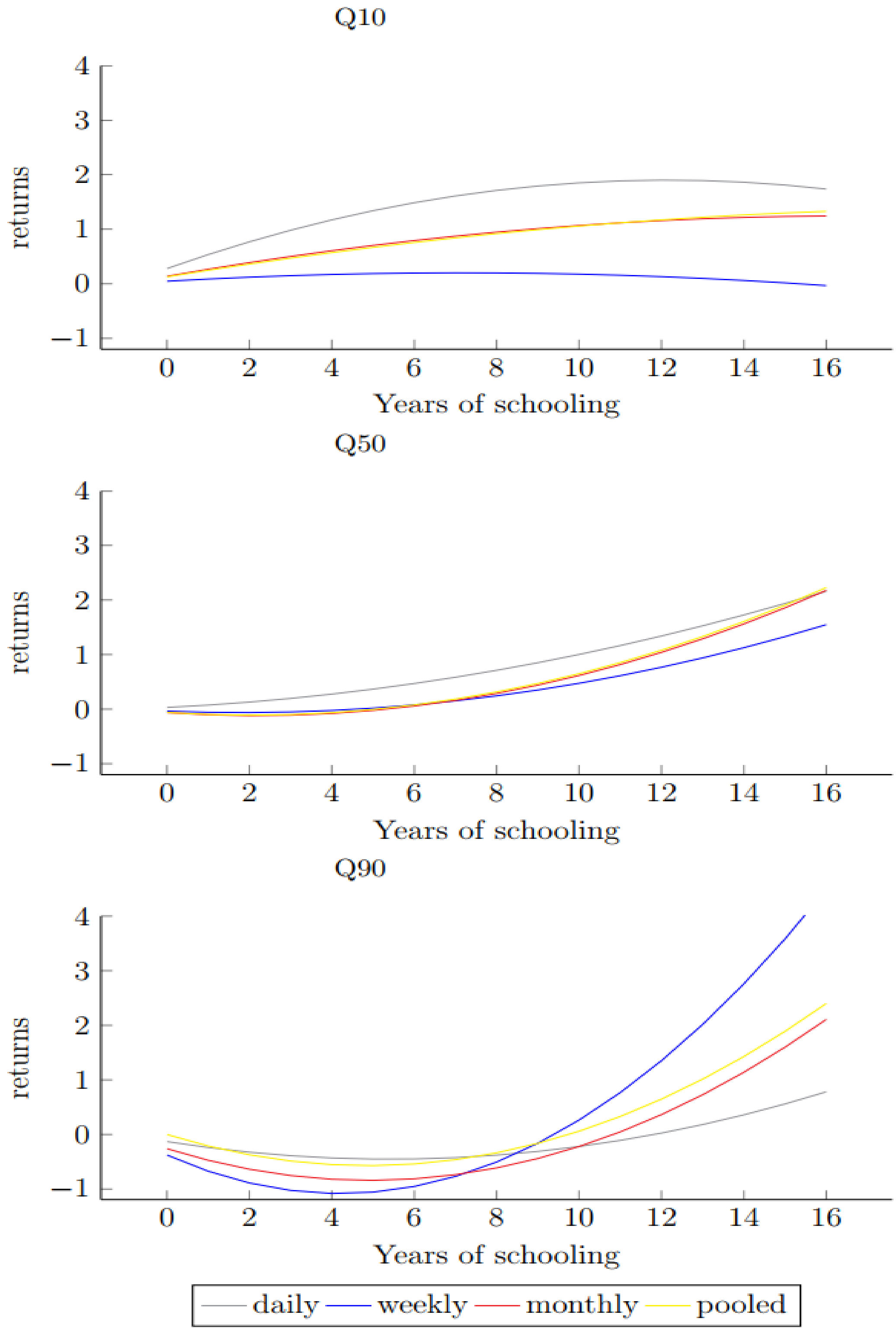


Figure 4.13: RIF coefficients of education (quadratic) by Pay Period - Tanzania

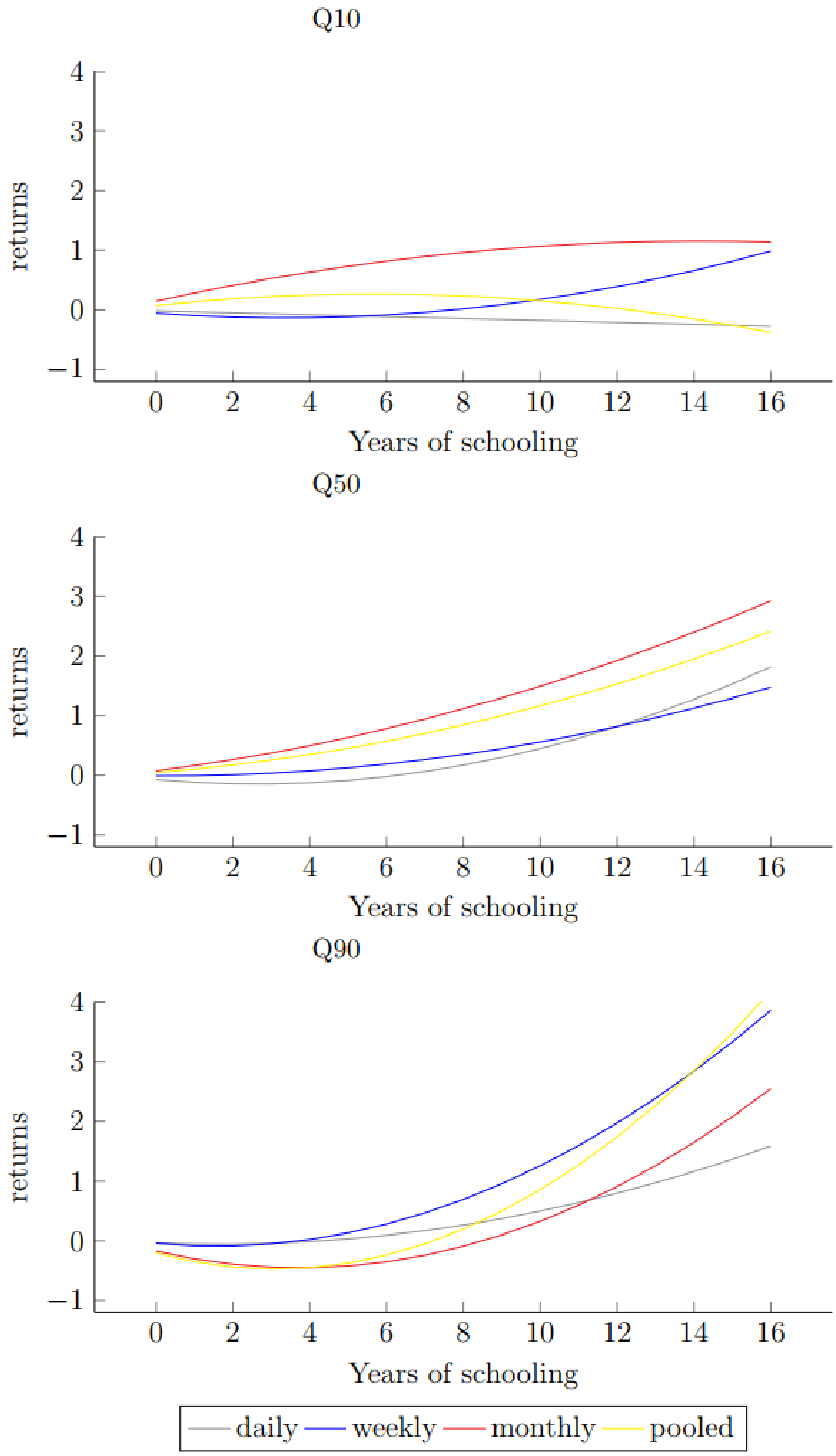


Figure 4.14: RIF coefficients of education (quadratic) by Pay Period - Uganda

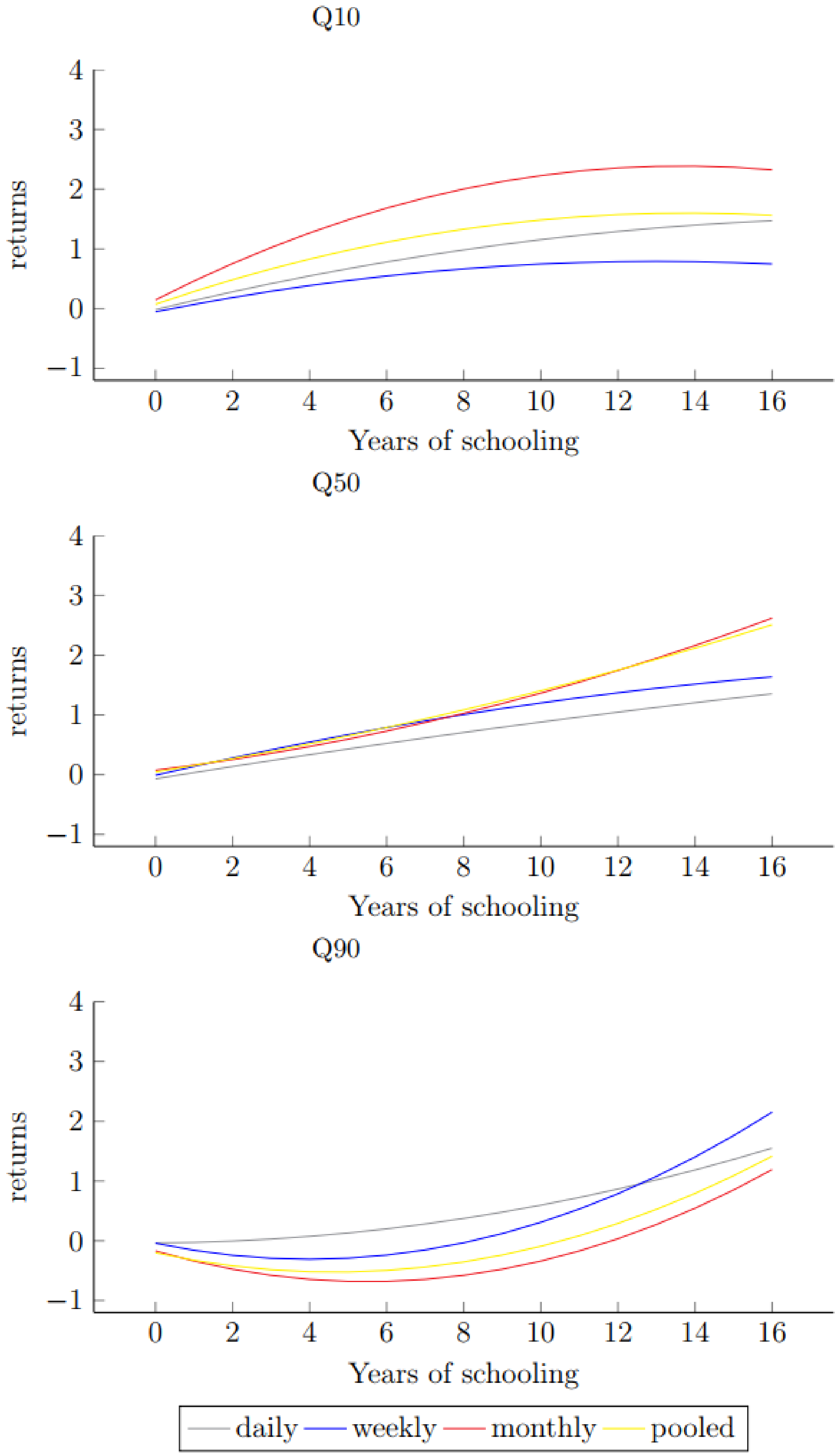


Table 4.12: Gender Differences in Earnings, Education and Returns to Education by Pay Period and Country

	Earnings (US$ per month)			Education (years)			Returns to Education		
	Daily	Weekly	Monthly	Daily	Weekly	Monthly	Daily	Weekly	Monthly
Malawi									
Male	77.68	100.02	99.15	9.48	7.45	9.13	0.091	0.115	0.148
Female	78.07	78.97	100.19	7.73	6.44	9.66	0.148	0.105	0.198
Difference	-0.39	21.05*	-1.04	1.75	1.01*	-0.53***	-0.057*	0.01	-0.05***
Obs. Male	133	361	3,864	133	361	3,864	133	361	3,864
Obs. Female	49	144	1,265	49	144	1,265	49	144	1,265
Obs. Total	182	505	5,129	182	505	5,129	182	505	5,129
Tanzania									
Male	15.69	11.37	74.13	5.79	5.83	8.18	0.054	0.077	0.133
Female	3.71	3.69	41.42	4.46	4.31	8.11	0.025	0.055	0.177
Difference	11.98***	7.68***	32.71***	1.34***	1.51***	0.07	0.029***	0.022**	-0.044***
Obs. Male	2,400	1,246	3,015	2,400	1,246	3,015	2,400	1,246	3,015
Obs. Female	1,338	683	1,815	1,338	683	1,815	1,338	683	1,815
Obs. Total	3,738	1,929	4,830	3,738	1,929	4,830	3,738	1,929	4,830
Uganda									
Male	43.98	41.83	65.22	6.59	7.21	9.69	0.085	0.091	0.14
Female	16.54	19.46	40.75	5.37	6.35	10.12	0.118	0.109	0.158
Difference	27.44***	22.37***	24.47***	1.22***	0.86*	-0.43*	-0.033***	-0.018	-0.018***
Obs. Male	981	422	1,743	981	422	1,743	981	422	1,743
Obs. Female	281	167	1,022	281	167	1,022	281	167	1,022
Obs. Total	1,262	589	2,765	1,262	589	2,765	1,262	589	2,765

Note: Difference for earnings is defined as geometric mean for males minus geometric mean for females. Difference for education is defined as arithmetic mean for males minus arithmetic mean for females. Difference for returns to education defined as the returns (AME(sch)) for males minus the corresponding value for females. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

be if they had the coefficients of male. Columns 2 - 4 of Table 4.13 reports the decomposition results for Malawi. Education has no significant effect on the gender difference in earnings for workers paid daily and weekly. These results also hold for the decomposition by quantiles of earnings (see appendix 4D). A significant proportion of the gender differences in earnings for workers paid monthly are attributable to gender differences in educational attainment. As discussed earlier, females paid monthly in Malawi have both more education and higher earnings, thus raising males educational endowment to the females' level would reduce the gender earnings gap.

Columns 5- 7 of Table 4.13 reports the decomposition results for Tanzania. As explained earlier, males in Tanzania earn higher than females across the pay periods. The results in suggest that education is among the significant factors explaining the gender wag gap in daily and weekly. Of the pure explained gender gap in earnings, differences in education explains approximately 7% and 14% for workers paid daily and weekly, respectively. This suggests that if females in daily and weekly had the same level of education endowments as males, their wage earnings would have been respectively 7% and 14% higher. In addition, the findings for Tanzania also show that gender differences in returns to education explains the earnings gap for workers paid weekly and monthly. Precisely, of the pure unexplained gender gap in earnings, difference in returns to education between males and females explains approximately 22% and 47% for those in weekly and monthly respectively. Note that for the monthly the coefficient on education for the unexplained part is negative implying that females have higher returns to education than males.

Columns 8 - 10 of Table 4.13 presents the corresponding results for Uganda. Like Tanzania, the mean wage for males in Uganda is higher than that of females across the pay periods, and gender earnings gap is wider for workers reporting daily earnings relative to their weekly and monthly counterparts. The findings show that while gender differences in education attainment play a significant role in explaining the gender earnings gap across the pay periods, gender differences in returns to education do not. Of the pure explained gender gap in earnings, differences in education explains approximately 37%, 38% and 33% for workers paid daily, weekly, and monthly respectively. This implies that if females in daily and weekly had the same level of education endowments as males, their wage earnings would have been more than a third higher. Because females paid monthly have higher educational attainment than their male counterparts, the coefficient is negative pointing out that if females had the same level of education as males then their wages would have been about a third lower.

Table 4.13: Reweighted RIF OB Gender Decomposition by Period and Country

	Malawi			Tanzania			Uganda		
	Daily	Weekly	Monthly	Daily	Weekly	Monthly	Daily	Weekly	Monthly
Overall									
Male	4.353***	4.605***	4.597***	2.753***	2.431***	4.306***	3.784***	3.734***	4.178***
Counterfactual	4.345***	4.478***	4.735***	1.959***	1.760***	4.088***	3.335***	3.491***	4.039***
Female	4.358***	4.369***	4.607***	1.312***	1.306***	3.724***	2.806***	2.968***	3.707***
Difference	-0.005	0.237*	-0.01	1.441***	1.125***	0.582***	0.978***	0.766***	0.470***
Explained	0.008	0.128	-0.138***	0.794***	0.671***	0.218***	0.449***	0.243**	0.138***
Unexplained	-0.013	0.109	0.128***	0.647***	0.454***	0.364***	0.529***	0.523***	0.332***
Pure explained	0.0000	0.128	-0.139***	0.810***	0.659***	0.215***	0.452***	0.235**	0.135***
education	0.018	-0.004	-0.112***	0.055***	0.094***	-0.008	0.167***	0.090**	-0.045*
covariates	-0.018	0.132	-0.026	0.755***	0.565***	0.223***	0.286***	0.145	0.181***
Pure unexplained	0.034	0.094	0.114***	0.726***	0.519***	0.340***	0.602***	0.548***	0.327***
education	-0.494	0.206	-0.118	0.034	0.112*	-0.159*	-0.086	-0.131	-0.148
covariates	-0.674	-1.17	0.027	0.872***	-0.037	0.125	1.596**	-0.475	-0.576
constant	1.202	1.058	0.205	-0.18	0.444	0.374	-0.908	1.154	1.051**
Specification error	0.008	0.000	0.001	-0.016**	0.012	0.003	-0.003	0.008	0.003
Reweight error	-0.047	0.015	0.014*	-0.079***	-0.064***	0.024**	-0.073	-0.025	0.005
Obs. Male	133	361	3,864	2,400	1,246	3,015	981	422	1,743
Obs. Female	49	144	1,265	1,338	683	1,815	281	167	1,022
Obs. Total	182	505	5,129	3,738	1,929	4,830	1,262	589	2,765

Note: The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean. P-values calculated from bootstrap (500 replications for Malawi and Uganda and 2000 replications for Tanzania) standard errors. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Some significant specification and reweight errors detected, warranting a cautious inference.

4.5.5 Distributional Effects of Education in *Ganyu* Labour

This subsection presents the analysis and results for *ganyu* labour. It follows the same approach employed in the main analysis. It begins by examining how an increase in education in the population affects the distribution of earnings at different unconditional quantiles of earnings. It then goes further to explore whether gender differences in education significantly explain the earnings gap between male and female workers in *ganyu* .

Table 4.14 reports the distribution effect of education for *ganyu* workers using both linear and quadratic specification of the earnings function. The coefficients from the linear specification shows that an increase in the population's average education by a year increases the mean wage of *ganyu* workers by 7 – 16% depending on the quantile of earnings distribution. Figure 4.15 shows that at the 10^{th} quantile the predicted returns increase at a constant rate across years of education, while at the 50^{th} and 90^{th} the predicted returns increase at increasing rates with years of education. At the 90^{th} quantile the predicted returns increase at a higher rate than the other quantiles implying that an increase in education is more likely to benefit the higher than the lower wage earners.

Table 4.14: Unconditional Quantile Regression (RIF) Results for *Ganyu*

	(1)	(2)	(3)	(4)	(5)
quantile	q(10)	q(25)	q(50)	q(75)	q(90)
Linear					
sch	0.022***	0.019***	0.008**	0.021***	0.026***
	(0.005)	(0.004)	(0.003)	(0.004)	(0.005)
Covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	0.523	1.537	2.569	3.544	4.423
R2	0.31	0.48	0.53	0.37	0.20
Obs.	16,528	16,528	16,528	16,528	16,528
Quadratic					
sch	0.019	0.002	-0.008	-0.001	-0.003
	(0.015)	(0.011)	(0.009)	(0.011)	(0.015)
sch2	0.000	0.002*	0.001*	0.002**	0.003*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Covariates	Yes	Yes	Yes	Yes	Yes
RIF mean	0.523	1.537	2.569	3.544	4.423
R^2	0.31	0.48	0.53	0.37	0.20
Obs.	16,528	16,528	16,528	16,528	16,528

Notes: Bootstrap standard errors computed by 500 replications in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Full results with all variables included in the RIF regression available upon request.

Tables 4.15 shows gender differences in earnings, education attainment and returns to education. Males have higher earnings, more schooling and higher returns to schooling. Table 4.16 reports the results for gender wage gap decomposition for workers in *ganyu* labour. Males in *ganyu* earn higher than females in both rural and urban areas. The results suggest that gender differences in education explain about 7% of the pure explained wage gap and the effect is slightly higher (8.5%) in urban areas. Since males are better endowed with education than females, the results imply that raising the female endowment of education to the male level would increase females' earnings and narrow the earnings gap in both rural and urban areas.

