**CREDIT Research Paper**

**No. 22/02**

# Universal Primary Education and Household Welfare in Tanzania

**by**

**Livini Donath, Oliver Morrissey and Trudy Owens**

**Abstract**

This is an update of Donath *et al*. (*CRP 21/04*) which assessed the effect of the 2001 Universal Primary Education (UPE) in Tanzania on the welfare difference between youth (aged 15-35 according to the official definition) and adult (aged over 35) headed households in 2018. As anybody aged over 25 in 2018 would not have derived full benefit of UPE from 2001, this paper examines whether the welfare difference comparing 2001 and 2018 between households headed by youth aged 15-25 and adults (aged over 35) is attributable to differences in educational attainment following the nationwide 2001 UPE. Household budget survey data for 2001 and 2018 are used to estimate the effect of education on household welfare (measured as consumption relative to the poverty line), availing of the fact that youth in 2018 (aged 15-25) will have benefitted from the UPE. Decomposition analysis reveals that the increase in youth educational attainment by 2018 is a significant factor explaining the increase in welfare of youth headed households between 2001 and 2018. If the youth in 2001 had the same education endowment as their 2018 counterparts, their relative welfare would have been about a quarter higher. The findings also show that differences in educational attainment are significant factors explaining differences in welfare between youth and adults in each year. If adults had the same level of educational attainment as the youth, their welfare would have been about a third higher in 2001 and 2018. Expanding access to education had a positive effect on welfare.

**JEL Classification:** I21, O55

**Keywords:** Universal Primary Education, youth, household welfare, Tanzania.

**Centre for Research in Economic Development and International Trade, University of Nottingham**

**CREDIT Research Paper**

**No. 22/02**

# Universal Primary Education and Household Welfare in Tanzania

**by**

**Livini Donath, Oliver Morrissey and Trudy Owens**

Outline



**The Authors**

Livini Donath is a Lecturer in the Department of Geography and Economics, Mkwawa University College of Education, University of Dar es Salaam (P.O. Box 2513, Iringa, Tanzania). Oliver Morrissey and Trudy Owens are in the School of Economics, University of Nottingham and are both CREDIT Fellows.

*Corresponding Author:* livini.donath@nottingham.ac.uk /livinitesha@gmail.com

**Acknowledgements**: The research was undertaken while Dr Donath was a PhD student at Nottingham (funded by a Commonwealth Scholarship).

# Universal Primary Education and Household Welfare in Tanzania

## 1. Introduction

The nationwide free and Universal Primary Education (UPE) of 2001 in Tanzania and the massive secondary school expansion program (known as 'ward secondary schools') that began in 2006 are the two most significant reforms to educational access in Tanzania in over 30 years. There was a significant increase in enrolment across all levels: the primary school gross enrolment ratio increased from 84% in 2001 to 98.6% in 2002 and then to 109.9% in 2005.[1] During the same period, the number of primary schools increased from 11,873 in 2001 to 12,286 (2002) and then to 14,257 in 2005 (United Republic of Tanzania [URT], 2005). Secondary school gross enrolment increased from 20% in 2006 to 30.5% in 2007 and 37% in 2012, while the number of secondary schools increased from 2,289 in 2006 to 3,485 (2007) and 4,528 in 2012 (URT, 2008, 2013, 2016). This large expansion potentially affected all school-age children. Although the merit of expanding and easing access to education is widely accepted (Colclough *et al.*, 2003), there is limited evidence that students benefit. Available evidence suggests that UPE has at best limited effects on educational attainment (Al-Samarrai, 2006; Kan & Klasen, 2021). However, Delesalle (2021) found that the earlier Tanzanian UPE reforms in the 1970s increased earnings in 2002 of heads of households likely to have benefited. This paper investigates if the more extensive UPE in 2001 increased incomes of the individuals likely to have benefited.