Figure 4.15: RIF coefficients of education (quadratic) - *Ganyu*

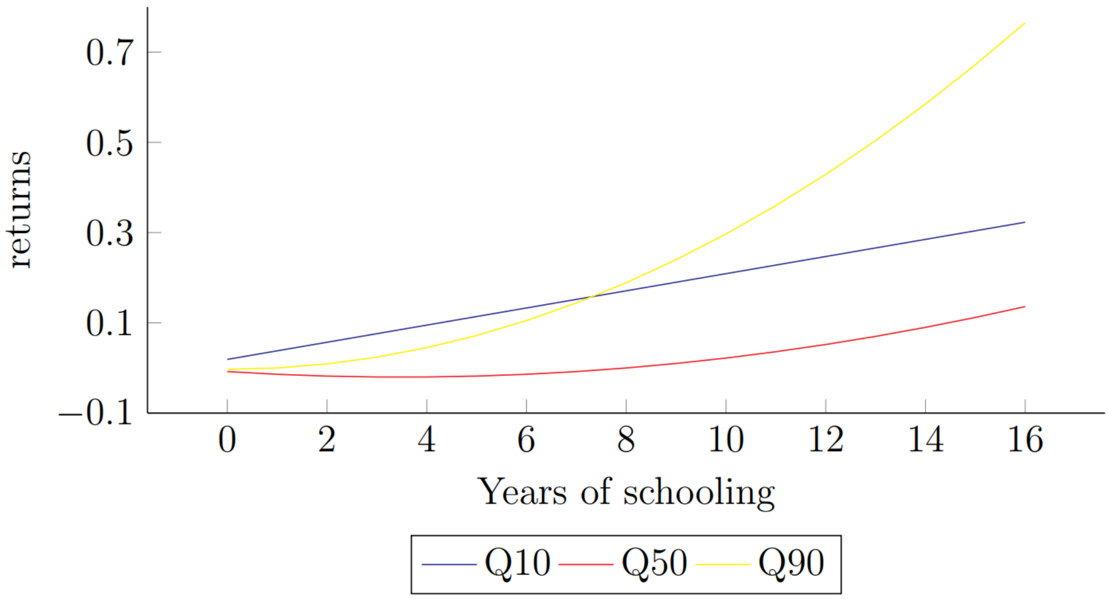


Table 4.15: Gender Differences in Earnings, Education Attainment and Returns to Education in *Ganyu*

	Earnings			Education (years)			Returns to Education		
	All	Rural	Urban	All	Rural	Urban	All	Rural	Urban
Male	16.74	15.51	37.65	5.51	5.33	7.3	0.022	0.019	0.053
Female	8.59	8.35	12.73	4.06	3.93	5.68	0.013	0.011	0.049
Difference	7.15***	6.16***	24.92***	1.45***	1.40***	1.62***	0.009***	0.008***	0.004
Obs. Male	8,282	7,570	712	8,282	7,570	712	8,282	7,570	712
Obs. Female	8,246	7,681	565	8,246	7,681	565	8,246	7,681	565
Obs. Total	16,528	15,259	1,277	16,528	15,259	1,277	16,528	15,259	1,277

Notes: Difference for earnings is the geometric mean for males minus geometric mean for females. Difference for education is arithmetic mean for males minus arithmetic mean for females. Difference for returns to education is the returns for males minus returns for females.* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4.16: Reweighted RIF OB Decomposition by Gender– *Ganyu*

	(1) All	(2) Rural	(3) Urban
Overall			
Male	2.898***	2.816***	3.713***
Counterfactual	2.528***	2.504***	2.865***
Female	2.250***	2.230***	2.507***
Difference	0.648***	0.586***	1.206***
Explained	0.370***	0.312***	0.848***
Unexplained	0.278***	0.275***	0.358**
Pure Explained	0.371***	0.312***	0.850***
education	0.027***	0.022***	0.073***
covariates	0.343***	0.289***	0.777***
Pure Unexplained	0.336***	0.321***	0.492***
education	0.043	0.036	0.170
covariates	0.364**	0.515***	0.589
constant	-0.07	-0.23	-0.268
Specification error	-0.001	0.000	-0.002
Reweight error	-0.058**	-0.046	-0.133
Obs. Male	8,282	7,570	712
Obs. Female	8,246	7,681	565
Obs. Total	16,528	15,259	1,277

Notes: P-values calculated from bootstrapped standard errors (500 replications). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All other OLS regressors included in the RIF regression.The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

4.6 Conclusion

Education is among the key factors that determine the levels of earnings among workers. But is the effect of education on earnings the same for low and high wage earners? Studies seeking to answer this question have mainly done so while aggregating various pay periods to a common period. This essay re-examined the relationship between education and earnings along the unconditional earnings distribution, taking into consideration the effects of pay period. Using nationally representative data from Malawi, Tanzania and Uganda and RIF regression techniques, we found that estimates significantly differ across the pay periods in all three countries. Generally, the effect of education is stronger for workers reporting monthly earnings compared to their daily and weekly counterparts, consistent with formal sector workers being more likely to be paid monthly.

Examination of the RIF means by the unconditional quantile of earnings revealed a considerable earnings inequality between low-wage and high-wage workers. We then investigated whether education is a significant factor that explains this wage inequality in each pay period. To measure wage inequality, we use the interquantile share ratio of the top decile of earnings to the bottom four (4) deciles. The findings from RIF regression reveal that education can either contribute to increasing or reducing wage inequality depending on the period in which the worker is paid. Education is found to increase inequality for workers paid monthly (suggesting higher wages for more skilled workers) and reduce inequality for those paid daily and monthly (perhaps because more educated workers are recent or temporary entrants with less on the job experience).

The essay investigated how much of the gender differences in earnings and inequality can be attributed to gender differences in educational attainment for each of the pay periods. Employing RIF OB decomposition method, we found that gender differences in education significantly explain the gender wage gap for workers paid daily and weekly for Tanzania, and in all pay periods for Uganda, while there was no significant gender wage gap for Malawi. This suggests that, for Tanzania and Uganda, policies targeting increasing female education attainment could narrow the gender wage gap. Further decomposition of the inequality within gender shows that inequality is higher among women compared to men, but the difference is mainly insignificant.

An extension examined the distributional effects of education on earnings for casual (*ganyu*) workers in Malawi. An increase in the population's average education by a year increases the mean wage of *ganyu* workers by 7 – 16% depending on the quantile of earnings distribution. RIF OB decomposition results further show that, in *ganyu* labour, the gender differences in education explain about 7% of the pure explained earnings gap and the effects is slightly higher in urban areas.

As discussed in the decomposition section, there are significant earnings differences between men and women in Tanzania and Uganda, across the pay periods with men generally earning more than women. For Malawi, gender gaps are significant only for ganyu workers. The discussion also shows that wages for women are more unequally distributed compared to that of males. Unfortunately, with the data in hand, we can not tell whether the wage penalties faced by women in Tanzania and Uganda can be accounted for by "glass ceilings" or "sticky floors". This deserves further exploration but is beyond the scope of this essay.

Appendices

Appendix 4A: RIF Regression by Quantile (Linear Schooling)

Table 4A.1: RIF Regression Results by Quantile of Earnings for Malawi (Daily)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	0.103**	0.137***	0.083*	0.090**	0.119***	0.110***	0.092**	0.060	0.039
	(0.043)	(0.045)	(0.043)	(0.043)	(0.042)	(0.039)	(0.036)	(0.037)	(0.042)
age	0.069	0.049	0.019	0.097	0.094	0.036	0.059	0.020	-0.032
	(0.117)	(0.112)	(0.137)	(0.133)	(0.109)	(0.103)	(0.091)	(0.087)	(0.083)
age2/100	-0.086	-0.054	-0.013	-0.085	-0.092	0.000	-0.027	-0.002	0.081
	(0.145)	(0.138)	(0.176)	(0.164)	(0.137)	(0.131)	(0.118)	(0.113)	(0.108)
female	-0.069	-0.039	0.227	0.601	-0.009	-0.091	-0.118	-0.178	-0.225
	(0.433)	(0.384)	(0.418)	(0.464)	(0.379)	(0.327)	(0.341)	(0.310)	(0.281)
rural	0.138	-0.233	-0.540	-0.644*	-0.664*	-0.353	-0.077	-0.320	-0.473
	(0.392)	(0.377)	(0.404)	(0.371)	(0.379)	(0.360)	(0.351)	(0.351)	(0.363)
year	0.483	0.077	1.163**	1.659***	1.433***	1.131***	0.821***	0.524*	0.456**
	(0.450)	(0.508)	(0.581)	(0.506)	(0.426)	(0.364)	(0.317)	(0.270)	(0.226)
weeks	1.863***	2.442***	2.195***	2.063***	1.655***	1.123***	0.789***	0.611***	0.298*
	(0.593)	(0.465)	(0.410)	(0.323)	(0.316)	(0.254)	(0.191)	(0.160)	(0.166)
RIF mean	2.19	2.785	3.509	3.998	4.583	5.026	5.497	5.782	6.298
R2	0.255	0.363	0.311	0.356	0.348	0.295	0.221	0.123	0.106
Obs.	182	182	182	182	182	182	182	182	182

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A.2: RIF Regression Results by Quantile of Earnings for Malawi (Weekly)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	0.007	0.041**	0.064***	0.063***	0.081***	0.109***	0.116***	0.167***	0.206***
	(0.021)	(0.020)	(0.016)	(0.016)	(0.016)	(0.018)	(0.023)	(0.028)	(0.040)
age	0.018	0.089*	0.143***	0.092**	0.094**	0.057	0.076**	0.055	0.055
	(0.057)	(0.046)	(0.048)	(0.044)	(0.040)	(0.038)	(0.037)	(0.040)	(0.052)
age2/100	-0.027	-0.116**	-0.171***	-0.110**	-0.101**	-0.039	-0.069	-0.037	-0.012
	(0.068)	(0.059)	(0.061)	(0.055)	(0.051)	(0.049)	(0.048)	(0.053)	(0.070)
female	-0.172	-0.058	-0.008	-0.062	-0.123	-0.140	-0.053	-0.012	0.234
	(0.169)	(0.161)	(0.145)	(0.140)	(0.133)	(0.133)	(0.143)	(0.155)	(0.233)
rural	-0.100	-0.473***	-0.664***	-0.525***	-0.456***	-0.497***	-0.353**	-0.183	-0.014
	(0.171)	(0.161)	(0.151)	(0.162)	(0.144)	(0.142)	(0.156)	(0.174)	(0.277)
year	0.698***	1.319***	1.600***	1.419***	1.243***	0.968***	0.973***	0.842***	1.098***
	(0.172)	(0.170)	(0.158)	(0.186)	(0.165)	(0.145)	(0.150)	(0.157)	(0.212)
weeks	2.088***	1.512***	1.179***	1.106***	0.970***	0.779***	0.678***	0.478***	0.409***
	(0.342)	(0.154)	(0.132)	(0.106)	(0.098)	(0.092)	(0.085)	(0.086)	(0.101)
RIF mean	2.946	3.395	3.867	4.296	4.592	4.942	5.255	5.608	6.243
R^2	0.50	0.42	0.44	0.40	0.38	0.36	0.31	0.27	0.23
Obs.	505	505	505	505	505	505	505	505	505

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A.3: RIF Regression Results by Quantile of Earnings for Malawi (Monthly)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	0.007	0.041**	0.064***	0.063***	0.081***	0.109***	0.116***	0.167***	0.206***
	(0.021)	(0.020)	(0.016)	(0.016)	(0.016)	(0.018)	(0.023)	(0.028)	(0.040)
age	0.018	0.089*	0.143***	0.092**	0.094**	0.057	0.076**	0.055	0.055
	(0.057)	(0.046)	(0.048)	(0.044)	(0.040)	(0.038)	(0.037)	(0.040)	(0.052)
age2/100	-0.027	-0.116**	-0.171***	-0.110**	-0.101**	-0.039	-0.069	-0.037	-0.012
	(0.068)	(0.059)	(0.061)	(0.055)	(0.051)	(0.049)	(0.048)	(0.053)	(0.070)
female	-0.172	-0.058	-0.008	-0.062	-0.123	-0.140	-0.053	-0.012	0.234
	(0.169)	(0.161)	(0.145)	(0.140)	(0.133)	(0.133)	(0.143)	(0.155)	(0.233)
rural	-0.100	-0.473***	-0.664***	-0.525***	-0.456***	-0.497***	-0.353**	-0.183	-0.014
	(0.171)	(0.161)	(0.151)	(0.162)	(0.144)	(0.142)	(0.156)	(0.174)	(0.277)
year	0.698***	1.319***	1.600***	1.419***	1.243***	0.968***	0.973***	0.842***	1.098***
	(0.172)	(0.170)	(0.158)	(0.186)	(0.165)	(0.145)	(0.150)	(0.157)	(0.212)
weeks	2.088***	1.512***	1.179***	1.106***	0.970***	0.779***	0.678***	0.478***	0.409***
	(0.342)	(0.154)	(0.132)	(0.106)	(0.098)	(0.092)	(0.085)	(0.086)	(0.101)
RIF mean	2.946	3.395	3.867	4.296	4.592	4.942	5.255	5.608	6.243
R^2	0.50	0.42	0.44	0.40	0.38	0.36	0.31	0.27	0.23
Obs.	505	505	505	505	505	505	505	505	505

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A.4: RIF Regression Results by Quantile of Earnings for Malawi (Pooled)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	0.070***	0.112***	0.138***	0.138***	0.128***	0.121***	0.155***	0.160***	0.123***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	(0.006)	(0.007)	(0.007)	(0.006)
age	0.052***	0.076***	0.097***	0.093***	0.070***	0.059***	0.057***	0.045***	0.041***
	(0.018)	(0.015)	(0.013)	(0.012)	(0.010)	(0.010)	(0.011)	(0.012)	(0.012)
age2/100	-0.057***	-0.084***	-0.098***	-0.090***	-0.063***	-0.047***	-0.044***	-0.029*	-0.028*
	(0.022)	(0.019)	(0.017)	(0.015)	(0.012)	(0.013)	(0.014)	(0.015)	(0.016)
female	-0.153***	-0.150***	-0.084*	-0.013	0.028	-0.076*	-0.102**	0.029	-0.026
	(0.057)	(0.046)	(0.044)	(0.046)	(0.040)	(0.039)	(0.045)	(0.052)	(0.053)
rural	-0.276***	-0.320***	-0.233***	-0.178***	-0.139***	-0.249***	-0.196***	-0.175***	-0.249***
	(0.047)	(0.043)	(0.043)	(0.047)	(0.034)	(0.033)	(0.039)	(0.044)	(0.059)
year	0.933***	1.175***	1.419***	1.419***	1.425***	1.485***	1.526***	1.389***	0.952***
	(0.054)	(0.049)	(0.054)	(0.055)	(0.049)	(0.053)	(0.060)	(0.061)	(0.056)
weeks	2.427***	1.720***	1.517***	1.297***	0.965***	0.798***	0.646***	0.528***	0.313***
	(0.142)	(0.072)	(0.063)	(0.052)	(0.035)	(0.035)	(0.036)	(0.035)	(0.026)
RIF mean	2.841	3.427	3.876	4.354	4.692	4.995	5.36	5.797	6.298
R^2	0.39	0.42	0.47	0.48	0.48	0.46	0.40	0.32	0.19
Obs.	5816	5816	5816	5816	5816	5816	5816	5816	5816

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A.5: RIF Regression Results by Quantile of Earnings for Malawi (*Ganyu*)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	0.022***	0.017***	0.019***	0.011***	0.008**	0.013***	0.019***	0.026***	0.026***
	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
age	0.027***	0.040***	0.048***	0.046***	0.048***	0.050***	0.044***	0.047***	0.044***
	(0.009)	(0.007)	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.007)
age2/100	-0.030***	-0.047***	-0.055***	-0.056***	-0.059***	-0.061***	-0.053***	-0.054***	-0.052***
	(0.012)	(0.009)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.008)	(0.009)
female	-0.126***	-0.176***	-0.266***	-0.325***	-0.392***	-0.437***	-0.467***	-0.513***	-0.523***
	(0.038)	(0.028)	(0.029)	(0.024)	(0.024)	(0.025)	(0.025)	(0.028)	(0.035)
rural	0.019	-0.06	-0.086**	-0.161***	-0.295***	-0.365***	-0.473***	-0.624***	-0.847***
	(0.060)	(0.048)	(0.042)	(0.041)	(0.042)	(0.042)	(0.048)	(0.062)	(0.085)
year	1.109***	1.375***	1.520***	1.605***	1.584***	1.393***	1.263***	1.144***	0.808***
	(0.047)	(0.042)	(0.039)	(0.043)	(0.038)	(0.044)	(0.029)	(0.033)	(0.030)
weeks	1.324***	1.215***	1.156***	1.112***	1.028***	0.964***	0.939***	0.917***	0.745***
	(0.047)	(0.032)	(0.025)	(0.022)	(0.023)	(0.019)	(0.019)	(0.021)	(0.023)
RIF mean	0.523	1.277	1.74	2.192	2.569	2.942	3.371	3.834	4.423
R^2	0.31	0.45	0.52	0.54	0.53	0.49	0.43	0.35	0.20
Obs.	16528	16528	16528	16528	16528	16528	16528	16528	16528

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A.6: RIF Regression Results by Quantile of Earnings for Tanzania (Daily)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	-0.014	0.006	0.027***	0.025**	0.030***	0.055***	0.077***	0.070***	0.042***
	(0.012)	(0.012)	(0.010)	(0.011)	(0.010)	(0.012)	(0.013)	(0.012)	(0.012)
age	0.035	0.025	0.057***	0.057***	0.075***	0.079***	0.076***	0.071***	0.056***
	(0.022)	(0.021)	(0.016)	(0.016)	(0.017)	(0.020)	(0.019)	(0.017)	(0.016)
age2/100	-0.033	-0.03	-0.069***	-0.074***	-0.101***	-0.101***	-0.094***	-0.085***	-0.064***
	(0.029)	(0.028)	(0.023)	(0.022)	(0.023)	(0.026)	(0.025)	(0.023)	(0.021)
female	-0.384***	-0.407***	-0.613***	-0.815***	-1.012***	-1.153***	-1.189***	-0.991***	-0.633***
	(0.130)	(0.102)	(0.073)	(0.076)	(0.080)	(0.089)	(0.095)	(0.086)	(0.062)
rural	0.037	0.051	-0.162**	-0.321***	-0.637***	-0.999***	-1.327***	-1.085***	-0.698***
	(0.094)	(0.080)	(0.068)	(0.068)	(0.080)	(0.102)	(0.131)	(0.132)	(0.119)
panel	-0.243**	-0.178**	-0.214***	-0.275***	-0.304***	-0.366***	-0.183*	-0.073	-0.141
	(0.100)	(0.086)	(0.077)	(0.073)	(0.077)	(0.091)	(0.098)	(0.100)	(0.097)
weeks	0.894***	1.339***	1.333***	1.347***	1.440***	1.553***	1.550***	1.142***	0.680***
	(0.111)	(0.059)	(0.052)	(0.047)	(0.054)	(0.059)	(0.061)	(0.053)	(0.033)
RIF mean	-0.274	0.500	1.129	1.631	2.091	2.812	3.475	4.194	4.922
R^2	0.19	0.37	0.46	0.50	0.53	0.52	0.49	0.38	0.20
Obs.	3738	3738	3738	3738	3738	3738	3738	3738	3738

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A.7: RIF Regression Results by Quantile of Earnings for Tanzania (Weekly)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	0.006	0.032**	0.019	0.039***	0.046***	0.071***	0.082***	0.113***	0.127***
	(0.016)	(0.013)	(0.013)	(0.013)	(0.011)	(0.013)	(0.015)	(0.016)	(0.019)
age	0.048*	0.049**	0.036*	0.039**	0.040**	0.050***	0.056***	0.039*	0.031
	(0.028)	(0.021)	(0.019)	(0.019)	(0.018)	(0.018)	(0.019)	(0.023)	(0.028)
age2/100	-0.056	-0.053*	-0.04	-0.045*	-0.048*	-0.062**	-0.069***	-0.045	-0.028
	(0.038)	(0.028)	(0.026)	(0.025)	(0.025)	(0.025)	(0.025)	(0.030)	(0.038)
female	-0.348***	-0.315***	-0.429***	-0.591***	-0.714***	-0.598***	-0.550***	-0.746***	-0.783***
	(0.107)	(0.101)	(0.092)	(0.078)	(0.088)	(0.091)	(0.096)	(0.097)	(0.103)
rural	0.698***	0.469***	0.380***	0.200**	-0.135	-0.488***	-0.886***	-1.166***	-1.554***
	(0.124)	(0.092)	(0.087)	(0.093)	(0.088)	(0.109)	(0.150)	(0.170)	(0.235)
panel	-0.003	0.057	-0.024	0.020	-0.119	-0.142	-0.126	-0.444***	-0.430***
	(0.114)	(0.104)	(0.099)	(0.100)	(0.097)	(0.100)	(0.117)	(0.128)	(0.154)
weeks	0.884***	1.042***	1.246***	1.348***	1.456***	1.360***	1.273***	1.068***	0.907***
	(0.060)	(0.061)	(0.057)	(0.063)	(0.057)	(0.061)	(0.059)	(0.061)	(0.072)
RIF mean	-0.232	0.457	0.999	1.526	2.048	2.503	2.948	3.642	4.509
R^2	0.23	0.38	0.51	0.57	0.60	0.57	0.53	0.41	0.27
Obs.	1929	1929	1929	1929	1929	1929	1929	1929	1929

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A.8: RIF Regression Results by Quantile of Earnings for Tanzania (Monthly)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	0.069***	0.110***	0.143***	0.178***	0.176***	0.175***	0.179***	0.175***	0.167***
	(0.010)	(0.009)	(0.008)	(0.009)	(0.007)	(0.007)	(0.007)	(0.008)	(0.009)
age	0.143***	0.181***	0.187***	0.160***	0.116***	0.079***	0.055***	0.024**	0.007
	(0.023)	(0.016)	(0.013)	(0.014)	(0.011)	(0.010)	(0.009)	(0.009)	(0.011)
age2/100	-0.177***	-0.205***	-0.196***	-0.155***	-0.103***	-0.058***	-0.031**	0.008	0.030*
	(0.027)	(0.019)	(0.016)	(0.017)	(0.013)	(0.013)	(0.012)	(0.013)	(0.016)
female	-0.218***	-0.298***	-0.511***	-0.547***	-0.455***	-0.351***	-0.261***	-0.163***	-0.086*
	(0.075)	(0.070)	(0.057)	(0.053)	(0.046)	(0.044)	(0.044)	(0.045)	(0.050)
rural	-0.260***	-0.391***	-0.349***	-0.393***	-0.317***	-0.224***	-0.187***	-0.175***	-0.185***
	(0.066)	(0.056)	(0.051)	(0.047)	(0.041)	(0.037)	(0.040)	(0.042)	(0.046)
panel	-0.013	0.037	0.022	0.118**	-0.029	-0.151***	-0.205***	-0.202***	-0.264***
	(0.060)	(0.052)	(0.048)	(0.051)	(0.044)	(0.043)	(0.045)	(0.048)	(0.056)
weeks	2.482***	1.797***	1.510***	1.175***	0.848***	0.607***	0.455***	0.300***	0.155***
	(0.155)	(0.087)	(0.063)	(0.053)	(0.033)	(0.025)	(0.023)	(0.022)	(0.021)
RIF mean	2.137	2.863	3.465	3.853	4.352	4.701	5.021	5.368	5.838
R^2	0.45	0.48	0.50	0.48	0.44	0.38	0.32	0.26	0.19
Obs.	4830	4830	4830	4830	4830	4830	4830	4830	4830

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A.9: RIF Regression Results by Quantile of Earnings for Tanzania (Pooled)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	-0.003	0.018***	0.043***	0.088***	0.125***	0.164***	0.177***	0.170***	0.171***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.008)	(0.006)	(0.006)	(0.006)	(0.007)
age	0.018	0.040***	0.064***	0.102***	0.125***	0.126***	0.100***	0.070***	0.034***
	(0.014)	(0.011)	(0.010)	(0.010)	(0.010)	(0.010)	(0.009)	(0.008)	(0.008)
age2/100	-0.030	-0.057***	-0.088***	-0.123***	-0.139***	-0.124***	-0.088***	-0.051***	-0.003
	(0.018)	(0.015)	(0.014)	(0.014)	(0.013)	(0.012)	(0.011)	(0.011)	(0.011)
female	-0.477***	-0.668***	-0.676***	-0.691***	-0.709***	-0.724***	-0.587***	-0.369***	-0.184***
	(0.059)	(0.061)	(0.044)	(0.049)	(0.045)	(0.043)	(0.037)	(0.039)	(0.035)
rural	0.124**	-0.100**	-0.302***	-0.532***	-0.700***	-0.695***	-0.581***	-0.361***	-0.286***
	(0.049)	(0.045)	(0.040)	(0.040)	(0.046)	(0.051)	(0.046)	(0.045)	(0.047)
panel	-0.138**	-0.117***	-0.043	-0.071*	0.003	0.021	-0.076**	-0.153***	-0.183***
	(0.054)	(0.044)	(0.041)	(0.038)	(0.041)	(0.039)	(0.037)	(0.036)	(0.040)
weeks	1.405***	1.712***	1.826***	1.659***	1.493***	1.154***	0.823***	0.509***	0.307***
	(0.054)	(0.046)	(0.045)	(0.039)	(0.041)	(0.028)	(0.023)	(0.015)	(0.014)
RIF mean	0.307	1.155	1.88	2.572	3.199	3.745	4.380	4.836	5.461
R^2	0.30	0.50	0.58	0.60	0.58	0.51	0.43	0.31	0.20
Obs.	11215	11215	11215	11215	11215	11215	11215	11215	11215

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A.10: RIF Regression Results by Quantile of Earnings for Uganda (Daily)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	0.107***	0.095***	0.110***	0.095***	0.094***	0.093***	0.088***	0.083***	0.074***
	(0.022)	(0.017)	(0.014)	(0.013)	(0.012)	(0.013)	(0.013)	(0.013)	(0.015)
age	-0.027	0.028	0.069***	0.105***	0.104***	0.095***	0.095***	0.116***	0.096***
	(0.041)	(0.028)	(0.025)	(0.025)	(0.023)	(0.020)	(0.020)	(0.020)	(0.020)
age2/100	0.033	-0.042	-0.090***	-0.142***	-0.135***	-0.123***	-0.121***	-0.143***	-0.115***
	(0.056)	(0.039)	(0.034)	(0.034)	(0.032)	(0.028)	(0.027)	(0.026)	(0.027)
female	-0.672***	-0.735***	-0.652***	-0.767***	-0.723***	-0.747***	-0.601***	-0.579***	-0.452***
	(0.215)	(0.151)	(0.117)	(0.112)	(0.103)	(0.100)	(0.094)	(0.087)	(0.088)
rural	-0.213	-0.239**	-0.283***	-0.334***	-0.416***	-0.422***	-0.337***	-0.264**	-0.360***
	(0.139)	(0.094)	(0.087)	(0.087)	(0.091)	(0.088)	(0.098)	(0.103)	(0.115)
panel	0.046	0.034	-0.080	-0.084	-0.090	-0.079	0.019	0.096	0.075
	(0.135)	(0.098)	(0.088)	(0.093)	(0.091)	(0.091)	(0.093)	(0.102)	(0.113)
weeks	2.388***	1.812***	1.394***	1.154***	0.989***	0.855***	0.705***	0.573***	0.421***
	(0.236)	(0.136)	(0.110)	(0.089)	(0.068)	(0.059)	(0.053)	(0.049)	(0.046)
RIF mean	1.786	2.543	3.029	3.348	3.719	4.104	4.389	4.801	5.270
R^2	0.36	0.41	0.38	0.35	0.31	0.28	0.22	0.18	0.11
Obs.	1262	1262	1262	1262	1262	1262	1262	1262	1262

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A.11: RIF Regression Results by Quantile of Earnings for Uganda (Weekly)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	0.071**	0.070***	0.084***	0.096***	0.111***	0.090***	0.107***	0.117***	0.127***
	(0.028)	(0.023)	(0.018)	(0.017)	(0.017)	(0.016)	(0.018)	(0.019)	(0.027)
age	0.018	0.042	0.069*	0.101***	0.105***	0.062**	0.110***	0.124***	0.120***
	(0.059)	(0.044)	(0.035)	(0.035)	(0.032)	(0.028)	(0.028)	(0.024)	(0.029)
age2/100	-0.01	-0.059	-0.088*	-0.127***	-0.133***	-0.071*	-0.125***	-0.145***	-0.133***
	(0.075)	(0.059)	(0.048)	(0.046)	(0.044)	(0.037)	(0.037)	(0.032)	(0.037)
female	-0.707**	-0.713***	-0.828***	-0.634***	-0.716***	-0.591***	-0.479***	-0.417***	-0.169
	(0.280)	(0.194)	(0.170)	(0.156)	(0.132)	(0.129)	(0.132)	(0.131)	(0.152)
rural	-0.482***	-0.614***	-0.687***	-0.528***	-0.405***	-0.388***	-0.299**	-0.314*	-0.225
	(0.178)	(0.140)	(0.134)	(0.145)	(0.135)	(0.138)	(0.146)	(0.164)	(0.196)
panel	0.217	0.153	-0.076	-0.044	-0.148	-0.124	-0.089	-0.137	-0.318*
	(0.220)	(0.182)	(0.173)	(0.166)	(0.146)	(0.138)	(0.140)	(0.172)	(0.180)
weeks	1.850***	2.123***	1.464***	1.192***	0.978***	0.810***	0.541***	0.517***	0.376***
	(0.336)	(0.246)	(0.156)	(0.111)	(0.096)	(0.071)	(0.074)	(0.064)	(0.075)
RIF mean	1.568	2.490	3.031	3.416	3.793	4.125	4.282	4.739	5.179
R^2	0.28	0.46	0.44	0.38	0.38	0.30	0.26	0.22	0.14
Obs.	589	589	589	589	589	589	589	589	589

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A.12: RIF Regression Results by Quantile of Earnings for Uganda (Monthly)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	0.136***	0.137***	0.187***	0.202***	0.171***	0.139***	0.115***	0.101***	0.091***
	(0.016)	(0.010)	(0.011)	(0.010)	(0.009)	(0.006)	(0.007)	(0.006)	(0.007)
age	0.125***	0.121***	0.141***	0.138***	0.124***	0.085***	0.040***	0.012	0.000
	(0.032)	(0.024)	(0.020)	(0.019)	(0.015)	(0.013)	(0.012)	(0.011)	(0.014)
age2/100	-0.156***	-0.140***	-0.157***	-0.153***	-0.132***	-0.080***	-0.025	0.004	0.020
	(0.040)	(0.030)	(0.025)	(0.024)	(0.020)	(0.018)	(0.016)	(0.015)	(0.019)
female	-0.386***	-0.321***	-0.336***	-0.416***	-0.315***	-0.253***	-0.272***	-0.280***	-0.266***
	(0.101)	(0.064)	(0.065)	(0.063)	(0.053)	(0.048)	(0.047)	(0.044)	(0.055)
rural	0.014	-0.094	-0.189***	-0.251***	-0.214***	-0.269***	-0.313***	-0.390***	-0.486***
	(0.098)	(0.063)	(0.065)	(0.065)	(0.055)	(0.048)	(0.046)	(0.047)	(0.056)
panel	0.321***	0.377***	0.323***	0.279***	0.252***	0.126**	0.037	-0.019	-0.06
	(0.108)	(0.079)	(0.077)	(0.075)	(0.060)	(0.053)	(0.050)	(0.050)	(0.059)
weeks	3.191***	1.724***	1.350***	1.026***	0.699***	0.499***	0.365***	0.303***	0.230***
	(0.279)	(0.111)	(0.080)	(0.062)	(0.047)	(0.040)	(0.034)	(0.028)	(0.028)
RIF mean	2.16	2.983	3.43	3.941	4.35	4.659	4.861	5.161	5.544
R^2	0.45	0.45	0.45	0.45	0.44	0.38	0.30	0.23	0.15
Obs.	2765	2765	2765	2765	2765	2765	2765	2765	2765

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4A.13: RIF Regression Results by Quantile of Earnings for Uganda (Pooled)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
sch	0.105***	0.115***	0.133***	0.155***	0.153***	0.146***	0.121***	0.109***	0.092***
	(0.010)	(0.007)	(0.006)	(0.008)	(0.007)	(0.006)	(0.005)	(0.005)	(0.006)
age	0.063**	0.064***	0.101***	0.119***	0.107***	0.101***	0.072***	0.054***	0.024**
	(0.026)	(0.017)	(0.014)	(0.014)	(0.013)	(0.011)	(0.009)	(0.009)	(0.010)
age2/100	-0.087***	-0.083***	-0.124***	-0.141***	-0.122***	-0.109***	-0.071***	-0.045***	-0.013
	(0.034)	(0.022)	(0.019)	(0.018)	(0.017)	(0.015)	(0.012)	(0.012)	(0.014)
female	-0.643***	-0.549***	-0.515***	-0.580***	-0.510***	-0.388***	-0.322***	-0.309***	-0.279***
	(0.096)	(0.063)	(0.053)	(0.050)	(0.046)	(0.044)	(0.038)	(0.038)	(0.043)
rural	-0.134*	-0.190***	-0.208***	-0.325***	-0.256***	-0.254***	-0.290***	-0.306***	-0.362***
	(0.076)	(0.050)	(0.049)	(0.050)	(0.046)	(0.042)	(0.041)	(0.042)	(0.049)
panel	0.340***	0.252***	0.215***	0.193***	0.182***	0.190***	0.115***	0.052	0.011
	(0.083)	(0.062)	(0.052)	(0.053)	(0.049)	(0.046)	(0.039)	(0.041)	(0.044)
weeks	2.978***	1.929***	1.381***	1.125***	0.855***	0.602***	0.443***	0.345***	0.260***
	(0.184)	(0.090)	(0.056)	(0.047)	(0.038)	(0.041)	(0.026)	(0.022)	(0.020)
RIF mean	1.931	2.77	3.251	3.678	4.055	4.375	4.705	5.054	5.446
R^2	0.41	0.44	0.41	0.40	0.38	0.36	0.29	0.24	0.14
Obs.	4631	4631	4631	4631	4631	4631	4631	4631	4631

Notes: Regression includes a constant. Bootstrap standard errors (computed by 500 replications) in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix 4B: RIF Regression by Quantile (Quadratic Schooling)

Table 4B.1: RIF Regression for Malawi (Quadratic Schooling)

quantile	q(10)	q(25)	q(50)	q(75)	q(90)
A. Daily					
sch	0.278*	0.042	0.031	-0.096	-0.130
	(0.165)	(0.144)	(0.140)	(0.132)	(0.126)
sch2	-0.011	0.003	0.006	0.011	0.011
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
RIF mean	2.190	3.205	4.583	5.618	6.298
R^2	0.26	0.33	0.35	0.17	0.12
Obs.	182	182	182	182	182
B. Weekly					
sch	0.046	0.030	-0.037	-0.179***	-0.376***
	(0.073)	(0.062)	(0.048)	(0.058)	(0.095)
sch2	-0.003	0.002	0.008***	0.023***	0.040***
	(0.004)	(0.004)	(0.003)	(0.004)	(0.008)
RIF mean	2.946	3.602	4.592	5.365	6.243
R^2	0.50	0.44	0.39	0.36	0.31
Obs.	505	505	505	505	505
C. Monthly					
sch	0.137***	0.103***	-0.064***	-0.177***	-0.260***
	(0.027)	(0.021)	(0.013)	(0.017)	(0.024)
sch2	-0.004***	0.002**	0.012***	0.021***	0.024***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
RIF mean	2.893	3.694	4.702	5.544	6.304
R^2	0.40	0.46	0.52	0.43	0.25
Obs.	5129	5129	5129	5129	5129
D. Pooled					
sch	0.126***	0.083***	-0.061***	-0.167***	-0.258***
	(0.026)	(0.020)	(0.014)	(0.017)	(0.024)
sch2	-0.003**	0.003**	0.012***	0.020***	0.024***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
RIF mean	2.841	3.672	4.692	5.558	6.298
R^2	0.39	0.45	0.49	0.41	0.25
Obs.	5816	5816	5816	5816	5816

Note: Bootstrap standard errors computed by 500 replications in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Full results with all variables available upon request.

Table 4B.2: RIF Regression for Tanzania (Quadratic Schooling)

quantile	q(10)	q(25)	q(50)	q(75)	q(90)
A. Daily					
sch	-0.016	0.030	-0.069**	-0.081**	-0.035
	(0.038)	(0.031)	(0.030)	(0.037)	(0.036)
sch2	0.000	0.000	0.011***	0.016***	0.008*
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)
RIF mean	-0.274	0.789	2.091	3.854	4.922
R^2	0.19	0.41	0.53	0.44	0.20
Obs.	3738	3738	3738	3738	3738
B. Weekly					
sch	-0.054	-0.011	-0.009	-0.003	-0.041
	(0.044)	(0.033)	(0.035)	(0.039)	(0.059)
sch2	0.007	0.005	0.006*	0.010**	0.019**
	(0.004)	(0.003)	(0.004)	(0.004)	(0.007)
RIF mean	-0.232	0.752	2.048	3.342	4.509
R^2	0.24	0.44	0.60	0.49	0.28
Obs.	1929	1929	1929	1929	1929
C. Monthly					
sch	0.147***	0.116***	0.076***	-0.065***	-0.170***
	(0.035)	(0.025)	(0.016)	(0.012)	(0.019)
sch2	-0.005***	0.000	0.006***	0.014***	0.020***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
RIF mean	2.137	3.178	4.352	5.203	5.838
R^2	0.45	0.51	0.45	0.32	0.24
Obs.	4830	4830	4830	4830	4830
D. Pooled					
sch	0.074***	0.050***	0.046***	-0.088***	-0.202***
	(0.019)	(0.016)	(0.014)	(0.010)	(0.012)
sch2	-0.006***	-0.002**	0.006***	0.020***	0.028***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
RIF mean	0.307	1.545	3.199	4.615	5.461
R^2	0.30	0.55	0.58	0.40	0.28
Obs.	11215	11215	11215	11215	11215

Note: Bootstrap standard errors computed by 500 replications in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Full results with all variables available upon request.

Table 4B.3: RIF Regression for Uganda (Quadratic Schooling)

quantile	q(10)	q(25)	q(50)	q(75)	q(90)
A. Daily					
sch	0.161**	0.223***	0.106***	0.040	-0.003
	(0.081)	(0.048)	(0.040)	(0.037)	(0.044)
sch2	-0.004	-0.008***	-0.001	0.004	0.006
	(0.005)	(0.003)	(0.003)	(0.003)	(0.004)
RIF mean	1.786	2.774	3.719	4.580	5.270
R^2	0.36	0.39	0.31	0.21	0.12
Obs.	1262	1262	1262	1262	1262
B. Weekly					
sch	0.135	0.154**	0.154***	0.027	-0.118*
	(0.088)	(0.061)	(0.054)	(0.045)	(0.070)
sch2	-0.005	-0.005	-0.003	0.006**	0.017***
	(0.006)	(0.004)	(0.004)	(0.003)	(0.006)
RIF mean	1.568	2.756	3.793	4.424	5.179
R^2	0.28	0.45	0.38	0.23	0.17
Obs.	589	589	589	589	589
C. Monthly					
sch	0.340***	0.240***	0.074***	-0.095***	-0.204***
	(0.060)	(0.034)	(0.022)	(0.017)	(0.022)
sch2	-0.012***	-0.005***	0.005***	0.012***	0.017***
	(0.003)	(0.002)	(0.001)	(0.001)	(0.002)
RIF mean	2.16	3.242	4.35	5.059	5.544
R^2	0.46	0.45	0.44	0.30	0.20
Obs.	2765	2765	2765	2765	2765
D. Pooled					
sch	0.229***	0.206***	0.103***	-0.071***	-0.154***
	(0.039)	(0.023)	(0.019)	(0.013)	(0.017)
sch2	-0.008***	-0.005***	0.003***	0.011***	0.015***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
RIF mean	1.931	3.052	4.055	4.846	5.446
R^2	0.41	0.42	0.38	0.28	0.18
Obs.	4631	4631	4631	4631	4631

Note: Bootstrap standard errors computed by 500 replications in parentheses; * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Full results with all variables available upon request.

Appendix 4C: Gender Differences in Returns to Education

Table 4C.1: Returns to Education by Gender and Period - Malawi

	Daily		Weekly		Monthly		Pooled	
	female	male	female	male	female	male	female	male
sch	0.012	-0.052	-0.162***	-0.082**	-0.095***	-0.056***	-0.095***	-0.057***
	(0.123)	(0.094)	(0.048)	(0.038)	(0.016)	(0.009)	(0.016)	(0.009)
sch2	0.008	0.008	0.017***	0.012***	0.015***	0.011***	0.014***	0.011***
	(0.008)	(0.006)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
age	0.088	-0.005	0.105***	0.067**	0.033***	0.062***	0.046***	0.060***
	(0.087)	(0.076)	(0.034)	(0.026)	(0.011)	(0.007)	(0.011)	(0.007)
age2	-0.085	0.021	-0.136***	-0.067**	-0.024	-0.058***	-0.043***	-0.056***
	(0.113)	(0.094)	(0.047)	(0.033)	(0.015)	(0.008)	(0.014)	(0.008)
rural	-0.919***	-0.085	-0.229*	-0.363***	-0.121***	-0.213***	-0.152***	-0.232***
	(0.334)	(0.248)	(0.132)	(0.096)	(0.039)	(0.022)	(0.038)	(0.022)
year	0.138	1.106***	1.171***	0.981***	1.302***	1.284***	1.248***	1.243***
	(0.361)	(0.258)	(0.121)	(0.090)	(0.037)	(0.021)	(0.036)	(0.021)
weeks	1.187***	1.306***	1.198***	0.977***	1.226***	1.094***	1.206***	1.084***
	(0.268)	(0.201)	(0.084)	(0.065)	(0.034)	(0.023)	(0.032)	(0.022)
constant	-2.020	-1.381	-2.156***	-0.874	-1.927***	-1.816***	-1.971***	-1.690***
	(1.892)	(1.588)	(0.630)	(0.537)	(0.232)	(0.143)	(0.223)	(0.143)
AME(sch)	0.148***	0.091***	0.105***	0.115***	0.198***	0.145***	0.182***	0.142***
	(0.043)	(0.032)	(0.014)	(0.012)	(0.006)	(0.003)	(0.005)	(0.003)
R^2	0.58	0.44	0.77	0.63	0.8	0.76	0.77	0.72
Obs.	49	133	144	361	1,265	3,864	1,458	4,358

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4C.2: Returns to Education by Gender and Period - Tanzania

	Daily		Weekly		Monthly		Pooled	
	female	male	female	male	female	male	female	male
sch	-0.032	-0.013	-0.114***	0.010	0.016	0.036***	-0.050***	-0.002
	(0.027)	(0.020)	(0.027)	(0.022)	(0.014)	(0.013)	(0.009)	(0.008)
sch2	0.006*	0.006***	0.019***	0.006***	0.009***	0.006***	0.014***	0.007***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
age	0.033***	0.064***	0.039***	0.049***	0.079***	0.097***	0.060***	0.069***
	(0.012)	(0.010)	(0.012)	(0.012)	(0.009)	(0.008)	(0.007)	(0.006)
age2	-0.043***	-0.077***	-0.046***	-0.054***	-0.062***	-0.094***	-0.060***	-0.073***
	(0.016)	(0.014)	(0.016)	(0.016)	(0.012)	(0.010)	(0.009)	(0.007)
rural	-0.571***	-0.493***	-0.092	-0.255***	-0.215***	-0.302***	-0.300***	-0.356***
	(0.072)	(0.050)	(0.075)	(0.063)	(0.035)	(0.030)	(0.032)	(0.025)
panel	-0.452***	-0.128***	-0.042	-0.178***	0.025	-0.122***	-0.051*	-0.113***
	(0.068)	(0.048)	(0.073)	(0.063)	(0.036)	(0.030)	(0.029)	(0.022)
weeks	1.036***	1.226***	1.018***	1.135***	1.107***	1.047***	1.083***	1.141***
	(0.024)	(0.018)	(0.022)	(0.020)	(0.025)	(0.023)	(0.012)	(0.010)
constant	-0.704***	-1.144***	-1.090***	-0.965***	-2.669***	-2.085***	-1.776***	-1.364***
	(0.227)	(0.190)	(0.228)	(0.222)	(0.159)	(0.154)	(0.118)	(0.103)
AME(sch)	0.148***	0.091***	0.105***	0.115***	0.198***	0.145***	0.182***	0.142***
	(0.043)	(0.032)	(0.014)	(0.012)	(0.006)	(0.003)	(0.005)	(0.003)
R^2	0.64	0.72	0.78	0.78	0.77	0.66	0.80	0.76
Obs.	1,338	2,400	683	1,246	1,815	3,015	4,073	7,142

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4C.3: Returns to Education by Gender and Period - Uganda