This period of expanded access to education was also a period of improvements in household welfare. Poverty incidence according to national poverty lines declined from 36% in 2001 to 26% in 2018 (URT, 2002, 2019). Poverty in terms of household consumption expenditure, a commonly used measure for income and welfare in developing countries (De Janvry & Sadoulet, 2016; Deaton, 2018), also declined. The ratio of per adult household consumption to the national poverty line improved from 1.79 in 2001 to 2.28 in 2018, equivalent to a 27% increase.[2] If and to what extent did the expansion of access to education in the early 2000s contribute to the improvement in welfare by 2018? To address this important question on the impact of UPE involves comparing a set of individuals who benefitted from the reforms to a set who did not, a difficult challenge given the limited data available for earnings of individuals. As a proxy to compare youth income before the reforms with youth income after, the approach adopted here compares the welfare of households according to the age cohort of the head using the 2001 and 2018 Tanzania Household Budget Surveys (HBS). The HBS are chosen because they have relatively good data for nationally representative samples of households (22,022 in 2001 and 9,552 in 2018 for the analysis).

The primary focus of this paper is to examine the extent to which differences in educational attainment between the youth cohort aged between 15 and 25 in 2001, who would have completed primary school before the UPE reforms, and the corresponding cohort in 2018, who should have benefited from the reforms (as few if any started primary school before 2002), explain the welfare difference between the two periods. This is a narrower definition of youth than the official Tanzanian definition of those aged 15 to 35, but is chosen to restrict analysis to those who could have derived full benefit from the 2001 reforms (Donath, Morrissey &

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

2 Authors' own calculations from the Tanzania HBS 2001 and 2018 (URT, 2002; URT, 2019).

Owens, 2021b consider the effect for the broader official definition).[3] Whilst it is expected that the youth aged 15–25 in 2018 (the cohort that mainly benefited from the UPE and subsequent expansion of secondary education) should have more education than the cohort aged over 25 (referred to as adults for convenience although the official definition for adults is over 35), it is not known if the increase in education was associated with increased earnings and welfare.[4]

In other words, it is unknown whether the increase in access and provision also increased educational attainment and earnings for the youth and also if there was an effect on returns to education. These two effects are not the same: beneficiaries of UPE should be more educated and therefore have higher earnings, but as the number of educated workers increases the returns to a given level of education may decline. Using data from the HBS, the analysis compares youth headed households in 2001 and 2018 and compares youth to adult headed households in both years. As consumption in HBS is measured at the 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 and education is the household head's level of education measured in years. Restricting the focus to household heads is a limitation but that is the welfare (income in terms of per capita consumption spending) measure in the HBS and available data are inadequate to estimate income or earnings of individuals compared to 2001.

Delesalle (2021) is the most comprehensive study using national household surveys to address the effect of education expansion on household welfare in Tanzania. She considers the UPE programme of 1974-78 aimed at levelling access to education across districts so expansion was concentrated in districts that were initially disadvantaged. As the timing and scale of that expansion varied across districts depending on initial status, with little or no expansion in districts initially well provided with schools, the potential benefit varied accrding to distrct. Identification is thus based on age and location at the time of reforms as determining the potential treatment effect, exploiting variation in the scale of expansion over time and residence district. She estimates that predicted consumption in 2002 of household heads exposed to 1974-78 programmes was increased by six per cent. Although we also consider effects for household heads, that identification strategy is not feasible as UPE 2001 did not vary intentionally in timing or intensity across districts. However, rather than using methods to predict consumption our approach avails of household consumption reported in the HBS.

There is evidence of positive returns to education in developing countries, through higher earnings and facilitating access to higher-paid wage employment, although estimates of the magnitude vary (Psacharopoulos & Patrinos, 2018). For Tanzania, recent estimated returns based on four waves of the Tanzania National Panel Surveys from 2008 to 2016 are 20-30% for completing primary education, about 80% for secondary education and up to 130% for post-secondary education (Donath, Morrissey & Owens, 2021a). Using the 2014 Tanzanian Integrated Labour Force Survey, Leyaro & Joseph (2019, Table 7) estimate that completing secondary education increases earnings by about 40%, and completing tertiary increases earnings by over 100%, compared to primary education. However, these studies do not capture

3 The age range considered as youth varies by institution and country, e.g., United Nations (15-24), African Union (15-35), Uganda (18-30). The official Tanzanian definition, which is in line with that of the African Union, defines youth as all males and females aged 15 to 35 years (URT, 2007). On this definition, youth account for approximately 65% of the total labour force (URT, 2015, 2018). We present statistics for this broad definition of youth in Appendix A.