	Daily		Weekly		Monthly		Pooled	
	female	male	female	male	female	male	female	male
sch	0.108***	0.111***	0.076	0.040	0.054**	0.071***	0.074***	0.066***
	(0.040)	(0.029)	(0.051)	(0.044)	(0.023)	(0.020)	(0.017)	(0.015)
sch2	0.001	-0.002	0.003	0.004	0.005***	0.003***	0.004***	0.003***
	(0.003)	(0.002)	(0.004)	(0.003)	(0.001)	(0.001)	(0.001)	(0.001)
age	-0.005	0.085***	0.092***	0.072***	0.093***	0.066***	0.078***	0.064***
	(0.026)	(0.016)	(0.034)	(0.023)	(0.016)	(0.012)	(0.012)	(0.009)
age2	0.011	-0.111***	-0.112**	-0.082***	-0.097***	-0.069***	-0.084***	-0.073***
	(0.035)	(0.022)	(0.046)	(0.030)	(0.021)	(0.016)	(0.017)	(0.012)
rural	-0.175	-0.282***	-0.298**	-0.377***	-0.171***	-0.276***	-0.142***	-0.288***
	(0.117)	(0.063)	(0.147)	(0.108)	(0.052)	(0.045)	(0.045)	(0.035)
panel	-0.151	0.044	-0.322	-0.001	0.249***	0.133***	0.143***	0.147***
	(0.128)	(0.062)	(0.204)	(0.119)	(0.059)	(0.048)	(0.048)	(0.036)
weeks	1.178***	1.260***	1.229***	1.043***	1.112***	1.132***	1.140***	1.151***
	(0.068)	(0.047)	(0.092)	(0.067)	(0.044)	(0.044)	(0.034)	(0.029)
constant	-1.461***	-2.469***	-3.117***	-1.452***	-3.247***	-2.331***	-3.022***	-2.005***
	(0.508)	(0.300)	(0.663)	(0.481)	(0.274)	(0.264)	(0.222)	(0.182)
AME(sch)	0.118***	0.085***	0.109***	0.091***	0.158***	0.139***	0.145***	0.110***
	(0.016)	(0.009)	(0.018)	(0.013)	(0.007)	(0.006)	(0.005)	(0.004)
R^2	0.58	0.52	0.64	0.50	0.70	0.56	0.69	0.54
Obs.	281	981	167	422	1,022	1,743	1,475	3,156

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4C.4: Returns to Education by Gender and Period - Ganyu

	female	male
sch	0.003	0.002
	(0.007)	(0.007)
sch2	0.001*	0.002***
	(0.001)	(0.001)
age	0.036***	0.046***
	(0.004)	(0.004)
age2	-0.042***	-0.055***
	(0.005)	(0.005)
rural	-0.206***	-0.432***
	(0.031)	(0.029)
year	1.168***	1.270***
	(0.017)	(0.016)
weeks	0.987***	1.023***
	(0.009)	(0.009)
constant	-1.205***	-1.002***
	(0.079)	(0.078)
AME(sch)	0.013***	0.022***
	(0.003)	(0.002)
R2	0.742	0.766
Obs.	8,246	8,282

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix 4D:RIF OB Gender Decomposition by Quantile and Country

Table 4D.1: RIF OB Gender Decomposition by Quantile for Malawi(Daily)

Quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	2.204***	2.800***	3.465***	3.782***	4.638***	5.136***	5.545***	5.924***	6.343***
Counterfactual	2.224***	2.851***	3.561***	3.935***	4.631***	5.058***	5.419***	5.762***	6.242***
Female	2.294***	2.888***	3.737***	4.230***	4.565***	5.005***	5.421***	5.664***	6.189***
Difference	-0.09	-0.088	-0.271	-0.448	0.073	0.131	0.123	0.26	0.154
Explained	-0.02	-0.051	-0.096	-0.153	0.007	0.078	0.126	0.162	0.101
Unexplained	-0.07	-0.037	-0.175	-0.294	0.066	0.053	-0.003	0.098	0.053
Pure explained	0.006	-0.024	-0.121	-0.139	-0.011	0.141	0.127	0.011	0.086
education	0.116	0.079	0.008	-0.055	0.002	0.025	0.024	-0.020	-0.005
covariates	-0.109	-0.103	-0.129	-0.084	-0.013	0.117	0.103	0.031	0.091
Pure unexplained	-0.001	0.037	-0.102	-0.22	0.115	0.108	0.032	0.114	0.087
education	0.834	-0.269	-1.232	-1.819	-2.042*	-1.138	-0.871	-0.054	1.699
covariates	0.508	-9.582	-7.637	-1.515	3.372	2.109	-0.171	3.201	2.367
_cons	-1.344	9.889	8.767	3.113	-1.215	-0.863	1.074	-3.034	-3.979
Specification error	-0.027	-0.027	0.026	-0.014	0.018	-0.063	-0.001	0.151	0.015
Reweight error	-0.068	-0.074	-0.074	-0.074	-0.049	-0.055	-0.035	-0.016	-0.035
Obs. Male	133	133	133	133	133	133	133	133	133
Obs. Female	49	49	49	49	49	49	49	49	49
Obs. Total	182	182	182	182	182	182	182	182	182

Notes: Bootstrap standard errors computed by 500 replications. * p < 0.10, ** p < 0.05, *** p < 0.01. The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

Table 4D.2: RIF OB Gender Decomposition by Quantile for Malawi(Weekly)

Quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	3.058***	3.462***	3.976***	4.269***	4.671***	5.043***	5.307***	5.693***	6.254***
Counterfactual	2.774***	3.401***	3.754***	4.233***	4.538***	4.898***	5.243***	5.611***	6.130***
Female	2.675***	3.342***	3.755***	4.171***	4.378***	4.744***	5.048***	5.398***	6.239***
Difference	0.383	0.120	0.221	0.098	0.292*	0.299*	0.258*	0.294	0.014
Explained	0.284	0.061	0.222	0.036	0.132	0.145	0.064	0.082	0.124
Unexplained	0.099	0.059	-0.001	0.062	0.16	0.154	0.195	0.213	-0.11
Pure explained	0.197	0.166	0.147	0.119	0.130	0.138	0.097	0.057	0.108
education	0.002	0.027	0.033	0.007	0.006	-0.001	-0.001	-0.035	-0.055
covariates	0.194	0.139	0.115	0.112	0.123	0.139*	0.098	0.093	0.163**
Pure unexplained	0.070	0.038	-0.012	0.048	0.150	0.144	0.182	0.198	-0.127
education	-0.512	0.177	0.618	0.391	0.680*	0.171	-0.083	-0.191	0.109
covariates	-0.634	-4.027	-1.528	-1.55	-0.696	-3.256*	-1.075	-2.938*	-0.571
_cons	1.216	3.889	0.898	1.207	0.166	3.229*	1.340	3.327*	0.335
Specification error	0.087	-0.104	0.075	-0.083	0.003	0.006	-0.033	0.024	0.016
Reweight error	0.030	0.020	0.012	0.014	0.010	0.010	0.013	0.015	0.018
Obs. Male	361	361	361	361	361	361	361	361	361
Obs. Female	144	144	144	144	144	144	144	144	144
Obs. Total	505	505	505	505	505	505	505	505	505

Notes: Notes: Bootstrap standard errors computed by 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

Table 4D.3: RIF OB Gender Decomposition by Quantile for Malawi(Monthly)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	2.939***	3.482***	3.921***	4.350***	4.670***	4.985***	5.362***	5.770***	6.283***
Counterfactual	2.911***	3.588***	3.988***	4.435***	4.799***	5.183***	5.566***	6.018***	6.514***
Female	2.709***	3.375***	3.870***	4.429***	4.718***	4.960***	5.417***	5.938***	6.370***
Difference	0.229***	0.108	0.051	-0.079	-0.048	0.025	-0.055	-0.168***	-0.087
Explained	0.028	-0.105	-0.067	-0.085	-0.128**	-0.198***	-0.204***	-0.249***	-0.231***
Unexplained	0.202***	0.213***	0.118*	0.006	0.080*	0.223***	0.149**	0.08	0.144**
Pure explained	0.054	-0.061	-0.087*	-0.117**	-0.156***	-0.194***	-0.238***	-0.253***	-0.245***
education	-0.029**	-0.069***	-0.097***	-0.111***	-0.119***	-0.119***	-0.146***	-0.171***	-0.161***
covariates	0.084*	0.007	0.01	-0.007	-0.038	-0.076**	-0.092**	-0.082**	-0.083***
Pure unexplained	0.190**	0.202***	0.105*	-0.006	0.067	0.209***	0.133**	0.063	0.126*
education	0.102	0.081	-0.607**	-0.05	-0.044	0.054	-0.233	-0.174	-0.252*
covariates	1.057	0.166	-1.156	-0.099	0.974*	0.866	-0.588	-1.441**	-0.696
_cons	-0.968	-0.045	1.868*	0.143	-0.863	-0.712	0.954	1.678**	1.074
Specification error	-0.027	-0.044	0.02	0.032	0.028	-0.004	0.034	0.004	0.013
Reweight error	0.012	0.011	0.013	0.013	0.013*	0.014*	0.016**	0.017**	0.018***
Obs. Male	3864	3864	3864	3864	3864	3864	3864	3864	3864
Obs. Female	1265	1265	1265	1265	1265	1265	1265	1265	1265
Obs. Total	5129	5129	5129	5129	5129	5129	5129	5129	5129

Notes: Notes: Bootstrap standard errors computed by 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

Table 4D.4: RIF OB Gender Decomposition by Quantile for Malawi(Pooled)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	2.914***	3.476***	3.907***	4.345***	4.670***	4.993***	5.353***	5.760***	6.281***
Counterfactual	2.837***	3.495***	3.943***	4.372***	4.769***	5.129***	5.521***	5.973***	6.475***
Female	2.706***	3.358***	3.831***	4.340***	4.754***	5.004***	5.367***	5.891***	6.349***
Difference	0.208***	0.118	0.076	0.005	-0.085**	-0.012	-0.014	-0.131**	-0.068
Explained	0.077	-0.019	-0.036	-0.027	-0.099**	-0.137***	-0.168***	-0.212***	-0.194***
Unexplained	0.131*	0.137**	0.112**	0.032	0.015	0.125***	0.154***	0.081	0.126**
Pure explained	0.076	-0.034	-0.06	-0.088**	-0.123***	-0.154***	-0.191***	-0.207***	-0.203***
education	-0.017	-0.047***	-0.074***	-0.087***	-0.095***	-0.097***	-0.121***	-0.145***	-0.144***
covariates	0.092*	0.014	0.014	-0.001	-0.028	-0.058*	-0.070**	-0.062**	-0.059***
Pure unexplained	0.116*	0.125**	0.098*	0.018	0.001	0.110**	0.138***	0.065	0.109**
education	0.053	0.101	-0.430**	-0.086	-0.004	-0.011	-0.231	-0.166	-0.020
covariates	0.215	-0.501	-1.573*	-0.476	0.686	0.492	-0.718	-1.498***	-0.432
_cons	-0.152	0.524	2.102**	0.580	-0.682	-0.370	1.088*	1.729***	0.561
Specification error	0.014	0.012*	0.014*	0.014**	0.014**	0.015**	0.016**	0.017***	0.018***
Reweight error	0.002	0.015	0.024	0.06	0.023	0.018	0.023	-0.005	0.009
Obs. Male	4358	4358	4358	4358	4358	4358	4358	4358	4358
Obs. Female	1458	1458	1458	1458	1458	1458	1458	1458	1458
Obs. Total	5816	5816	5816	5816	5816	5816	5816	5816	5816

Notes: Notes: Bootstrap standard errors computed by 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

Table 4D.5: RIF OB Gender Decomposition by Quantile for Malawi(*Ganyu*)

	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	0.872***	1.566***	2.020***	2.481***	2.929***	3.312***	3.689***	4.161***	4.715***
Counterfactual	0.410***	1.093***	1.567***	2.082***	2.555***	2.944***	3.378***	3.850***	4.471***
Female	0.221***	0.989***	1.495***	1.875***	2.231***	2.657***	2.958***	3.409***	3.931***
Difference	0.651***	0.578***	0.526***	0.607***	0.698***	0.655***	0.731***	0.753***	0.784***
Explained	0.462***	0.473***	0.453***	0.400***	0.374***	0.368***	0.311***	0.311***	0.244***
Unexplained	0.190***	0.105**	0.073**	0.207***	0.324***	0.287***	0.420***	0.442***	0.539***
Pure explained	0.532***	0.496***	0.465***	0.407***	0.379***	0.372***	0.363***	0.335***	0.280***
education	0.035***	0.052***	0.036***	0.018**	0.027***	0.037***	0.034***	0.037***	0.033***
covariates	0.497***	0.444***	0.429***	0.389***	0.352***	0.335***	0.329***	0.298***	0.247***
Pure unexplained	0.246***	0.175***	0.145***	0.278***	0.386***	0.345***	0.475***	0.497***	0.585***
education	-0.110	0.045	0.069	0.071	0.026	0.044	0.067*	0.020	-0.072
covariates	-0.675	0.018	0.641**	0.964***	0.563**	0.240	0.274	0.611**	-0.308
_cons	1.032*	0.112	-0.565*	-0.756**	-0.203	0.06	0.134	-0.133	0.965***
Specification error	-0.071	-0.023	-0.012	-0.007	-0.005	-0.004	-0.052*	-0.024	-0.036**
Reweight error	-0.056***	-0.071***	-0.073***	-0.071***	-0.062***	-0.057***	-0.055***	-0.056***	-0.046***
Obs. Male	8282	8282	8282	8282	8282	8282	8282	8282	8282
Obs. Female	8246	8246	8246	8246	8246	8246	8246	8246	8246
Obs. Total	16528	16528	16528	16528	16528	16528	16528	16528	16528

Notes: Notes: Bootstrap standard errors computed by 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

Table 4D.6: RIF OB Gender Decomposition by Quantile for Tanzania(Daily)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	0.090*	0.925***	1.664***	2.241***	2.817***	3.481***	4.165***	4.730***	5.218***
Counterfactual	-0.358***	0.141	0.732***	1.395***	1.842***	2.445***	3.013***	3.854***	4.765***
Female	-0.661***	-0.242***	0.475***	0.906***	1.237***	1.670***	2.132***	2.755***	3.603***
Difference	0.751***	1.167***	1.190***	1.335***	1.580***	1.811***	2.033***	1.975***	1.615***
Explained	0.448***	0.784***	0.932***	0.846***	0.975***	1.035***	1.152***	0.876***	0.454***
Unexplained	0.303***	0.383***	0.258***	0.489***	0.605***	0.776***	0.881***	1.099***	1.161***
Pure explained	0.809***	0.996***	0.964***	1.062***	1.123***	1.133***	0.885***	0.622***	0.377***
education	0.029	0.068***	0.046**	0.040*	0.092***	0.103***	0.092***	0.073***	0.032*
covariates	0.780***	0.928***	0.918***	1.022***	1.031***	1.030***	0.793***	0.549***	0.345***
Pure unexplained	0.340***	0.461***	0.355***	0.583***	0.701***	0.875***	0.983***	1.201***	1.228***
education	-0.296**	0.135	0.227*	0.138	0.053	0.022	0.102	0.005	0.026
covariates	0.261	0.426	1.690**	0.661	1.080*	1.294**	1.645**	1.811**	1.520**
_cons	0.375	-0.099	-1.562*	-0.215	-0.433	-0.442	-0.764	-0.615	-0.319
Specification error	-0.361***	-0.212***	-0.032	-0.216***	-0.148**	-0.097	0.267***	0.254***	0.077
Reweight error	-0.037***	-0.078***	-0.097***	-0.094***	-0.096***	-0.099***	-0.102***	-0.102***	-0.067***
Obs. Male	2400	2400	2400	2400	2400	2400	2400	2400	2400
Obs. Female	1338	1338	1338	1338	1338	1338	1338	1338	1338
Obs. Total	3738	3738	3738	3738	3738	3738	3738	3738	3738

Notes: Notes: Bootstrap standard errors computed by 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

Table 4D.7: RIF OB Gender Decomposition by Quantile for Tanzania(Weekly)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	0.124*	0.823***	1.477***	2.032***	2.447***	2.895***	3.563***	4.123***	4.899***
Counterfactual	-0.176**	0.295***	0.736***	1.175***	1.645***	2.222***	2.718***	3.321***	4.158***
Female	-0.417***	0.014	0.490***	0.837***	1.173***	1.650***	2.170***	2.724***	3.337***
Difference	0.541***	0.809***	0.987***	1.195***	1.274***	1.246***	1.393***	1.400***	1.562***
Explained	0.300***	0.528***	0.741***	0.857***	0.802***	0.673***	0.845***	0.803***	0.741***
Unexplained	0.240*	0.281***	0.247***	0.338***	0.472***	0.573***	0.548***	0.597***	0.822***
Pure explained	0.457***	0.620***	0.788***	0.796***	0.783***	0.817***	0.811***	0.730***	0.611***
education	0.027	0.058*	0.095***	0.080***	0.093***	0.127***	0.122***	0.140***	0.162***
covariates	0.430***	0.562***	0.693***	0.716***	0.690***	0.690***	0.689***	0.590***	0.449***
Pure unexplained	0.257**	0.323***	0.306***	0.408***	0.555***	0.655***	0.629***	0.684***	0.902***
education	0.083	-0.013	0.024	0.196	0.280**	0.059	0.031	0.124	0.123
covariates	-0.371	0.086	1.189	0.305	0.471	-0.833	-1.072	-0.583	-0.473
_cons	0.544	0.249	-0.907	-0.093	-0.196	1.429*	1.669**	1.144	1.252
Specification error	-0.157	-0.091	-0.047	0.061	0.019	-0.143**	0.034	0.073	0.129
Reweight error	-0.016	-0.042**	-0.059**	-0.070***	-0.083***	-0.082***	-0.080***	-0.087***	-0.081***
Obs. Male	1246	1246	1246	1246	1246	1246	1246	1246	1246
Obs. Female	683	683	683	683	683	683	683	683	683
Obs. Total	1929	1929	1929	1929	1929	1929	1929	1929	1929

Notes: Notes: Bootstrap standard errors computed by 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

Table 4D.8: RIF OB Gender Decomposition by Quantile for Tanzania(Monthly)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	2.437***	3.230***	3.708***	4.220***	4.559***	4.803***	5.128***	5.499***	5.900***
Counterfactual	2.148***	2.761***	3.446***	3.918***	4.374***	4.720***	5.014***	5.333***	5.783***
Female	1.774***	2.565***	2.982***	3.415***	3.746***	4.227***	4.663***	5.067***	5.702***
Difference	0.663***	0.665***	0.726***	0.804***	0.814***	0.576***	0.465***	0.432***	0.198***
Explained	0.290***	0.469***	0.262***	0.302***	0.185***	0.083	0.114**	0.166***	0.117***
Unexplained	0.374***	0.196***	0.464***	0.503***	0.629***	0.493***	0.351***	0.266***	0.082*
Pure explained	0.382***	0.395***	0.329***	0.230***	0.195***	0.153***	0.124***	0.112***	0.092***
education	-0.001	-0.004	-0.006	-0.007	-0.008	-0.009	-0.010	-0.012	-0.014
covariates	0.383***	0.399***	0.335***	0.237***	0.203***	0.162***	0.134***	0.124***	0.106***
Pure unexplained	0.343***	0.154**	0.428***	0.473***	0.603***	0.470***	0.332***	0.254***	0.073
education	0.019	0.445*	0.441**	0.17	-0.243	-0.646***	-0.512***	-0.402***	-0.331***
covariates	-2.052	1.947**	1.100	-0.195	-1.221**	-0.917*	0.092	0.513	1.053***
_cons	2.375	-2.238**	-1.114	0.497	2.066***	2.033***	0.753	0.144	-0.650
Specification error	-0.092	0.074	-0.067	0.072**	-0.010	-0.070	-0.010	0.054*	0.024
Reweight error	0.031	0.043*	0.036*	0.030*	0.026*	0.023**	0.019**	0.012	0.009
Obs. Male	3015	3015	3015	3015	3015	3015	3015	3015	3015
Obs. Female	1815	1815	1815	1815	1815	1815	1815	1815	1815
Obs. Total	4830	4830	4830	4830	4830	4830	4830	4830	4830

Notes: Notes: Bootstrap standard errors computed by 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

Table 4D.9: RIF OB Gender Decomposition by Quantile for Tanzania(Pooled)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	0.732***	1.644***	2.345***	3.001***	3.663***	4.155***	4.664***	5.079***	5.565***
Counterfactual	0.230***	1.089***	1.795***	2.496***	3.120***	3.714***	4.368***	4.831***	5.460***
Female	-0.131***	0.642***	1.181***	1.824***	2.440***	2.998***	3.572***	4.239***	5.104***
Difference	0.863***	1.002***	1.165***	1.176***	1.223***	1.157***	1.092***	0.840***	0.461***
Explained	0.502***	0.555***	0.550***	0.505***	0.543***	0.442***	0.296***	0.248***	0.105***
Unexplained	0.361***	0.448***	0.615***	0.671***	0.680***	0.716***	0.796***	0.592***	0.355***
Pure explained	0.543***	0.588***	0.596***	0.592***	0.501***	0.364***	0.240***	0.167***	0.075***
education	0.026**	0.026***	0.032***	0.051***	0.054***	0.048***	0.029**	0.018	-0.004
covariates	0.517***	0.562***	0.564***	0.541***	0.447***	0.316***	0.210***	0.150***	0.079***
Pure unexplained	0.362***	0.449***	0.614***	0.668***	0.673***	0.707***	0.785***	0.579***	0.340***
education	0.358***	0.271**	0.186*	0.079	-0.060	-0.107	-0.317***	-0.691***	-0.681***
covariates	0.662	1.531***	0.990**	0.637	0.923**	-0.063	-1.192***	-0.994***	-0.191
_cons	-0.658	-1.354***	-0.563	-0.048	-0.19	0.877**	2.294***	2.264***	1.213***
Specification error	-0.041	-0.033	-0.046	-0.087***	0.042	0.078**	0.056**	0.081***	0.031
Reweight error	-0.001	-0.001	0.001	0.003	0.007	0.009	0.011	0.013**	0.015***
Obs. Male	7142	7142	7142	7142	7142	7142	7142	7142	7142
Obs. Female	4073	4073	4073	4073	4073	4073	4073	4073	4073
Obs. Total	11215	11215	11215	11215	11215	11215	11215	11215	11215

Notes: Notes: Bootstrap standard errors computed by 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

Table 4D.10: RIF OB Gender Decomposition by Quantile for Uganda (Daily)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	2.115***	2.785***	3.246***	3.617***	3.966***	4.278***	4.576***	4.957***	5.398***
Counterfactual	1.426***	2.381***	2.763***	3.245***	3.562***	3.957***	4.285***	4.649***	5.214***
Female	1.147***	1.800***	2.307***	2.516***	3.033***	3.288***	3.549***	3.958***	4.478***
Difference	0.968***	0.986***	0.940***	1.101***	0.933***	0.990***	1.027***	1.000***	0.920***
Explained	0.689***	0.404**	0.484***	0.372***	0.404***	0.321***	0.290***	0.309***	0.185*
Unexplained	0.279	0.582***	0.456***	0.729***	0.529***	0.670***	0.737***	0.691***	0.736***
Pure explained	0.703***	0.598***	0.505***	0.482***	0.413***	0.383***	0.319***	0.271***	0.204***
education	0.185**	0.242***	0.201***	0.197***	0.189***	0.176***	0.142***	0.128***	0.093**
covariates	0.517***	0.356***	0.305***	0.285***	0.224***	0.207***	0.177***	0.144***	0.110**
Pure unexplained	0.420**	0.669***	0.528***	0.793***	0.580***	0.711***	0.776***	0.723***	0.760***
education	0.031	-0.319	-0.034	-0.199	-0.276	-0.179	-0.031	-0.114	-0.167
covariates	2.169	1.049	1.510	1.536	1.340	0.888	0.407	1.279	0.912
_cons	-1.780	-0.061	-0.948	-0.544	-0.484	0.002	0.399	-0.443	0.015
Specification error	-0.014	-0.195	-0.022	-0.110	-0.009	-0.062	-0.028	0.038	-0.019
Reweight error	-0.141	-0.088	-0.072	-0.064	-0.051	-0.041	-0.039	-0.032	-0.024
Obs. Male	981	981	981	981	981	981	981	981	981
Obs. Female	281	281	281	281	281	281	281	281	281
Obs. Total	1262	1262	1262	1262	1262	1262	1262	1262	1262

Notes: Notes: Bootstrap standard errors computed by 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on ‘Male’ and ‘female’ implies that the mean for each group is significantly different from their combine mean.

Table 4D.11: RIF OB Gender Decomposition by Quantile for Uganda (Weekly)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	1.914***	2.763***	3.210***	3.720***	3.954***	4.320***	4.401***	4.842***	5.380***
Counterfactual	1.563***	2.467***	3.041***	3.348***	3.798***	4.156***	4.304***	4.727***	5.301***
Female	1.283***	1.791***	2.369***	2.712***	3.125***	3.539***	3.804***	4.253***	4.756***
Difference	0.631***	0.972***	0.841***	1.008***	0.828***	0.781***	0.597***	0.589***	0.623***
Explained	0.351	0.296	0.169	0.371**	0.155	0.163	0.096	0.115	0.079
Unexplained	0.280	0.676***	0.671***	0.637***	0.673***	0.617***	0.500***	0.474***	0.545***
Pure explained	0.494*	0.381**	0.312**	0.257**	0.165*	0.159*	0.159*	0.141*	0.071
education	0.116	0.126**	0.130**	0.130**	0.071	0.061	0.084	0.079	0.024
covariates	0.377	0.254	0.182	0.127	0.094	0.098	0.075	0.061	0.048
Pure unexplained	0.303	0.706***	0.695***	0.663***	0.698***	0.640***	0.529***	0.500***	0.577***
education	0.406	-0.169	0.057	0.07	0.271	-0.319	-0.650*	-0.439	-0.765*
covariates	1.587	-0.908	-0.759	0.342	-0.208	-1.581	-1.032	-0.245	0.054
_cons	-1.689	1.783	1.396	0.251	0.635	2.539*	2.212	1.184	1.288
Specification error	-0.143	-0.085	-0.143	0.114	-0.01	0.004	-0.063	-0.025	0.007
Reweight error	-0.024	-0.029	-0.023	-0.026	-0.026	-0.022	-0.029	-0.026	-0.032
Obs. Male	422	422	422	422	422	422	422	422	422
Obs. Female	167	167	167	167	167	167	167	167	167
Obs. Total	589	589	589	589	589	589	589	589	589

Notes: Notes: Bootstrap standard errors computed by 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

Table 4D.12: RIF OB Gender Decomposition by Quantile for Uganda (Monthly)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	2.546***	3.206***	3.684***	4.100***	4.427***	4.714***	5.034***	5.234***	5.588***
Counterfactual	2.127***	2.920***	3.411***	4.021***	4.388***	4.693***	4.969***	5.228***	5.609***
Female	1.724***	2.582***	3.048***	3.525***	3.999***	4.418***	4.640***	4.976***	5.332***
Difference	0.822***	0.623***	0.636***	0.575***	0.428***	0.296***	0.394***	0.258***	0.256***
Explained	0.419***	0.286***	0.273***	0.078	0.039	0.021	0.065	0.005	-0.021
Unexplained	0.403***	0.338***	0.364***	0.497***	0.388***	0.275***	0.329***	0.253***	0.276***
Pure explained	0.350***	0.203***	0.137**	0.130**	0.100**	0.058	0.040	-0.004	-0.026
education	-0.033*	-0.048*	-0.064*	-0.063*	-0.055*	-0.045*	-0.038*	-0.034*	-0.033*
covariates	0.383***	0.251***	0.201***	0.193***	0.155***	0.103***	0.078***	0.03	0.007
Pure unexplained	0.400***	0.334***	0.357***	0.489***	0.377***	0.265***	0.323***	0.250***	0.275***
education	0.122	-0.612	0.493	-0.028	-0.386	-0.276	-0.073	-0.125	0.020
covariates	-0.247	-0.784	-1.130	-0.362	-0.569	-0.254	0.159	0.135	0.155
_cons	0.525	1.730	0.994	0.879	1.332*	0.794	0.238	0.240	0.100
Specification error	0.069	0.083	0.135***	-0.052	-0.061	-0.038	0.026	0.009	0.006
Reweight error	0.003	0.004	0.007	0.008	0.011	0.01	0.006	0.003	0.001
Obs. Male	1743	1743	1743	1743	1743	1743	1743	1743	1743
Obs. Female	1022	1022	1022	1022	1022	1022	1022	1022	1022
Obs. Total	2765	2765	2765	2765	2765	2765	2765	2765	2765

Notes: Notes: Bootstrap standard errors computed by 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

Table 4D.13: RIF OB Gender Decomposition by Quantile for Uganda (Pooled)

quantile	q(10)	q(20)	q(30)	q(40)	q(50)	q(60)	q(70)	q(80)	q(90)
Overall									
Male	2.370***	3.034***	3.455***	3.900***	4.166***	4.532***	4.816***	5.103***	5.502***
Counterfactual	1.931***	2.803***	3.333***	3.736***	4.179***	4.512***	4.831***	5.168***	5.515***
Female	1.422***	2.304***	2.808***	3.228***	3.586***	3.975***	4.420***	4.783***	5.224***
Difference	0.949***	0.731***	0.647***	0.671***	0.580***	0.557***	0.396***	0.320***	0.278***
Explained	0.439***	0.231***	0.122**	0.163***	-0.013	0.02	-0.014	-0.065*	-0.013
Unexplained	0.510***	0.500***	0.525***	0.508***	0.593***	0.537***	0.411***	0.385***	0.291***
Pure explained	0.335***	0.212***	0.159***	0.083**	0.060	0.020	-0.018	-0.051**	-0.092***
education	0.007	0.005	-0.011	-0.031	-0.051**	-0.065***	-0.076***	-0.080***	-0.087***
covariates	0.328***	0.208***	0.169***	0.114***	0.111***	0.085***	0.058***	0.029**	-0.005
Pure unexplained	0.505***	0.492***	0.513***	0.493***	0.579***	0.522***	0.397***	0.374***	0.282***
education	0.230	-0.322	-0.247	-0.240	-0.502***	-0.636***	-0.547***	-0.313***	-0.162*
covariates	0.383	-0.652	-0.529	-0.671	-0.533	-0.757	-0.486	-0.061	0.076
_cons	-0.108	1.466	1.290	1.405*	1.614***	1.914***	1.430***	0.748*	0.368
Specification error	0.104*	0.019	-0.037	0.081**	-0.073	0.000	0.004	-0.014	0.079***
Reweight error	0.005	0.007	0.012	0.014	0.014	0.015**	0.014**	0.012**	0.009**
Obs. Male	3156	3156	3156	3156	3156	3156	3156	3156	3156
Obs. Female	1475	1475	1475	1475	1475	1475	1475	1475	1475
Obs. Total	4631	4631	4631	4631	4631	4631	4631	4631	4631

Notes: Notes: Bootstrap standard errors computed by 500 replications. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on 'Male' and 'female' implies that the mean for each group is significantly different from their combine mean.

Chapter 5

Youth Education and Household Welfare

5.1 Introduction

In the past two decades, Tanzania witnessed a considerable increase in education investment by households and the government, leading to a significant increase in enrolment across all levels (primary, secondary and tertiary). The free and Universal Primary Education (UPE) 2001 and the massive secondary school expansion program (known as 'ward secondary schools') that began in 2006 are the two significant reforms/initiatives in the 2000s that had a substantial effect on enrolment.

Following free UPE 2001, the primary school gross enrolment ratio increased from 84% in 2001 to 98.6% in 2002 and then to 109.9%[1] in 2005. During the same period, the number of primary schools increased from 11,873 to 12,286 and then to 14,257 in 2001, 2002 and 2005, respectively (URT, 2005). Similarly, following the 'ward secondary schools' initiative, secondary school gross enrolment increased from 20.2% in 2006 to 30.5% in 2007 and then to 36.9% in 2012. The number of secondary schools increased from 2,289 to 3,485 and then to 4,528 in 2006, 2007 and 2012, respectively (URT, 2008, 2013, 2016).

Household consumption expenditure remains popular among the measures of income and welfare in developing countries. Using the per adult household consumption relative to the national poverty line to proxy welfare in Tanzania, this ratio[2] improved significantly from 1.79 in 2001 to 2.28 in 2018, equivalent to a 27% increase. Within the period, poverty incidence (using national poverty lines)

[1]Enrolment went above 100% due to older children taking advantage of the opportunity for free schooling.

[2]Authors' own calculations from the Tanzania Household Surveys 2001 and 2018.

declined from to 36% to 26.4% (URT, 2002, 2019).

This essay's primary focus is to examine whether differences in educational attainment between the youth[3] cohort in 2001 (those aged between 15 and 35 who should have completed primary school before the reforms) and their corresponding cohort in 2018 (who should have benefited from the reforms) explain the welfare difference between the two periods. While many researchers have investigated the relationship between education and household welfare in developing countries, the most recent studies have focused on the role of gender, employment status, and rural/urban categories in that relationship (Delesalle, 2019; Khan and Morrissey, 2020; Belghith et al., 2020). This essay explores the relationship between education and welfare for this important demographic group which accounts for approximately 65% of the total labour force in Tanzania (URT, 2005, 2018).

With such a massive share in the labour force, there is no doubt that youth contribute substantially to the livelihood of the household in which they live (Arsalan et al., 2019), either through the income they earn from employment or through supplying labour to the household production activities and enterprises. For that reason in this paper we explore the role of youth in household welfare over time.

The essay makes two contributions to the literature on the link between education and household welfare. First, although considerable research has been devoted to welfare differences and their determinants across gender, employment, and rural/urban categories rather less attention has been paid to age groups (e.g. youth vs adults). Secondly, it examines how schooling gains between 2001 and 2018 are associated with household welfare changes over the period. Contrasted with previous studies which mainly examined the association between education and welfare at any given point in time, this essay examines how much of the welfare differences between 2001 and 2018 can be attributed to changes in the association

[3] What age range is considered as youth varies by institution and country e.g. United Nations (15 - 24), European Union (15 - 29), African Union (15 - 35), Uganda (18 - 30), Nigeria (18-35), Ethiopia (15-29). This study uses the Tanzania's definition which is in line with that of the African Union and define youth as all males and females aged 15 to 35 years (URT, 2007).

between education and welfare over this period. It also explores how much of the welfare differences can be attributed to changes in educational attainment between 2001 and 2018 (i.e. the welfare effects of change in education distribution between the period).

Using data from the Tanzania household budget surveys (HBS) for 2001 and 2018 and recentered influence function (RIF) based decomposition, the findings reveal that differences in educational attainment between youth and adults are significant factors in explaining the difference in welfare between the two groups in both years. Precisely, if adults had the same level of educational attainment as the youth, their welfare would have been about 40% and 32% higher in 2001 and 2018 respectively. The findings also suggest that if the youth in 2001 had the same education endowment as their 2018 counterparts, their welfare would have been about 20% higher. Although there appears to have been a decline in returns to education for the youth, we do not find evidence that this reduced welfare.

We do not control for endogeneity of education from unobserved ability or for endogenous selection of youth to be heads of households in our welfare decomposition, which is a limitation of the analysis. Absence of good instrumental variables from the data is one factor, and another is that most of the instrument free methods are not suitable given the presence of multiple endogenous regressors and the normal distribution nature of the dependent variable. Most importantly, even with good instruments, the debate on the combination of RIF and the standard methods for endogeneity control remains inconclusive. We do show that education is a minor factor determining whether a young person is a head of household: an additional year of schooling increases a young member's probability to head the household by only 0.5%. Such a small effect is unlikely to significantly affect the estimated coefficients of education in our main results.

The rest of the essay is organised as follows: Section 2 provides a review of the selected related literature, followed by a detailed description of the methodology in section 3. Section 4 describes the data used in our analysis and provides descriptive statistics for the main variables. Section 5 presents the results and discussion, and section 6 concludes.

5.2 Literature Review

The determinants of household welfare in developing countries have attracted many studies in the past 40 years. Factors such as education, age, gender, race, household shocks, employment status, sector of employment, place of residence and rural-urban migration have been found to have a significant contribution to welfare (e.g. Arouri et al. (2015); Arsalan et al. (2019); Delesalle (2019); Khan and Morrissey (2020); Belghith et al. (2020)) . In exploring these factors, previous studies have mainly categorised households in terms of gender, sector of employment and place of residence (rural/urban).

This essay can be linked to three broad categories of literature on household welfare. The first category consists of studies that focus on the determinants of household welfare. Within this category, we will focus on only such studies that include education as one of the determinants of household welfare. In Tanzania, these studies include Khan and Morrissey (2020), Arsalan et al. (2019) and Delesalle (2019). Using households data from the first three waves of the Tanzania National Panel Survey (TNPS), Khan and Morrissey (2020) found that an extra year of education of the head of household is associated with about 1.2% higher level of consumption (fixed effects estimation; no significant effect in the instrumental variable (IV) regression). Delesalle (2019) on the other hand used the same waves of the survey in combination with the 2002 Tanzania Population and Housing Census (TPHC) and estimated the consumption returns to head education of between 7.3% and 9.3% for rural households, much larger estimates than those by Khan and Morrissey (2020). Having employed different estimation strategies, samples, and dependent variables one would expect differences in the estimates of the association between education and consumption between the two studies, though not as large, warranting more investigation using other approaches.

Using the proportion of members with at least secondary education, Arsalan et al. (2019) explored the association between education and household log per capita expenditure and poverty status (based on the international poverty line). The study combined population density data, satellite data and household surveys from 12 developing countries, including Tanzania. They found that an increase in the number of working-age household members with secondary schooling by one person was associated with a 23% increase in expenditure for younger households and a 34% increase for older households; and a 7% and 6% decrease in poverty, respectively. In a closely related study on Vietnam, Arouri et al. (2015)[4] found that a percentage point increase in household members with an upper-secondary degree was associated with 36% and 55% higher household income and consumption respectively; and a decrease of the likelihood of being poor of 20%. More strikingly, they found, a percentage point increase in members with college/university degree was associated with a 92% and 71% higher level of income and consumption respectively and 19% lower likelihood of being poor.

Elsewhere in other developing countries, Himaz and Aturupana (2018) proxied household education by year of schooling of the most educated adult member in the household to estimate the association between education and household per capita expenditure in Sri Lanka. The study applied quantile IV regression on a sample of 72,811 households (18, 203 per survey) from the Household Income and Expenditure Surveys 1990/1, 1995/6, 2001/2 and 2005/6. The findings suggest that while, in general, an extra year of schooling increases welfare by about 3.8%, the effect varies considerably across the welfare distribution with the effect declining by quantile.

Fulford et al. (2014) estimated youth consumption returns to education for different cohorts in India. Using data from the Indian National Sample Surveys (INSS) 1983, 1987, 1993, 1999, and 2005 the study found that an extra year of education brings male cohorts 4% more consumption but provide no additional

[4] They categorized household as younger if the proportion of youth in the household is greater or equal to the proportion of youth in the population, and older if less than that in the population.

consumption for female cohorts. Alem and Söderbom (2012) found that although only higher education had a significant effect on consumption in 2004 and 2008 in Ethiopia, all levels (primary, secondary, and higher) significantly explain the growth of consumption between the years.

The second category consists of the studies examining welfare differences between various household categories such as gender (of the head or composition), sector of employment, and residence (rural/urban). The most common strategy in this literature is to include dummies for the different categories in their regressions to capture the welfare differences (see for example Arouri et al. (2015); Himaz and Aturupana (2018); Khan and Morrissey (2020); and Ayyash and Sek (2020)). Other studies analyse welfare and its determinants separately for each category (e.g. Van de Walle (2013); Delesalle (2019)).

For gender, findings from these types of studies generally suggest that female-headed households and households with higher proportions of female members tend to have lower welfare than their male counterparts (e.g. Ayyash and Sek (2020)). For residence, households residing in rural areas tend to have lower welfare than their urban counterparts. Furthermore, households with the majority of members employed in agriculture have lower welfare levels than those with the majority of members in wage employment or self-employment.

The third category of studies related to ours focuses on decomposing the welfare differences/inequality between groups or periods. The most relevant are those with a methodology based on the seminal works of Oaxaca (1973) and Blinder (1973) and their extensions (for most recent extensions, see Firpo et al. (2018) and Fortin et al. (2011)). Within this strand, our study is most closely related to Belghith et al. (2020) which employed Oaxaca-Blinder decomposition to examine what amount of poverty reduction can be attributed to changes in the endowments of household characteristics and the amount due to changes in the returns to these characteristics between 2012 and 2018 in Tanzania. Using data from HBS 2012 and 2018, the study's findings suggest that between 2012 and 2018 gains in education have benefited the better-off more than the poor and that the returns to education, while increased for the better-off, significantly declined for the poor.

Ramadan et al. (2018) applied RIF regression technique to decompose the welfare gap between various socio-demographic groups (male vs female-headed, rural vs urban households, educated vs uneducated head) for four Arab countries, namely Egypt, Tunisia, Jordan and Palestine. The study used household expenditure from household surveys[5] from the countries spanning between 2005 and 2015 to measure welfare. The study's findings revealed significant welfare gaps between female and male; rural and urban; and educated and uneducated within the countries. Differences in educational attainment remained one of the main determinants of the welfare gaps between male and female-headed and rural and urban households. Households with an educated head fared better in terms of welfare compared to their uneducated counterparts regardless of gender of the head or location. Agyire-Tettey et al. (2018) applied a similar approach to examine the rural-urban welfare gap for Ghana and obtained similar results: differences in educational attainment significantly explained the welfare gaps between rural and urban households.

Another study by Skoufias and Katayama (2011) examined the welfare difference between metropolitan, urban, and rural households in Brazil's five regions. The study employed the Oaxaca-Blinder method on a sample of households from the 2003-2004 Household Budget Survey to decompose welfare differences both at the mean and at different quantiles of the welfare distribution. The findings revealed that the welfare differences between metropolitan, urban, and rural households are mainly attributed to differences in endowments between households residing in these geographical areas. Differences in the household head's education explained about 40% of the welfare difference between metropolitan

[5]These are the 2008/2009, 2010/2011, 2012/2013 and 2014/2015 Household Income, Expenditure and Consumption Surveys (HIECS) for Egypt, the 2005 and 2010 National Survey on Household Budget, Consumption and Standard of Living (EBCNV) for Tunisia, the 2006, 2010 and 2013 Household Expenditure and Income Survey (HEIS) for Jordan and 2007, 2010 and 2011 Palestine Expenditure and Consumption Surveys (PECS) for Palestine

and urban areas. Nevertheless, quantile decomposition revealed that this effect of education on welfare differences is heterogenous along the welfare distribution.

This essay aims not to replicate the reviewed studies above on the link between education and welfare. Instead, it contributes to this literature by focusing on a rarely exploited socio-demographic dimension, namely the age group (youth), and second by examining how differences in educational attainment and returns to education are associated with differences in welfare.

5.3 Empirical Strategy

This Chapter's empirical methodology follows Firpo et al. (2009, 2018)'s RIF based decomposition for the mean difference between two groups, derived in Chapter 4. As explained in Chapter 4, for a given dependent variable Y and independent variables X, RIF decomposition uses RIF regression in combination with reweighting to decompose any statistic of interest into two parts: the difference due to endowments (characteristics/explained/composition effect) and the difference attributed to the relationship between Y and X (coefficient/unexplained/return effect). It goes further to decompose the contribution of each explanatory variable on the two parts.

In this essay, the aim is to decompose the welfare differences between cohorts to two parts: the part that can is attributable to the cohort differences in characteristics (in our case education is the focus) and the part attributable to returns to these characteristics (again the focus is on returns to education). It is the decomposition of welfare between youth in 2001 (pre reform) and their counterpart in 2018 (post reform cohort) that shed light on the effectiveness of the large expansion of education. If a significant part of the differences in welfare between the two youth cohorts can be attributed to the differences in educational attainment between the cohorts, then we can argue that the program succeeded to improve welfare through increasing educational attainment.

The baseline regression in this case is the standard household consumption model of the form:

$$lnC_{it} = \alpha S_{it} + \beta X_{it} + \varepsilon_{it} \tag{5.1}$$

Where C is the household consumption to poverty line ratio (CPL) — our preferred measure of welfare (described in detail in the next section); S a vector of schooling of the household head and its square (in years); X is a vector (including a constant) of individual/household characteristics; α and β are regression parameters; ε is standard error term; and i and t index individual and time, respectively. With exogeneity assumption, (5.1) is usually estimated using OLS.

For any two groups, RIF decomposition uses the reweighted parameter estimates from (5.1) to decompose the statistic of interest into two parts as explained earlier in Chapter 4. As a recap, we briefly explain the decomposition method here, and refer the reader to Chapter 4 Section 4.3 for the detailed derivation.

For simplicity and mathematical convenience, rewrite (5.1) in the form

$$Y = X'\beta + \varepsilon \tag{5.2}$$

Where X here is a vector of covariates, including years of education and its square. Suppose there is some categorical variable R or T such that the joint distribution function of Y, X and R is given by $f_{(Y,X,R)}(y_1, x_i, R_i)$ and that of Y, X and T by $f_{(Y,X,T)}(y_1, x_i, T_i)$.