4 In Tanzania, students typically have seven years of primary school beginning at age 7 so the age group 15-25 in 2018 are considered as the 'treated' group. As some youth aged over 25 in 2018 may have benefitted, by repeating a year or two of primary school or from secondary school expansion, Appendix B presents results of analysis for broad youth (15-35) and the older youth (26-35) to compare with the analysis for treated youth (15-25) on which we focus in the paper.

the effects of large-scale reforms in access to education, and the labour force surveys provide limited samples of youth that benefitted from UPE and do not go back to the early 2000s (see Appendix A). The HBS provides reasonable and nationally representative samples in 2001 and 2018; although the sample size is smaller in the latter survey it is representative for Mainland Tanzania (URT, 2019).

The paper makes two contributions to the literature on the link between education and household welfare. First, studies of trends in welfare in Tanzania have focussed on differences according to gender, employment, and rural/urban categories (Belghith, Karamba, Talbert & de Boisseson, 2020; Khan & Morrissey, 2020), and less attention has been paid to age groups (specifically, youth versus adults). This paper examines how the expansion of education between 2001 and 2018 is associated with household welfare changes for youth (aged 15-25) compared to adult headed households over the period. Secondly, this paper examines at the household head level the attribution of welfare differences between 2001 and 2018 to changes in amount of schooling completed (endowments) and to changes in returns to education. As expansion, especially UPE, increases the average education level of entrants to the labour market there may be effects on returns to education.

To separate the effect of endowments and returns, the analysis employs decomposition analysis using the recentered influence function (RIF); this shows that differences in educational attainment between youth and adults significantly explain the difference in welfare between the two groups in both years. If adults had the same level of educational attainment as the youth, their welfare would have been a third 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 a quarter higher. Although there appears to have been a decline in returns to education for the youth, consistent with a growing educated labour force, this adverse effect was more than compensated by increased endowments, so welfare rose.

Section 2 reviews selected related literature from developing countries, with a focus on Tanzania, 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.

## 2. Related Literature

The literature on the effects of UPE tends to focus on educational outcomes and most studies do not address effects on earnings or households. For example, Kan & Klasen (2021) investigate the effect of Uganda's free primary education (the main component of UPE) from 1997 on educational attainment using data from household surveys over 2005 to 2014. They found no significant effect on the number of years in, or likelihood of completing, primary school, although there was some improvement in progression to secondary school. This is consistent with literature showing that UPE, even if enrolment increases, has little if any effect on educational attainment, and public education expenditure is not clearly correlated with educational outcomes (Al-Samarrai, 2006).

The various studies on determinants of household welfare in developing countries include some that focus on the effects of shocks (Alem and Söderbom, 2012; Arouri, Nguyen & Youssef, 2015), some on inequalities and distribution across households (Agyire-Tettey, Ackah & Asuman, 2018; Ramadan, Hlasny & Intini, 2018; Skoufias & Katayama, 2011), and some on education (Himaz & Aturupana, 2018). Studies for SSA often focus on the effects of household labour diversification into non-farm employment (Davis, Di Guiseppe and Zezza, 2017; Van den Broeck and Kilic, 2019). Factors such as education, age, gender, race, employment status, sector of employment, place of residence and rural-urban migration have been found to be

correlates or determinants of welfare, although typically individual characteristics are only for the household head. Female-headed households, households with higher proportions of female members, residing in rural areas and/or engaged in agriculture have lower welfare.