When there are only two groups in R and T, such that $R \in [0, 1]$ and $T \in [0, 1]$, e.g. in our case R and T are indicator variables for the groups of interest defined by

$$R = \begin{cases} 1, & \text{if } youth \\ 0, & \text{if } adult \end{cases}$$

and

$$T = \begin{cases} 1, & \text{if } 2018 \\ 0, & \text{if } 2001 \end{cases}$$

For simplicity of derivation and without loss of generalisation, we will stick to one categorical variable, T. The joint distribution function between the measure of welfare, the covariates and T for $T\!=\!k \in [0,1]$ is given as:

$$f^k_{Y,X}(y,x) = f^k_{Y|X}(Y|X)f^k_X(X) \tag{5.3}$$

and its its cumulative distribution function conditional on T as:

$$F^k_Y(y) = \int f^k_{Y|X}(Y|X)dF^k_X(X) \tag{5.4}$$

The cumulative distribution of Y conditional on T can then be used to decompose the difference in the distribution of statistic v between the two groups. Accordingly,

$$\Delta v = v_1 - v_0 = v(f^1_Y) - v(f^0_Y) \tag{5.5}$$

With some counterfactual statistic v_c, we can rewrite (5.5) as

$$\Delta v = v_1 - v_c + v_c - v_0$$

$\Delta v_S = v_1 - v_c$ is the difference attributed to the relationship between Y and X; and

$\Delta v_X = v_c - v_0$ the difference arising due to differences in characteristics, the Xs.

From $v(F_Y) = X'\beta$,

After estimating the counterfactual statistic from the data, the final decomposition can then be rewritten as

$$\Delta v = \Delta v^p_S + \Delta v^e_S + \Delta v^p_X + \Delta v^e_X \tag{5.6}$$

The component $\Delta v^p_S + \Delta v^e_S$ in (5.6) is called the coefficients/explained/characteristics effect which constitutes of the pure(coefficients/explained/characteristics) effect (Δv_S^p) and the reweighting error (Δv_S^e). The component $\Delta v_X^p + \Delta v_X^e$ is called the composition/unexplained/returns effect and constitutes the pure(composition/unexplained/returns) effect (Δv_X^p) and specification error (Δv_X^e). The empirical estimation of the RIF decomposition for the mean of log consumption to poverty line ratio is performed in Stata using user-written command *Oaxaca_rif* (Rios-Avila, 2020b).

5.4 Data and Descriptive Statistics

5.4.1 Data Source and Sample

Consumption and income remain the most popular measures of welfare in economics. In developing countries where data on income from agricultural and non-wage informal (self-employed) employment are rarely available, welfare measures based on consumption are the most suitable Deaton (2018). Unlike the labour force surveys that do not have information on consumption, HBS and the Tanzania National Panel Survey (TNPS) collect information on household consumption. However, TNPS are smaller surveys relative to HBS and only available from 2008 – 2019 making them less suitable for our analysis since the aim is to compare welfare of the youth before and after UPE reform. HBS on the other hand are available from 1992/93 – 2018, and the 2001 and 2018 HBS allow us to compare youth 'pre and post treatment'.

As consumption in HBS is measured at household level, we assign it to the head of the household and thus comparison is between groups of households distinguished by the age of the head – youth (aged 15-35) who benefited from UPE by 2018 and adults who didn't. In addition, education in our case is the household head's level of education measured in years. Our analysis focuses on the household head, considering the household head follows the literature given the absence of suitable household-level measures of education, especially as we wish to separate those who benefitted from UPE. A limitation of distinguishing households based on the age of the head is that education may be endogenous to household formation by youth. Unfortunately, the HBS does not include suitable data to model the formation of households (thus we do not pursue this in our analysis).

This study, therefore, uses data from the Tanzanian Household Budget Surveys for 2001 and 2018, which we obtained from the National Bureau of Statistics. HBS is among the largest household surveys in Tanzania, covering all regions of the Mainland[6]. Data collection for HBS 2001 took place from May 2000 to June 2001 and for HBS 2018 from December 2017 to November 2018. Both surveys employed a multi-stage cluster sampling to obtain representative samples of 22,176 and 9,552 households in 2001 and 2018, respectively. Despite the sample for 2018 being significantly smaller than its 2001 counterpart, the sampling mechanism still ensured representativeness at the national (Mainland) level (URT, 2019). A total of 154 households in 2001 had missing information on assets ownership and were excluded from the analysis, leaving us with a sample of 22,022 households. All households in 2018 had complete information.

5.4.2 Definition of the Main Variables

Household characteristics

- *CPL*: As stated earlier, household consumption per adult equivalent is widely used as a proxy of well-being. In this study we use the ratio of household consumption per adult equivalent to the national poverty line (CPL). We employ this approach to account for inflation between the survey period given the absence of good price deflators. Its logarithm is used as the dependent variable in the base OLS regression and used to construct the RIF for the RIF decomposition.

- *poor*: a dummy variable = 1 for households below the basic needs poverty line and 0 otherwise

[6]Tanzania (also the United Republic of Tanzania) includes the Tanzania Manland (Tanganyika) and the islands of Zanzibar. The Mainland covers about 99% of the total area and about 98% of the total population

- *rural*: a dummy variable = 1 for households resides in rural area and 0 otherwise

- *hhsize*: Total number of usual members in the household

Household head characteristics

- *education*: years of schooling of the household head

- *noeducation*: a dummy variable = 1 if household head completed less than three years of primary education and 0 otherwise

- *someprimary*: a dummy variable = 1 if household head completed at least four and at most six years of primary education and 0 otherwise

- *primary*: a dummy variable = 1 if household head completed the seven years of primary education and 0 otherwise

- *somesecondary*: a dummy variable = 1 if household head completed at least two and at most three years of secondary education and 0 otherwise

- *secondary*: a dummy variable = 1 if household head completed the four years of lower secondary education and 0 otherwise

- *postsecondary*: a dummy variable = 1 if household head has more than lower secondary education and 0 otherwise

- *age*: age of the household head in years

- *female*: a dummy variable = 1 if the head of the household is female and 0 otherwise

- *married*: a dummy variable = 1 if the head of the household is married and 0 otherwise

5.4.3 Descriptive Statistics

Tables 5.1 and 5.2 show the means for the continuous variables, and the percentages of the respective group's observations for the dummy variables, included in the analysis. Table 5.1 shows the characteristics of youth and adult headed households for each year. The share of households headed by a youth decreased by ten percentage points from 36% in 2001 to 26% in 2018. Youth headed households have significantly higher consumption and lower poverty rates than adult-headed households in both years, although the differences are smaller in 2018.

Table 5.1: Summary Statistics for the Main Variables by Age Group and Year (Within Year)

Variable Name	2001			2018		
	Youth	Adult	Difference	Youth	Adult	Difference
Hh Characteristics						
CPL	1.71	1.40	-0.31***	2.12	1.85	-0.27***
poor	0.28	0.41	0.13***	0.20	0.28	0.09***
rural	0.80	0.80	0.00	0.66	0.69	0.03**
hhsize	5.18	6.89	1.70***	4.75	6.59	1.84***
Head Characteristics						
education	6.04	4.24	-1.80***	6.73	5.59	-1.14***
noeducation	0.18	0.42	0.24***	0.19	0.27	0.08***
someprimary	0.05	0.18	0.13***	0.08	0.10	0.02**
primary	0.71	0.33	-0.37***	0.49	0.52	0.03*
somesecondary	0.01	0.00	-0.01***	0.06	0.01	-0.05***
secondary	0.05	0.05	0.00	0.14	0.08	-0.07***
postsecondary	0.01	0.02	0.01***	0.04	0.02	-0.01**
age	29.86	51.46	21.60***	30.10	52.27	22.17***
female	0.18	0.20	0.02*	0.19	0.25	0.06***
married	0.85	0.80	-0.05***	0.83	0.78	-0.06***
Observations	8,039	13,983	-	2,507	6,945	-

Source: Author's calculations from HBS 2001 and 2018 data. Dummy variables are in proportions. 'Difference' is the mean for adult that heads household minus the corresponding value for youth that heads household. Statistics are weighted using survey weights. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ based on Lincom's test of mean differences.

Density plots in Figure 5.1 also show the differences in consumption between the two age groups. In terms of education endowment, heads defined as youth have more schooling than their adult counterparts in both years. The difference in post-primary education attainment between the two age groups significantly increased between the years. These results reflect the benefits of the expansion of secondary education in the mid and late 2000s.

Figure 5.1: Distribution of Household Consumption between Age Groups by Year

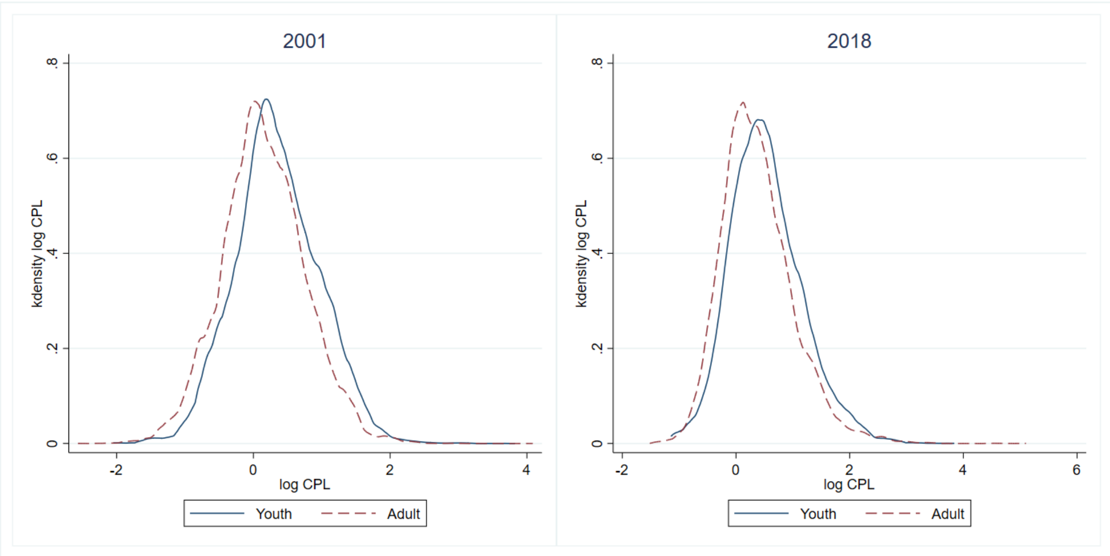


Table 5.2 presents exactly the same means but in this table compares the characteristics of household heads between the years (for ease of reading the means have been repeated). Table 5.2 shows that between 2001 and 2018, the youth's average household consumption increased, and poverty rates declined significantly. The increase in youth consumption between the two periods is shown graphically using a density plot by the left panel of Figure 2.

Table 5.2 further shows that youth education increased significantly, with the most pronounced increase at post-primary education levels. The share of youth with completed secondary education increased by about threefold, from 5% in 2001 to 14% in 2018, and the share with higher education by fourfold from 1% to 4%. There is a significant proportion of secondary school students who drop out of school (6% of the youth in 2018). The national qualifying exam at the second year of secondary school which requires students who fail to repeat the year may be one of the main factors that explain this but further exploration is beyond this essay's scope.

Table 5.2: Summary Statistics for the Main Variables by Age and Year (Between Years)

Variable Name	Youth			Adults		
	2001	2018	Difference	2001	2018	Difference
Hh Characteristics						
CPL	1.71	2.12	0.41***	1.40	1.85	0.45***
Poor	0.28	0.20	-0.08***	0.41	0.28	-0.13***
Rural	0.80	0.66	0.16***	0.80	0.69	-0.11***
Hhsize	5.18	4.75	0.43***	6.89	6.59	-0.30***
Head Characteristics						
education	6.04	6.73	0.69***	4.24	5.59	1.35***
noeducation	0.18	0.19	0.01	0.42	0.27	-0.15***
someprimary	0.05	0.08	0.03***	0.18	0.10	-0.08***
primary	0.71	0.49	-0.22***	0.33	0.52	0.19***
somesecondary	0.01	0.06	0.05***	0.00	0.01	0.01***
secondary	0.05	0.14	0.09***	0.05	0.08	0.03***
postsecondary	0.01	0.04	0.03***	0.02	0.02	0.00
Age	29.86	30.10	0.24	51.46	52.27	0.81*
Female	0.18	0.19	0.01	0.20	0.25	0.05***
Married	0.85	0.83	-0.02	0.80	0.78	-0.02*
Observations	8,039	2,507	-	13,983	6,945	-

Source: Author's calculations from HBS 2001 and 2018 data. Dummy variables are in proportions. Difference is the mean for 2018 minus the corresponding value for 2001. Statistics are weighted using survey weights. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ based on Lincom's test of mean differences.

A complication in the table for adult headed households is the fact that adults in 2018 may have been in the youth category in 2001. For completeness the mean difference is included in Table 5.2 but Table 5.3 provides a better and more intuitive comparison for the adult households by grouping them into two. The first group consists of those in the youth category in 2001 (aged 35 to 53 years in 2018). The second group consists of those in the adult category in 2001 (aged 54 years and above in 2018). Table 5.3 shows how these two groups fare relative to adult headed households in 2001. We know from Table 5.1 that the youth in 2001 had more educational attainment than their adult counterparts and so we would expect adults aged 35 to 53 years in 2018 to have more education than those aged 54 years and above, which happens to be the case. The right panel of Figure 5.2 shows the distribution of consumption reported in Table 5.3. Both groups of adult headed households in 2018 have significantly higher consumption than adult headed households in 2001. However, the 35 to 53 years age group enjoys slightly higher consumption than their 54 years and above counterparts.

Table 5.3: Differences in the Means of the Variables between Survey Years for the Youth

	(1)	(2)	(3)	(4)
Variable Name	2001 Age >35	2018 35<Age<=53	2018 Age>53	Difference (1)-(3)
Hh Characteristics				
CPL	1.40	1.86	1.83	0.43***
poor	0.41	0.29	0.28	-0.13***
rural	0.80	0.67	0.71	-0.09***
hhsize	6.89	6.56	6.64	-0.25**
Head Characteristics				
education (years)	4.24	6.28	4.51	-0.27**
noeducation	0.42	0.18	0.40	-0.02*
someprimary	0.18	0.06	0.15	-0.03**
primary	0.33	0.64	0.35	-0.02
somesecondary	0.00	0.01	0.01	0.01*
secondary	0.05	0.08	0.07	0.02**
postsecondary	0.02	0.03	0.02	0.00
age	51.46	43.92	65.27	13.81**
female	0.20	0.22	0.31	0.11***
married	0.80	0.84	0.68	-0.12***
Observations	13,983	3,966	2,979	NA

Source: Author's calculations from HBS 2001 and 2018 data. Dummy variables are in proportions. Difference is the value in column (3) minus the corresponding value in (1) comparing youth in 2001 to 2018 adults that were adult also in 2001. Statistics are weighted using survey weights. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ based on Lincom's test of mean differences.

Similarly, we investigated whether youth who are heads of households differ significantly in characteristics from those who are not. Table 5.4 compares the characteristics of these two groups of youth by year. In both years, youth who do not head their household live in households with lower consumption and higher poverty rates. However, the educational endowment in the two groups seems to have changed between the years. Unlike 2001, youth who are not household heads in 2018 have more schooling than household heads.

Figure 5.2: Distribution of Household Consumption between Years by Age Group

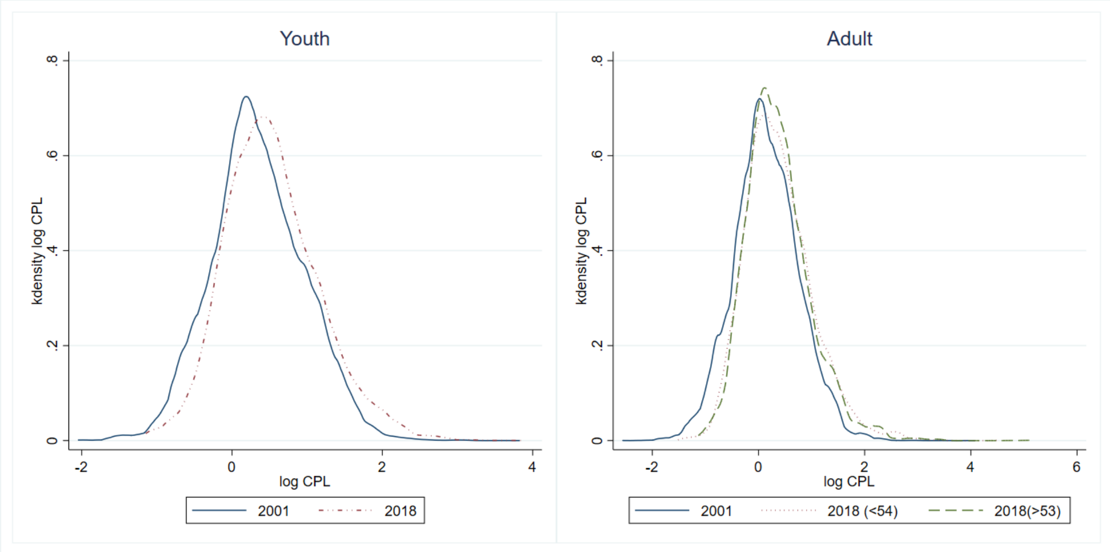


Table 5.4: Characteristics of the Youth by Year and Headship Status

Var Name	2001			2018		
	Head	Other	Difference	Heads	Other	Difference
Hh Characteristics						
CPL	1.71	1.47	0.24***	2.12	1.95	0.17***
poor	0.28	0.38	-0.10***	0.20	0.25	-0.05***
rural	0.80	0.78	0.02***	0.66	0.63	0.03***
hhsize	5.18	6.83	-1.65***	4.75	6.51	-1.76***
Youth Characteristics						
education	6.04	5.48	0.56***	6.73	7.00	-0.27***
noeducation	0.18	0.22	-0.04***	0.19	0.16	0.03***
someprimary	0.05	0.15	-0.10***	0.08	0.08	0.00
primary	0.71	0.58	0.13***	0.49	0.47	0.02***
somesecondary	0.01	0.02	-0.01***	0.06	0.10	-0.04***
secondary	0.05	0.03	0.02***	0.14	0.17	-0.03***
postsecondary	0.01	0.00	0.01***	0.04	0.02	0.02***
age	29.86	22.74	7.12***	30.10	22.38	7.72***
female	0.18	0.65	-0.47***	0.19	0.62	-0.43***
married	0.85	0.46	0.39***	0.83	0.38	0.45***
Observations	8,039	31,503	-	2,507	11,468	-

Source: Author's calculations from HBS 2001 and 2018 data. Dummy variables are in proportions. Difference is between youth headed households and other (youth who are not household heads) in 2001 and 2018. Statistics are weighted using survey weights. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ based on Lincom's test of mean differences.

The findings in 5.4 are consistent with the fact that the share of households headed by youth decreased by ten percentage points between 2001 and 2018, implying that as post-primary school enrolment rose, more youth are spending more years in education, and as a result, the youth who move out to establish households (and become heads) at early ages are those with relatively low education. As this might bring about a selection problem and affect our results, we also examine how education affects youth's likelihood to head their households. The results for that analysis are included in Section 5.

5.5 Results and Discussion

5.5.1 Main Results and Discussion

The first part of our analysis explores whether welfare returns to education for youth are different from that of adults; and whether they have changed between 2001 and 2018. For each year and by age group OLS regression estimates for model (1) were obtained. Table 5.5 presents the regression results in each category. In 2001 schooling is positively and significantly correlated with welfare for both youth and adult heading households, but negatively correlated for 2018. The coefficient of schooling squared is positive and highly statistically significant across age groups and years, implying a strong convex relationship between education and welfare—each extra year of schooling is associated with higher welfare than the previous year. All other included regressors have the expected sign.

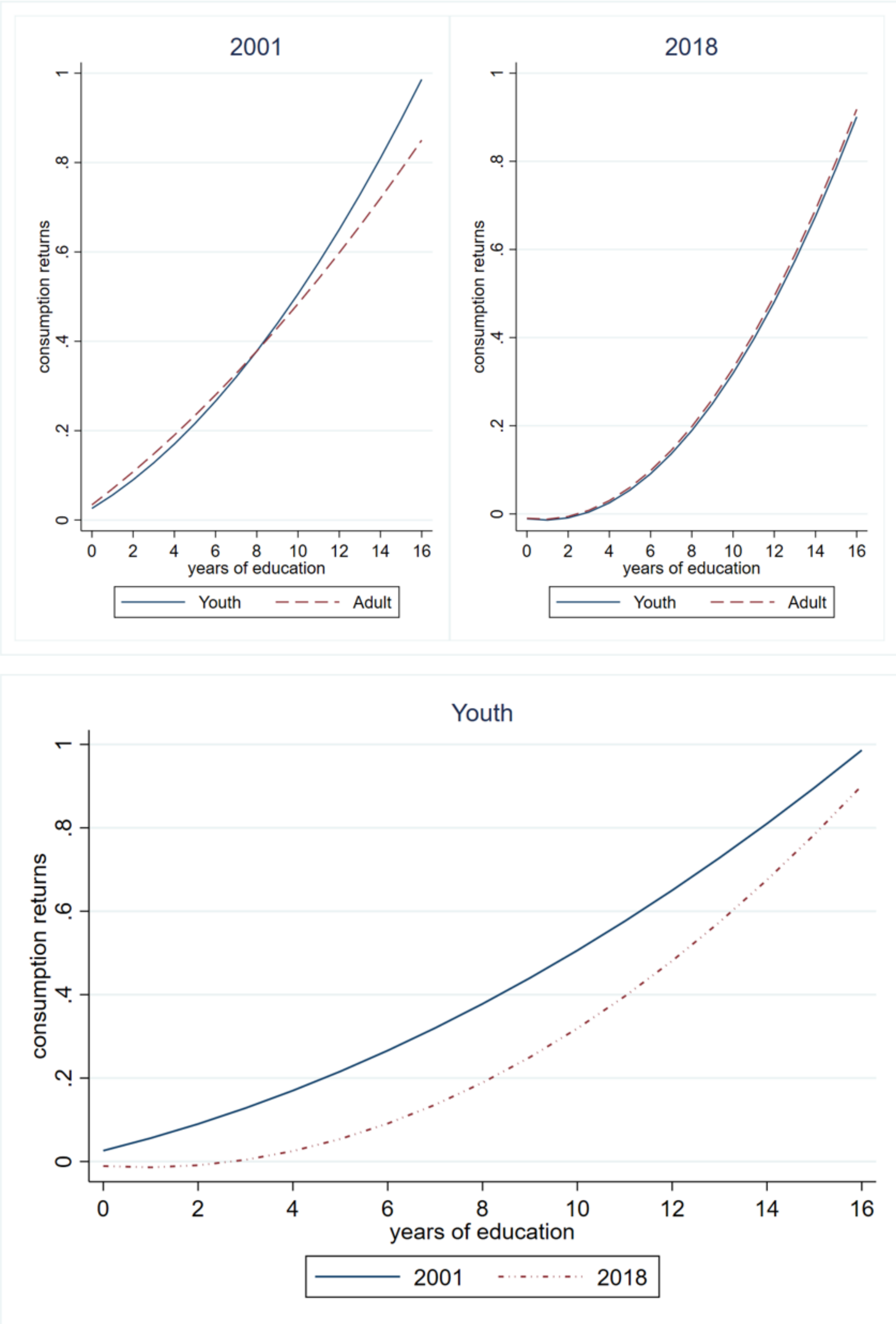
As the presence of the square term may complicate the interpretation of the coefficients of schooling variables in Table 5.5, the predicted average marginal effects of schooling is added in the table. In addition, we plot the implied welfare returns to each year of education from the results in Table 5.5 in Figure 5.3 and focus the interpretation on it. The top panel of Figure 5.3 shows that youth heading households in 2001 had higher returns to post-primary education than adults heading households, but this advantage disappeared in 2018. The bottom panel of Figure 5.3, on the other hand, shows that the welfare returns to education for the youth declined significantly between 2001 and 2018. These results may be attributed to the significant gains in schooling for the youth over this period—one would expect the returns to education to decline as education attainment increases in the population.