Few studies focus explicitly on youth. In an exception, Aslan, Tschirley & Egger (2021) investigate the relative welfare of youth (in the age range 15-24) using data on 85 low and middle income countries for 2017 to identify the conditions of the region in which they live, with a specific comparison of rural versus urban areas. This provides demographic trends, noting that Africa is the only region where the population share of youth is increasing. The core analysis uses national household surveys for 12 countries (including Tanzania) to compare households with and without young members, including whether the head is young as one of the controls, according to expenditure per capita (2011 USD) and poverty status. Households with young members tend to be poorer, in part because returns to education are lower, especially in rural areas (where education is measured as the share of members aged 18-63 who completed secondary education).

Most evidence of positive returns to education in developing countries is based on data on earnings of individuals (Psacharopoulos & Patrinos, 2018). There is considerable heterogeneity: returns tend be higher for those on lower incomes in Ethiopia (Girma & Kedir, 2005); public sector returns are higher than the private sector for Rwanda (Lassibille & Tan, 2005); in Tanzania, wage employees have higher returns than the self-employed and agricultural workers (Al-Samarrai & Reilly, 2008). For Tanzania, Serneels, Beegle & Dillon (2017), controlling for endogeneity and selection, found that estimated returns ranged between zero and 20% for men and 30-50% for women for a year of post-primary school depending on what type of survey data was used. Donath *et al.* (2021a, pp 18-23) note that returns per year of schooling vary according to the pay period: workers paid monthly, likely to be in wage employment, have positive returns after two years of education and these increase steadily up to about 150% after 16 years of school, whereas workers paid daily or weekly require up to six years of school before returns are positive and although returns then increase steadily they remain at least ten percentage points lower than for workers paid monthly. However, available surveys on earnings are not suitable for our purpose of assessing the impact of UPE due to small samples and/or limited data for the early 2000s.

Some studies consider the household level and find that a measure of household education (often for the head or a selected member) is positively associated with household welfare although the effects are heterogeneous; see Himaz and Aturupana (2018) for Sri Lanka, Alem and Söderbom (2012) for Ethiopia, Arouri *et al.* (2015) for Vietnam. Using the proportion of members with at least secondary education in cross-country analysis, Arsalan *et al.* (2019) 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 with a 7% and 6% decrease in poverty, respectively.

In exploring these factors, recent studies on Tanzania have focussed on classifying households in terms of gender, sector of employment and rural/urban residence; education is often included but is not a specific focus. Belghith *et al.* (2020, p7) note the increase in primary school enrolment from 71% to 85% and secondary enrolment from 3%1to 34% between 2012 and 2018, and observe that while education is positively associated with earnings and household welfare the expansion in access and attainment means that 'that the rewards for years of schooling below a certain level have declined [so that] the gains in income, and consumption, associated with primary education have become minimal' (Belghith *et al.* 2020, p8). Khan and Morrissey (2020) use data from the first three waves (2008/09, 2010/11 and 2012/13) of the Tanzania National Panel Survey (TNPS) to examine changes in household welfare (food consumption expenditure per adult equivalent) with a focus on the effect of income

diversification (the number and different types of work household members engage in). The more educated were more likely to be in wage and self-employment, associated with higher welfare, whereas the less educated were in agriculture and agricultural wage employment associated with poorer households.

To investigate how household factors including education influence the type of work, Khan and Morrissey (2020) estimate regressions for determinants of entry into and continued engagement in self-employment and both types of wage employment. The head's years of education is negatively associated with entry to and continuing in agricultural wage but positively associated with entry to and continuing in non-agricultural wage employment in rural areas (in urban areas it is positively associated with continuing non-agricultural wage). Thus, education does facilitate access to higher earning work: an extra year of education for the household head is associated with about 1.2% higher level of consumption. This effect is weak and not robust (insignificant) when lagged income diversification is used. However, education affects diversification, especially in rural areas, through association with non-agricultural wage employment.