Table 5.5: OLS Regression Estimates of Returns to Education by Age Group and Year

	2001		2018	
	Youth	Adult	Youth	Adult
sch	0.026***	0.034***	-0.011*	-0.010***
	(0.005)	(0.003)	(0.006)	(0.004)
sch2	0.002***	0.001***	0.004***	0.004***
	(0.000)	(0.000)	(0.000)	(0.000)
age	0.059***	-0.011***	0.030	0.002
	(0.016)	(0.003)	(0.028)	(0.003)
age2/100	-0.098***	0.008***	-0.045	-0.002
	(0.028)	(0.002)	(0.050)	(0.003)
female	0.102***	0.010	0.085***	0.054***
	(0.015)	(0.014)	(0.028)	(0.018)
rural	-0.131***	-0.147***	-0.160***	-0.183***
	(0.014)	(0.011)	(0.025)	(0.016)
married	0.103***	0.071***	0.049	0.052***
	(0.016)	(0.014)	(0.030)	(0.019)
lnhhsize	-0.516***	-0.451***	-0.523***	-0.473***
	(0.011)	(0.008)	(0.021)	(0.011)
Constant	-0.005	1.153***	0.331	0.656***
	(0.219)	(0.082)	(0.396)	(0.099)
Others controls	Yes	Yes	Yes	Yes
AME(sch)	0.056***	0.043***	0.042***	0.027***
	(0.002)	(0.001)	(0.003)	(0.002)
Obs.	8,039	13,983	2,507	6,945
R2	0.41	0.40	0.50	0.44

Notes: AME(sch) is the marginal effects of schooling. Other controls included are livestock per capita, dummies for region of residence and ownership of assets. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 5.3: Implied Returns to Education by Age Group and Year



The second part focuses on the results from the Reweighted RIF Oaxaca-Blinder decomposition within the years. The difference in mean household welfare between youth and adult headed households for each year is decomposed into two parts as explained in section 5.3: the part due to differences in characteristics/endowment (also called the explained part) and the part due to differences in returns to these characteristics (also called the unexplained part). Each of the two parts are then broken down into two subparts: the explained part into pure explained and the specification error; and the unexplained part into pure unexplained and reweighing error. A explained earlier in section 5.3, for more robust results both the specification and the reweighing error should be small and insignificant (Firpo et al., 2018; Rios-Avila, 2020b) implying that the model is correctly specified and reweighed.

Table 5.6 presents the decomposition results by year. To simplify interpretation, the coefficients of the education variables (*sch* and *sch2*) are aggregated[7] into one variable 'education'; the coefficient of (*age* and *age2*) into 'headage'; and ownership of assets, livestock per capita, and dummies for regions of residence into 'other controls'[8]. The top panel of Table 5.6 shows the contribution of the explained and unexplained parts to the total difference in log welfare. Only the explained component is significant in both years implying that it is only the difference in characteristics/endowment that explains differences in welfare between the two age groups. Both the specification and reweighing errors are insignificant as expected.

The breakdown of the *Pure_explained* component in Table 5.6 reveals that the coefficient on education in the explained component is positive and significant, suggesting that the youths heading households have significantly better education attainment than adults heading households, consistent with what we observed in Table 5.1 in the previous section.

[7]The Stata command ***oaxaca_rif*** is calibrated for that option.
[8]This is common approach in Oaxaca-Blinder decomposition literature and fits with the specification.

Table 5.6: Reweighted RIF Oaxaca-Blinder Decomposition Within Years

	(1) 2001	(2) 2018
Overall		
Youth	0.346***	0.531***
Counterfactual	0.093*	0.378***
Adult	0.149***	0.374***
Difference	0.197***	0.157***
Explained	0.252***	0.153***
Unexplained	-0.055	0.004
Pure_explained	0.251***	0.151***
education	0.101***	0.048***
headage	0.000	-0.002
female	0.000	-0.001**
rural	0.002*	0.002**
married	0.000	0.004***
lnhhsize	0.154***	0.145***
Other controls	-0.008	-0.049***
Pure_Unexplained	-0.422	0.643
education	0.026	0.014
age	0.207	-0.321
female	-0.011	-0.001
rural	-0.008	0.028
married	-0.023	0.081*
lnhhsize	-0.244*	-0.157
Other controls	-0.015	-0.024
constant	-0.404	1.057
Specification error	0.002	0.002
Reweight error	0.366	-0.639
N1	8,039	2,507
N2	13,983	6,945
N	22,022	9,452

'Other controls' is the aggregate effect of livestock per capita, dummies for region of residence and ownership of assets. Dummy variables are normalized. N1 is sample size for youth, N2 sample size for adult and N total sample/observations (youth + adults). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on 'Youth' and 'Adults' implies that the mean for each group is significantly different from their combine mean.

The findings in 5.6 show that a significant portion of the welfare difference between youth and adult headed households is attributable to differences in educational attainment between youths and adults that head households. Precisely, of the *pure_explained* welfare differences of 0.251 and 0.151 in 2001 and

2018 respectively, approximately 40% and 32%[9] are attributed to differences in educational attainment between youths heading households and adults heading households. In other words, if an adult had the same level of educational attainment as a youth heading a household, their welfare would have been about 40% higher in 2001 and 32% higher in 2018. The difference in returns to education, however, does not have a significant effect on welfare. This is consistent with the regression results in Table 5.5 and the top panel of Figure 5.3, which shows small differences in returns to education between the two age groups.

The third part of our analysis focuses on the Reweighted RIF Oaxaca-Blinder decomposition results for the youth heading households between 2001 and 2018. It is this part of the analysis, that is the decomposition of welfare between youth in 2001 (pre reform) and their counterpart in 2018 (post reform cohort) that shed light on the effectiveness of the large expansion of education. If a significant part of the differences in welfare between the two youth cohorts can be attributed to the differences in educational attainment between the cohorts, then we can argue that the program succeeded to improve welfare through increasing educational attainment.

To assess if there is heterogeneity of the association between the difference in education and difference in welfare between the two periods, we perform the RIF decomposition by gender and place of residence. Table 5.7 reports the estimated coefficients of education for this decomposition. The results in Table 5.7 suggest that the difference in welfare between the two periods is mainly attributed to differences in characteristics[10]. Also, the results provide evidence of heterogeneity of the effects of education across groups.

The results in column (1) of Table 5.7 suggest that if the youth in 2001 had the same education endowment as their 2018 counterparts, their welfare would have been about 20% higher. Therefore, it implies that other things equal, policies that contributed to the increase in education attainment led to improved youth

[9] i.e., 0.101 out of 0.251 and 0.048 out of 0.151 in 2001 and 2018 respectively

[10] The covariates included in the baseline regression.

welfare. The results in columns (2) and (3) suggest that although between 2001 and 2018 welfare increased more for males than females, the welfare increase attributed to the increase in education was significantly higher for females than males. Furthermore, results in columns (4) and (5) suggest that education played a more significant role in increasing the rural youth's welfare than their urban counterparts.

5.5.2 Robustness Checks

Endogeneity of education from unobserved ability is one of the primary concerns of our model. Another concern is the potential endogenous youth selection into the households, whether the factors associated with higher welfare such as education are also associated with a higher likelihood of the youth to head the household they live in. Whereas the methods to address both issues are well documented in the literature, the debate on how (if at all possible) to combine these methods with the RIF decomposition methods remains inconclusive (Firpo et al., 2018; Rios-Avila, 2020a). This shortcoming notwithstanding, and without trying to include the selection equation in the RIF decomposition model, we use a linear probability model (LPM) to assess whether education increases/reduces youth's likelihood to head the household. While the problems of LPM are well documented in the literature, we nonetheless, prefer it as it allows the inclusion of household fixed effects.

Tables 5.8 shows that, after controlling for household fixed effects, more educated youth are significantly more likely to be head of household. Precisely, an extra year of education is associated with about 0.005 increase in youth's probability of heading the household in which they live. However, the results in Table 5.9 reveals that after disaggregating the results by year, gender and place of residence, more educated youth are less likely to be household heads in 2018 than in 2001 (an extra year increases the probability of being head by 0.003 and 0.006 in 2018 and 2001 respectively). Table 5.9 further reveals that the significant association between education and headship in 2018 is generally driven by youths

Table 5.7: Reweighted RIF Oaxaca-Blinder Decomposition for Youth

	(1) Pooled	(2) Female	(3) Male	(4) Rural	(5) Urban
Overall					
2018	0.531***	0.547***	0.527***	0.379***	0.820***
Counterfactual	0.299***	0.326***	0.314***	0.127***	0.552***
2001	0.346***	0.379***	0.339***	0.257***	0.711***
Difference	0.185***	0.168*	0.188***	0.121***	0.108
Explained	0.232***	0.221**	0.212***	0.251***	0.268***
Unexplained	-0.047	-0.053	-0.024	-0.130***	-0.160
Pure_explained	0.269***	0.171*	0.271***	0.225***	0.349***
education	0.055***	0.075***	0.055***	0.023**	0.094***
age	0.000	0.001	0.001	0.001	0.003
female	0.000			-0.001	0.000
rural	-0.001	0.001	-0.004		
married	0.002	-0.005	0.001	0.003	-0.001
lnhhsize	0.016	0.006	-0.001	0.021	-0.032
Other controls	0.196***	0.096	0.222***	0.176***	0.287***
Pure_Unexplained	-0.100*	-0.095	-0.108*	-0.115***	-0.18
education	-0.190	-0.305*	-0.178	-0.261***	-0.181
age	0.490	0.657	0.096	0.991	-0.887
female	-0.002			0.005	0.008
rural	0.006	-0.059	0.036		
married	-0.003	0.014	0.062	0.064	-0.034
lnhhsize	0.111	-0.046	-0.034	0.025	0.003
Other controls	0.168	-0.115	0.420*	0.384***	-0.017
constant	-0.689	-0.248	-0.498	-1.289	0.943
Specification error	-0.037	0.050	-0.059	0.027	-0.081
Reweight error	0.053	0.042	0.083	-0.015	0.020
N1	2,507	473	2,034	1,653	854
N2	8,039	1,920	6,119	2,687	5,352
N	10,546	2,393	8,153	4,340	6,206

'Other controls' is the aggregate effect of livestock per capita, dummies for region of residence and ownership of assets. Dummy variables are normalized. N1 is sample size for 2018, N2 sample size for 2001 and N total sample/observations (2001 + 2018). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The significance of coefficients on '2018' and '2001' implies that the mean for each group is significantly different from their combine mean.

residing in rural areas.

Despite the significance of the coefficients of education in Tables 5.8 and 5.9 , their sizes are noteworthy. The coefficients of education that average at 0.005 suggest that the probability of becoming head of household increases by only 0.5% for every year increase in schooling. Such a small estimated effect is less likely to significantly affect the estimated coefficients of education in our main results. While this signals a potential selection problem, which we acknowledge that we do not have a remedy for, we argue the low coefficient gives us cautious confidence in the main results.

Table 5.8: Pooled LPM Regression Results by Gender and Place of Residence

	(1) Pooled	(2) Female	(3) Male	(4) Rural	(5) Urban
sch	0.005***	0.003**	0.003	0.006***	0.004***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)
youth<26	-0.239***	-0.113***	-0.190***	-0.235***	-0.236***
	(0.005)	(0.009)	(0.015)	(0.008)	(0.007)
female	-0.316***			-0.368***	-0.269***
	(0.004)			(0.007)	(0.006)
married	0.085***	-0.030***	0.317***	0.108***	0.071***
	(0.006)	(0.008)	(0.018)	(0.010)	(0.009)
AGR	-0.004	-0.001	0.002	0.001	-0.024***
	(0.005)	(0.007)	(0.011)	(0.008)	(0.007)
WAGE	0.183***	0.099***	0.069***	0.164***	0.191***
	(0.009)	(0.017)	(0.022)	(0.020)	(0.010)
SELF	0.154***	0.106***	0.084***	0.070***	0.185***
	(0.008)	(0.017)	(0.018)	(0.016)	(0.010)
Constant	0.410***	0.130***	0.298***	0.426***	0.390***
	(0.008)	(0.012)	(0.019)	(0.012)	(0.012)
Obs.	53,517	29,231	24,286	23,206	30,311
R2	0.72	0.89	0.93	0.70	0.74

Note: AGR, WAGE and SELF are dummies for sectors of main employment, standing for agriculture, wage employment and self-employment out of agriculture, respectively. <26 is an age dummy =1 if aged less than 26 years and 0 otherwise.
Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5.9: LPM Regression Results by Year, Gender and Place of Residence

	2001					2018				
	All	Female	Male	Rural	Urban	All	Female	Male	Rural	Urban
sch	0.006***	0.003**	0.003	0.007***	0.005***	0.003**	0.001	0.001	0.004**	0.003
	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
youth<26	-0.244***	-0.118***	-0.196***	-0.243***	-0.240***	-0.220***	-0.093***	-0.152***	-0.220***	-0.215***
	(0.006)	(0.010)	(0.017)	(0.010)	(0.008)	(0.011)	(0.021)	(0.027)	(0.013)	(0.021)
female	-0.302***			-0.374***	-0.257***	-0.348***			-0.356***	-0.332***
	(0.005)			(0.008)	(0.006)	(0.009)			(0.011)	(0.016)
married	0.082***	-0.033***	0.351***	0.117***	0.069***	0.091***	-0.021	0.220***	0.095***	0.086***
	(0.007)	(0.009)	(0.022)	(0.012)	(0.009)	(0.013)	(0.018)	(0.032)	(0.015)	(0.027)
AGR	-0.019***	-0.001	-0.001	-0.017*	-0.034***	0.042***	0.012	0.026	0.029**	0.081**
	(0.006)	(0.008)	(0.012)	(0.011)	(0.007)	(0.012)	(0.018)	(0.021)	(0.012)	(0.035)
WAGE	0.182***	0.109***	0.070***	0.172***	0.189***	0.163***	0.051	0.062	0.152***	0.179***
	(0.010)	(0.018)	(0.025)	(0.026)	(0.011)	(0.020)	(0.044)	(0.044)	(0.032)	(0.025)
SELF	0.156***	0.109***	0.097***	0.045**	0.184***	0.134***	0.094***	0.023	0.097***	0.173***
	(0.010)	(0.020)	(0.021)	(0.023)	(0.011)	(0.016)	(0.035)	(0.034)	(0.022)	(0.023)
Constant	0.419***	0.139***	0.296***	0.454***	0.397***	0.382***	0.103***	0.294***	0.387***	0.371***
	(0.010)	(0.014)	(0.023)	(0.016)	(0.012)	(0.017)	(0.025)	(0.036)	(0.019)	(0.033)
Obs.	39,542	21,873	17,669	13,486	26,056	13,975	7,358	6,617	9,720	4,255
R^2	0.72	0.89	0.92	0.69	0.73	0.72	0.90	0.95	0.71	0.75

Note: AGR, WAGE and SELF are dummies for sectors of main employment, standing for agriculture, wage employment and self-employment out of agriculture, respectively. <26 is an age dummy =1 if aged less than 26 years and 0 otherwise. Robust standard errors in parentheses * p < 0.10, ** p < 0.05, *** p < 0.01.

5.6 Conclusion

In this essay, we examined how much of the welfare differences between youths and adults, and between the youth in 2001 and their 2018 counterparts, can be attributed to differences in educational attainment and differences in returns to education. Samples of household heads from the 2001 and 2018 HBSs were investigated using RIF decomposition of the mean.

We find evidence that youths, having more education than adults, enjoy higher welfare levels than adults in both years. The difference in educational attainment between the two groups significantly explains the differences in welfare, while the difference in education returns does not. We also find, compared to their 2001 counterparts, the youth in 2018 have higher education and welfare levels. The findings revealed that the difference in welfare is significantly attributed to differences in educational attainment between 2001 and 2018. However, despite the evidence of the substantial decline in returns to education between 2001 and 2018, we find no evidence that such a decline reduced welfare.

We did not control for endogeneity of education from ability bias in our welfare decomposition as no good proxies for ability are available in the HBS data. Combining IV strategy with RIF decomposition is not warranted either, even if we had good instruments for education from the data. Moreover, the methods to deal with sample selection problem in the context of RIF decomposition are not explicit in the literature (Firpo et al., 2018; Rios-Avila, 2020a). Nonetheless, as a robustness check, we investigate if education is associated with a young household member's likelihood to head the household and find that the strength of the association is minimal. This finding leads us to believe that our results are robust to selection.

Note that there are other methods are potentially available to determine the effect of the UPE reform on both education attainment and welfare. A difference-in-differences approach which compared young and old cohorts before and after the reform, or a regression discontinuity which exploited the sharp cut-off at particular ages are possibilities, but we leave that for future research.

Chapter 6

Conclusion

6.1 Summary

This thesis employed different approaches in analysing the link between education and the labour market in East Africa. The main contribution of the thesis is to investigate if returns to education and the effect of education on the distribution of earnings and gender wage gaps varies according to workers' pay period. In East Africa, workers report earnings over three main pay periods: daily, weekly, or monthly. These periods are associated with the type and duration of employment whereby workers paid monthly are more likely to work in formal and regular employment while those paid daily and weekly are more likely to work in informal, casual and piece-rate jobs. As pay periods may indicate distinct labour markets, these three pay periods are used to categorise workers into three groups (with separate analysis of *ganyu* workers in Malawi).

To address the well-known endogeneity concerns from ability bias and sample selection in Chapter 3, Gaussian Copula (GC) and Heckman with Gaussian Copula (HGC) estimators were employed. Gaussian Copula (Park and Gupta, 2012) is an instrument free method for recovering estimates which are free from endogeneity by directly modelling the correlation between the endogenous regressor and the error term in the regression. By modelling the endogenous education within the sample selection model (Heckman, 1979), the ability bias and selection problems are simultaneously addressed. The GC and HGC estimators, therefore, offer a valid method to address the concerns and the results appear robust.

The empirical evidence from the first essay suggests that pooling workers paid over different periods as in previous studies for Africa leads to biased estimates of returns to education. Pooling/aggregating earnings to different common measures (daily, monthly or annualised expressed monthly) produce different estimates of returns. Estimating separately for each period, the findings reveal that returns to education differ by pay period: for Malawi, daily has the highest, and weekly has the lowest returns to education. For Tanzania and Uganda, the weekly has the highest and monthly has the lowest returns to education. These results hold regardless of the estimation strategy employed (OLS, GC and HGC) or the earnings function specification (quadratic schooling or dummies for levels of education). The evidence, therefore, suggests that estimating returns separately for workers paid over different periods is a better strategy than pooling all workers/periods together to capture the segment of workers not paid monthly (or more generally to capture informal sector workers).

An extension provides a separate analysis of returns to education for the casual and piece-rate labour market in Malawi, referred to as *ganyu* labour. In line with the above analysis, this section explored how using different measures of earnings affects the estimates of returns to education in *ganyu* labour. In line with earlier results, the findings for *ganyu* show that, generally, converting to monthly yields larger and inefficient estimates of returns to education than converting to daily or annualised. Furthermore, like the pooled results, the difference between converting to daily and converting to annualised estimates is small.

Recentered Influence Function (RIF) regression and decomposition techniques remain among the modern econometric tools for analysing the effects of a change in the distribution of a variable on the statistic of interest (Rios-Avila, 2020b). The second essay employed RIF regression and decomposition strategies (Firpo et al., 2009, 2018) to examine three essential aspects. Firstly, for each pay period, it begins by showing how an increase in the population's education by one year would affect earnings at different points of the wage distribution. Secondly, it goes further to investigate how such an increase in education would affect the earnings gap between the high-wage and the low-wage earners. Lastly, it decomposes the gender wage and inequality differences to ascertain the proportion that can be attributed to both gender differences in educational attainment and in returns to education.

Evidence from the RIF regression (assuming linear schooling) suggests that while estimates significantly differ within pay periods depending on the quantile of earnings, they also differ across the pay periods. Only the weekly sample has a consistent pattern within pay periods whereby returns increase monotonically by quantile of earnings for all three countries. Across the pay periods, generally, the effect of an additional year of education in the population on average earnings is larger for workers reporting earnings monthly compared to their daily and weekly counterparts. This suggests that returns to education (or skill) are more important for formal sector workers.

The RIF regression for inequality shows that education can either increase or reduce wage inequality depending on the period in which the worker is paid. Education is associated with an increase in inequality for workers paid weekly and reduced inequality for those paid daily and monthly. That is, conditional on working and paid weekly, education is likely to benefit more those in high than in low paying jobs, while for the other pay periods it benefits more those in low than in high paying jobs.

Allowing for non-linearity in schooling, the results reiterate that an increase in education in the population generates different earnings outcomes depending on the pay period. However, this specification reveals that the outcomes differ depending on the level of education. Precisely, the relationship between earnings and education is concave for workers whose earnings are in the bottom 10% and strongly convex for workers whose earnings are in the top 10% of the earnings distribution. This suggests that, in all three countries, an increase in education in the population is more likely to benefit the high-wage than the low-wage workers and hence likely to increase inequality. For workers in the top 10% of the earnings distribution quantiles, the effects of education are very small (even negative for

Malawi) for early years of schooling but increase rapidly after about the 6th year, regardless of the pay period.

The gender earnings gap was decomposed to assess how much of it can be attributed to gender gaps in educational attainment as well as in returns to education, again by pay period. Significant gender gaps in earnings were found for Tanzania and Uganda but not for Malawi. Results from reweighted RIF OB decomposition suggest that gender differences in educational attainment significantly explain the gender wage gap for Tanzania and Uganda, implying that policies that increase women's educational attainment are vital in narrowing the gender wage gap.

Chapter 4 also extended the analysis to include a section for *ganyu* labour in Malawi. Examining the distributional effects of education on earnings for *ganyu* workers, the findings reveal a pattern of results like those in the primary analysis. An increase in the population's average education by a year significantly increases the mean wage of *ganyu* workers by an amount that varies depending on the quantile of the earnings distribution. Reweighted RIF OB gender decomposition results further show that a significant proportion of the wage gap in *ganyu* labour can be attributed to the gender differences in educational attainment in both urban and rural areas.

The third essay (Chapter 5) focuses on investigating how much of the welfare difference between both youth-headed households in 2001 and 2018 and between youth and adult-headed households in each of the years can be attributed to differences in educational attainment and returns to education. The aim was to assess the impact of increased participation in education, especially following the Universal Primary Education (UPE) introduced in 2001, which mainly benefited the youth aged 15 – 35 years in 2018. Contrasted with previous studies which mainly examined the association between education and welfare at any given point in time, this essay examined both how much of the welfare differences between 2001 and 2018 can be attributed to changes in the association between education and welfare over this period, and how much can be attributed to changes in educational attainment between 2001 and 2018 (i.e., effects of change in education distribution between the period).

Proxying welfare by household (per adult equivalent) consumption expenditure relative to the national poverty line, the study used a reweighted RIF OB decomposition, a similar methodological approach to that employed in Chapter 4, to decompose the welfare differences between youth in 2001 and their 2018 counterparts as well as between youth and adults in each year. As expected, the analysis shows that youths, having more education than adults, enjoy higher welfare levels than adults in both years. The difference in educational attainment between the two groups significantly explains the differences in welfare, but the difference in the returns to education does not. Comparing youth cohorts across years, the youth in 2018 have higher education and welfare levels than their 2001 counterparts. The findings revealed that the difference in welfare is significantly attributed to differences in educational attainment between 2001 and 2018. Differences in returns to education explain the welfare gap only for young women and youth residing in rural areas.

6.2 Policy Implications

The thesis' findings provide some important policy implications. The findings of all empirical chapters provide new evidence to justify that the returns to education in East Africa are positive, and thus the efforts to increase both education attainment and achievement should be upheld. The findings suggest that returns to education in the region are convex, implying that higher education levels are becoming more important. With all countries having UPE policies in place (and tuition-free secondary education in Tanzania and Uganda), the governments of the respective countries should ensure that school enrolment, completion and grade progression rates remain high and ideally increase.

The findings from Chapter 4 also show that females still lie behind males in terms of educational attainment and raising female's educational attainment can significantly reduce the gender earnings gap. Education policies should therefore be geared towards increasing girls' enrolment into all levels of education.

We can also point out some policy implications with regards to the role of the informal sector in inequality; the inequality between men and women; and the inequality which arises due to education reform although these need further analyses. First, the higher inequality among*ganyu* workers who are mainly in informal employment compared to the other workers suggests that moving the workers out of the informal employment may reduce earnings inequality in the region. Second, there is a significant gender earnings gap for *ganyu* workers in Malawi but not for the other category of workers. This suggests that women are more disadvantaged while working in the informal sector and thus policies targeted at moving women out of the informal employment may curb the gender earnings gap. Third, the UPE reform that is found to have increased earnings and welfare in Tanzania may be associated with increased earnings inequality between the pre reform cohort and the post reform cohort. In designing policies to reduce earnings inequality, policy makers should therefore take this reform into consideration.

6.3 Limitations of the Study

Due to limitations with the LSMS data, we lack variables that we could have used as controls[1] for unobservable ability or instruments for endogenous schooling. Consequently, while the results appear robust we could not compare or check robustness using common strategies such as two-stage least squares and control functions. Furthermore, although data for Tanzania and Uganda have a panel dimension, the proportion of workers observed more than once is relatively small. Using panel data estimation strategies like fixed effects to address the endogeneity would have created another sample selection bias problem.

In Chapters 4 and 5, no attempt was made to control for endogeneity of

[1] Such as IQ tests and standardized test grades or parental education.

education from the unobserved ability or endogenous selection to the different pay periods, which is a limitation of the analyses. As stated earlier, the absence of suitable instrumental variables from the data is one factor. Most importantly, even with good instruments or instrument-free methods like GC, the debate on the combination of RIF and the endogeneity control methods remains inconclusive (Rios-Avila, 2020a; Firpo et al., 2018). In that regard, we leave that open to future studies when such methods become widely available.

The analysis in Chapter 5 is based on comparisons of youth and adult headed households, given the difficulty in measuring education and consumption of individuals at the household level. While youth make the largest proportion of the labour force, less than a third of them head the households they live in. As a result, we could not analyse the welfare gain resulting from gains in schooling for the youth who live in households headed by adult.

6.4 Future Research

This thesis is the first study in Sub-Saharan Africa (SSA) to empirically show that the pay period matters, and care should be taken when estimating returns to education for workers paid over different pay periods. Building on the findings of this thesis, further exploration on the topic could be carried out using data covering more countries and with more detailed labour market information. This should help to propose standard adjustment factors that could be used by all researchers to convert earnings from one period to another or pooling for comparison, thereby making studies across the region more comparable.

The findings from Chapter 3 point that moving the population to wage employment (formal employment) with monthly payments may see greater returns to education. Future research could extend the current analysis to investigate the differences in returns to education between workers paid monthly who work in the formal sector and those in the informal sector.

The findings from Chapter 4 show that education can significantly reduce the gender earnings gap. However, we lacked appropriate data that would be required to investigate how policies geared towards increasing girls' enrolment into all levels of education, including free education and allowing pregnant girls back to school should be enhanced, and to what extent the policies would bring about gender equality in education and eventually labour market outcomes. Future research could extend the current analysis if better household and individual data become available.

As stated earlier, the analysis in Chapter 5 was based on comparisons of youth (and adult) headed households, given the difficulty in measuring education and consumption of individuals at the household level. Future studies could use other surveys to analyse how the gains in schooling over this period affected the wage earnings and their distribution, with analysis at the individual level.

References

Agesa, R. U., Agesa, J., and Dabalen, A. (2013). Sources of the Persistent Gender Wage Gap along the Unconditional Earnings Distribution : Findings from Kenya. *Oxford Development Studies*, 41(1):76–103.

Agyire-Tettey, F., Ackah, C. G., and Asuman, D. (2018). An Unconditional Quantile Regression Based Decomposition of Spatial Welfare Inequalities in Ghana. *Journal of Development Studies*, 54(3):537–556.

Al-Samarrai, S. and Reilly, B. (2008). Education, employment and earnings of secondary school and university leavers in Tanzania: Evidence from a tracer study. *Journal of Development Studies*, 44(2):258–288.

Alem, Y. and Söderbom, M. (2012). Household-Level Consumption in Urban Ethiopia: The Effects of a Large Food Price Shock. *World Development*, 40(1):146–162.

Arouri, M., Nguyen, C., and Youssef, A. B. (2015). Natural Disasters, Household Welfare, and Resilience: Evidence from Rural Vietnam. *World Development*, 70:59–77.

Arsalan, A., Tschirley, D., and Egger, E.-M. (2019). What Drives Rural Youth Welfare? The Role of Spatial, Economic, and Household Factors. *International Fund for Agricultural Development (IFAD)*, 42.

Aslam, M., Bari, F., and Kingdon, G. (2012). Returns to schooling, ability and cognitive skills in Pakistan. *Education Economics*, 20(2):139–173.

Ayyash, M. and Sek, S. K. (2020). Decomposing Inequality in Household Consumption Expenditure in Malaysia. *Economies*, 8(4):83.

Baffour, P. T. (2013). Determinants of Urban Worker Earnings in Ghana and Tanzania: The Role of Education. Credit research paper 13/01, School of Economics, University of Nottingham, Nottingham.

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to schooling*. Columbia University Press, New York.

Belghith, N. B. H., Karamba, W., Talbert, E., and de Boisseson, P. (2020). *TANZANIA MAINLAND POVERTY ASSESSMENT: Tanzania's Path to Poverty Reduction and Pro-Poor Growth*. The World Bank, Washington D.C.

Blinder, A. S. (1973). Wage Discrimination : Reduced Form and Structural Estimates. *The Journal of Human Resources*, 8(4):436–455.

Borjas, G. J. (2016). *Labor economics*. McGraw-Hill Education, New York, 7 edition.

Bound, J., Brown, C., Duncan, G. J., and Rodgers, W. L. (1994). Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data. *Journal of Labor Economics*, 12(3):345–368.

Bridges, S., Fox, L., Gaggero, A., and Owens, T. (2017). Youth unemployment and earnings in Africa: Evidence from Tanzanian Retrospective Data. *Journal of African Economies*, 26(2):119–139.

Bryceson, D. F. (2006). Ganyu casual labour, famine and HIV/AIDS in rural Malawi: Causality and casualty. *Journal of Modern African Studies*, 44(2):173–202.

Cameron, C. A. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press, New York.

Cameron, C. A. and Trivedi, P. K. (2009). *Microeconometrics using Stata*. Stata Press., Texas.

Card, D. (1999). The Causal Effect of Education on Earnings. In Ashenfelter, O. and Card, D., editors, *Handbook of Labour Economics*, volume 3, chapter 30, pages 1801–1863. Elsevier B.V, Amsterdam.

Card, D. (2001). Estimating the returns to schooling: progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004). *Copula Methods in Finance*, volume 53. John Wiley & Sons Ltd, West Sussex.

CHuang, Y. and Lai, W. (2017). Returns to Human Capital and Wage Inequality: The Case of Taiwan. *Journal of Economic Development*, 42(3):61–88.

Cobham, A., Schlögl, L., and Sumner, A. (2016). Inequality and the Tails: The Palma Proposition and Ratio. *Global Policy*, 7(1):25–36.

Cuaresma, J. C. and Raggl, A. (2016). The dynamics of returns to education in Uganda: National and subnational trends. *Development Policy Review*, 34(3):385–422.

Deaton, A. (2018). *The analysis of household surveys.* The World Bank, Washington D.C, reissue edition.

Delesalle, E. (2019). THE EFFECT OF THE UNIVERSAL PRIMARY EDUCATION PROGRAM ON LABOR MARKET OUTCOMES: EVIDENCE FROM TANZANIA. NBER Working Paper No. 25789, NBER, Cambridge, MA.

Ebbes, P., Wedel, M., Böckenholt, U., and Steerneman, T. (2005). Solving and Testing for Regressor-Error (in) Dependence When no Instrumental Variables are Available : With New Evidence for the Effect of Education on Income. *Quantitative Marketing and Economics*, 3:365–392.

Essama-Nssah, B., Paul, S., and andre Bassolé, L. (2013). Accounting for heterogeneity in growth incidence in cameroon using recentered influence function regression. *Journal of African Economies*, 22(5):757–795.

Farré, L., Klein, R., and Vella, F. (2013). A parametric control function approach to estimating the returns to schooling in the absence of exclusion restrictions: An application to the NLSY. *Empirical Economics*, 44(1):111–133.

Fichtenbaum, R. (2006). Labour market segmentation and union wage gaps. *Review of Social Economy*, 64(3):387–420.

Firpo, S. P., Fortin, N. M., and Lemieux, T. (2009). Unconditional Quantile Regressions. *Econometrica*, 77(3):953–973.

Firpo, S. P., Fortin, N. M., and Lemieux, T. (2018). Decomposing wage distributions using recentered influence function regressions. *Econometrics*, 6(2):1–41.

Fortin, N., Lemieux, T., and Firpo, S. (2011). Decomposition Methods in Economics. In Ashenfelter, O. and Card, D., editors, *Handbook of Labor Economics*, volume 4.A, pages 1–102. Elsevier Inc., Amsterdam.

Fulford, S., College, B., and Hill, C. (2014). Returns to education in India. *World Development*, 59:434–450.

Girma, S. and Kedir, A. (2005). Heterogeneity in returns to schooling : Econometric evidence from Ethiopia. *The Journal of Development Studies*, 41(8):1405–1416.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Applied Econometrics*, 31(3):129–137.

Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006). Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond. In Hanushek, E. A., Machin, S. J., and Woessmann, L., editors, *Handbook of the Economics of Education*, volume 1, chapter 7, pages 307–458. Elsevier B.V.

Heckman, J. J., Lochner, L. J., and Todd, P. E. (2008). Earnings Functions and Rates of Return. *Journal of Human Capital*, 2(1):1–31.

Heckman, J. J. and Vytlacil, E. (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica*, 73(3):669–738.

Heckman, J. J. and Vytlacil, E. J. (2007). Econometric Evaluation of Social Programs, Part I: Causal Models, Structural Models and Econometric Policy Evaluation. In *Handbook of Econometrics*, volume 6B, pages 4779–4874.

Himaz, R. and Aturupana, H. (2018). Schooling and household welfare: The case of Sri Lanka from 1990 to 2006. *Review of Development Economics*, 22(2):592–609.

Hult, G., Proksch, D., Sarstedt, M., Pinkwart, A., and Ringle, C. (2018). Addressing Endogeneity in International Marketing Applications of Partial Least Squares Structural Equation Modeling. *Journal of International Marketing*, 26(3):1–21.

Joseph, C. and Leyaro, V. (2019). Gender Differential Effects of Technical and Vocational Training : Empirical Evidence for Tanzania. Credit research paper 19/04, School of Economics, University of Nottingham, Nottingham.

Kahyarara, G. and Teal, F. (2008). The Returns to Vocational Training and Academic Education: Evidence from Tanzania. *World Development*, 36(11):2223–2242.

Kavuma, S. N., Morrissey, O., and Upward, R. (2015). Private returns to education for wage-employees and the self-employed in Uganda. Wider working paper 2015/021, UN-WIDER, Helsinki.

Khan, R. and Morrissey, O. (2020). Income diversification and household welfare in Tanzania 2008 - 13. Wider working paper 2020/110, UNU-WIDER, Helsinki.

Kilic, T., Palacios-López, A., and Goldstein, M. (2015). Caught in a Productivity Trap: A Distributional Perspective on Gender Differences in Malawian Agriculture. *World Development*, 70:416–463.

Kim, S. B. (2020). Gender earnings gap among the youth in Malawi. *African Development Review*, 32(2):176–187.

Klein, R. and Vella, F. (2009). Estimating the Return to Endogenous Schooling Decisions via Conditional Second Moments. *The Journal of Human Resources*, 44(4):1047–1065.

Klein, R. and Vella, F. (2010). Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics*, 154(2):154–164.

Kuepié, M. and Nordman, C. J. (2016). Where Does Education Pay Off in Sub-Saharan Africa? Evidence from Two Cities of the Republic of Congo. *Oxford Development Studies*, 44(1):1–27.

Kwenda, P. and Ntuli, M. (2018). A detailed decomposition analysis of the public-private sector wage gap in South Africa. *Development Southern Africa*, 35(6):815–838.

Lassibille, G. and Tan, J. P. (2005). The Returns to Education in Rwanda. *Journal of African Economies*, 14(1):92–116.

Lewbel, A. (2012). Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models. *Journal of Business & Economic Statistics*, 30(1):67–80.

Mincer, J. (1974). *Schooling, experience, and earnings. Human behavior and social institutions.* Columbia University Press, New York.

Mishra, V. and Smyth, R. (2015). Estimating returns to schooling in urban China using conventional and heteroskedasticity-based instruments. *Economic Modelling*, 47:166–173.

National Statistical Office (2019). 2018 Malawi Population and Housing Census Main Report. Technical report, National Statistical Office, Zomba.

National Statistical Office (2020). Fourth Integrated Household Survey (IHS4). Retrieved 3 June 2020, from http://www.nsomalawi.mw/index.php?option=com_content&view.

Nielsen, H. S. and Rosholm, M. (2001). The public-private sector wage gap in Zambia in the 1990s: A quantile regression approach. *Empirical Economics*, 26(1):169–182.

Nikolov, P. and Jimi, N. (2018). What factors drive individual misperceptions of the returns to schooling in Tanzania ? Some lessons for education policy. *Applied Economics ISSN:*, 50(44):4705–4723.

Nix, E., Gamberoni, E., and Heath, R. (2016). Bridging the Gender Gap: Identifying What Is Holding Self-employed Women Back in Ghana, Rwanda, Tanzania, and the Republic of Congo. *World Bank Economic Review*, 30(3):501–521.

Nordman, C. J., Robilliard, A. S., and Roubaud, F. (2011). Gender and ethnic earnings gaps in seven West African cities. *Labour Economics*, 18(SUPPL. 1):S132–S145.

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review*, 14(3):693–709.

Park, S. and Gupta, S. (2012). Handling endogenous regressors by joint estimation using copulas. *Marketing Science*, 31(4):567–586.

Patrinos, H. A. and Psacharopoulos, G. (2010). Returns to education in developing countries. *International Encyclopedia of Education*, pages 305–312.

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.

Peichl, A. and Pestel, N. (2015). *Earnings Inequality*, volume 6. Elsevier, second edi edition.

Pischke, J.-s. (1995). Measurement Error and Earnings Dynamics : Some Estimates From the PSID Validation Study. 13(3):305–314.

Psacharopoulos, G. and Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics*, 12(2):111–134.

Psacharopoulos, G. and Patrinos, H. A. (2018). Returns to investment in education: a decennial review of the global literature. *Education Economics*, 26(5):445–458.

Ramadan, R., Hlasny, V., and Intini, V. (2018). Inter-Group Expenditure Gaps In The Arab Region And Their Determinants: Application To Egypt, Jordan, Palestine And Tunisia. *Review of Income and Wealth*, 64(October):S145–S188.

Rios-Avila, F. (2020a). Oaxaca_rif & standard vs. reweighted model. Retrieved 25 January 2021, from https://www.statalist.org/forums/forum/general-stata-discussion/general/1564820.

Rios-Avila, F. (2020b). Recentered influence functions (RIFs) in Stata: RIF regression and RIF decomposition. *Stata Journal*, 20(1):51–94.

Rutz, O. J. and Watson, G. F. (2019). Endogeneity and marketing strategy research: an overview. *Journal of the Academy of Marketing Science*, 47(3):479–498.

Salisbury, T. (2016). Education and inequality in South Africa: Returns to schooling in the post-apartheid era. *International Journal of Educational Development*, 46:43–52.

Schultz, T. P. (1961). Investment in Human Capital. *The American Economic Review*, 51(1):1–17.

Schultz, T. P. (2004). Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education. *Journal of African Economies*, 13(suppl_2):ii95–ii148.

Serneels, P., Beegle, K., and Dillon, A. (2017). Do returns to education depend on how and whom you ask? *Economics of Education Review*, 60:5–19.

Skoufias, E. and Katayama, R. S. (2011). Sources of welfare disparities between and within regions of Brazil: Evidence from the 2002-2003 household budget survey (POF). *Journal of Economic Geography*, 11(5):897–918.

Söderbom, M., Teal, F., Eberhardt, M., Quinn, S., and Zeitlin, A. (2014). *Emperical Development Economics*. Routledge, London and New York, 1 edition.

Stefani, P. C. and Biderman, C. (2009). The evolution of the returns to education and wage differentials in Brazil : a quantile approach. *Applied Economics*, 41(11):1453–1460.

URT (2002). *Household Budget Survey 2000/01*. National Bureau of Statistics, Dar es Salaam.

URT (2005). *Basic Education Statistics in Tanzania (BEST) 1995 - 2005*. Ministry of Education and Culture, Dar es Salaam.

URT (2007). *NATIONAL YOUTH DEVELOPMENT POLICY*. MINISTRY OF LABOUR, EMPLOYMENT AND YOUTH DEVELOPMENT, Dar es Salaam.

URT (2008). *The Tanzania Economic Survey 2007*. The Ministry of Finance and Economic Affairs, Dar es Salaam.

URT (2013). *The economic survey 2012*. Ministry of Finance, Dar es Salaam.

URT (2016). *National Basic Education Statistics in Tanzania (BEST) 2012 - 2016*. Ministry of Education, Science and Technology, Dar es Salaam.

URT (2017). *National Panel Survey Wave 4.* National Bureau of Statistics, Dar es Salaam.

URT (2018). *National population projections.* National Bureau of Statistics and Office of the Chief Government Statistician, Dar es Salaam, Zanzibar.

URT (2019). *Household Budget Survey 2017-18: Key Indicators Report.* National Bureau of Statistics, Dar es Salaam.

Van de Walle, D. (2013). Lasting welfare effects of widowhood in mali. *World Development*, 51:1–19.

Verbeek, M. (2004). *A Guide to Modern Econometrics.* John Wiley & Sons Ltd, West Sussex, 2 edition.

Wang, L. (2013). Estimating returns to education when the IV sample is selective. *Labour Economics*, 21:74–85.

Whiteside, M. (2000). Ganyu Labour in Malawi and its Implications for Livelihood Security Interventions – An Analysis of Recent Literature and Implications for Poverty Alleviation. *Agricultural Research and Extension Network*, 99:1–10.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, Cambridge, Massachusetts, London.

World Bank (2020a). Living Standards Measurement Study. Retrieved 23 April 2021, from https://www.worldbank.org/en/programs/lsms/overview.

World Bank (2020b). *Poverty & Equity Brief Sub-Saharan Africa.* The World Bank, Washington D.C.