Delesalle (2021) used the same waves of TNPS in combination with the 2002 Tanzania Population and Housing Census (TPHC) to, as noted above, estimate the effect of the UPE programme of 1974-78 (using variation in age at the time of reforms and in the scale of expansion over time and residence district) on consumption in 2002. As the TPHC does not include household consumption, variables (mostly dwelling characteristics) available in both the TPHC and TNPS are employed – the estimated effect of the variables on consumption in TNPS is matched with the census to generate predicted consumption in 2002. She estimates 'that the returns to education for the entire Tanzanian population are about 6 per cent' (p2), the UPE had a significant effect increasing years of education (p7), and the consumption returns to head's education are between 2.6 and 7.3 per cent overall (p9); returns are highest for self and wage employment but positive for agriculture (p10). These are much larger estimates than in Khan and Morrissey (2020) but this is unsurprising given different time periods, estimation strategies, samples, and dependent variables.

*Decomposition Analysis*

To estimate the contribution of education to welfare over time and between groups the decomposition methodology based on the seminal work of Oaxaca (1973) and Blinder (1973) as extended by Firpo, Fortin & Lemieux (2009, 2018) and Fortin, Lemieux & Firpo (2011) is employed, specifically the recentered influence function (RIF). Belghith *et al.* (2020) employed Oaxaca-Blinder decomposition to examine how much 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 using data from HBS 2012 and 2018. The 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 decomposition for four Arab countries and found that differences in educational attainment was one of the main determinants of the welfare gaps between male and female-headed and rural and urban households. Agyire-Tettey *et al.* (2018) for Ghana found that differences in educational attainment significantly explained the welfare gaps between rural and urban households. Skoufias & Katayama (2011), for Brazil, found that differences in the household head's education explained about 40 per cent of the welfare difference between metropolitan and urban areas.

## 3. Empirical Strategy

As noted above, available data do not permit estimating returns to education for individuals or applying the identification strategy of Delesalle (2021). As the structure of the data permit a comparison of youth and adult headed households in two periods, a decomposition method based on the characteristics of each household type is appropriate as it allows comparison of types of households 'as if' they had features (characteristics and returns on those characteristics) of another type. The empirical methodology follows Firpo *et al.* (2009, 2018) RIF based decomposition for the mean difference between two groups. 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 or composition effect) and the difference attributed to the relationship between *y* and *x* (coefficient effect or return effect). It goes further to decompose the contribution of each explanatory variable on the two parts.

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

$$lnC_{it} = \boldsymbol{\alpha s_{it}} + \boldsymbol{\beta x_{it}} + \varepsilon_{it} \quad (1)$$

Where $\boldsymbol{C}$ is the household consumption to poverty line ratio (CPL), our preferred welfare measure to account for the price differences (inflation) between surveys given the absence of good price deflators; $\boldsymbol{s}$ a vector of schooling of the household head and its square (in years); $\boldsymbol{x}$ is a vector (including a constant) of individual/household characteristics; $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are regression parameters; $\varepsilon$ is standard error term; and $i$ and $t$ index individual and time, respectively. With the exogeneity assumption, (1) is usually estimated using OLS. For any two groups, RIF decomposition uses the reweighted parameter estimates from (1) to decompose the statistic of interest into two parts as explained below. Following Rios-Avila (2020), the RIF decomposition (1) can be written as

$$Y = \boldsymbol{x'\beta} + \varepsilon \quad (2)$$

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

$$R = \begin{cases} 1 \; if \; youth \\ 0 \; if \; adult \end{cases} \text{; and}$$

$$T = \begin{cases} 1 \; if \;\; 2018 \\ 0 \; if \;\; 2001 \end{cases}$$

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:

$$f^k_{Y,X}(y,x) = f^k_{Y|X}(Y|X) f^k_X(X) \quad (3)$$

and its cumulative distribution function conditional on $T$ as:

$$F^k_Y(y) = \int f^k_{Y|X}(Y|X) dF^k_X(X) \quad (4)$$

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,

$$\Delta v = v_1 - v_0 = v(f_Y^1) - v(f_Y^0) \quad (5)$$

Which implies

$$\Delta v = v\left(f_{Y|X}^1(Y|X)dF(X)\right) - v\left(f_{Y|X}^0(Y|X)dF(X)\right)$$

We can rewrite (5) 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

$$v_c = v(f_Y^c) = v\left(f_{Y|X}^0(Y|X)dF_X^1(X)\right) \quad (6)$$

$\Delta v_S = v_1 - v_c$ 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 $X$s.

From $v(F_Y) = \boldsymbol{x}'\boldsymbol{\beta}$ it follows that

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

$$v_0 = E\left(RIF\left(y_i; v(f_Y^0)\right)\right) = \bar{X}^{0\prime}\hat{\beta}^0; \text{ and}$$

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

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

$$F_Y^c = \int f_{Y|X}^0(Y|X)dF_X^1(X) \cong \int f_{Y|X}^0(Y|X)dF_X^0(X)\,\omega(X) \quad (7)$$

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

$$\omega(X) = \frac{1-p}{p}\frac{P(T=1|X)}{1-P(T=1|X)} \quad (8)$$

with $p$ is the sample share in group 1 and $P(T=1|X)$ is 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 (8) yields

$$v_c = E\left(RIF\left(y_i; v(f_Y^c)\right)\right) = \bar{X}^{c\prime}\hat{\beta}^c \quad (9)$$

The decomposition can then be rewritten as

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

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

Then

$$\Delta v = \Delta v_S^p + \Delta v_S^e + \Delta v_X^p + \Delta v_X^e \quad (10)$$

The component $\Delta v_S^p + \Delta v_S^e$ is called the coefficient effect which comprises the pure coefficient effect ($\Delta v_S^p$) and the reweighting error ($\Delta v_S^e$). The component $\Delta v_X^p + \Delta v_X^e$ is called the aggregate composition effect and constitutes the pure composition effect ($\Delta v_X^p$) and specification error ($\Delta v_X^e$). The error components help assess the quality of the reweighting and specification of the regression function; smaller and insignificant coefficients of the error components indicate more robust results (Firpo *et al.*, 2018; Rios-Avila, 2020). 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, 2020).

## 4. Data and Descriptive Statistics

The Tanzanian Household Budget Surveys (HBS) for 2001 and 2018, conducted by the National Bureau of Statistics, are among the most extensive household surveys in Tanzania, collecting rich individual and household information, including consumption, covering all regions of the Mainland.[5] 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.

Households are categorised according to the head's age group: youth and adults. Youth households include all households headed by a youth aged between 15 and 25, who may have benefitted from the 2001 UPE by 2018, and adult households include all households headed by an adult aged over 35 (who won't have benefitted from UPE).[6] The welfare indicator is measured at the household level as the ratio of the per adult equivalent household consumption to the national poverty line (CPL), both as reported in the respective HBS (URT, 2002; URT, 2019), and is assigned to the head of household. Given the absence of good price deflators covering 2001 to 2018, especially to capture spatial price variation, the ratio is an appropriate if simple way to represent the relative welfare of the household at the time of the survey and helps to account for inflation and trends in earnings between the surveys. This is useful for comparing welfare in 2001 to 2018.

Education is the household head's level of education measured in years (in the context of estimating returns to education, Donath *et al.*, 2021a show that this gives results consistent with using level of education completed). This is appropriate as our comparison is between groups of households distinguished by the age of the head – youth who benefited from UPE by 2018 and adults who didn't. A limitation of distinguishing households based on the age of the head

5 Tanzania (also the United Republic of Tanzania) includes the Tanzania Mainland (Tanganyika) and the island of Zanzibar. The Mainland covers about 99% of the total area and about 98% of the total population.

6 As HBS is a general household survey there is very little information on how households were formed, such as due to marriage or migration, so we are unable to investigate determinants of youth becoming a head of household.

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. 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.

Section 4.1 provides definitions of the main variables used in the analysis, followed by Section 4.2 with descriptive statistics comparing youth (aged 15-25) and adult (aged over 35) headed households in 2001 and 2018. Further descriptive data can be found in Appendix A, including for the broader youth (aged 15-35). The broad patterns are very similar, notably the increase in educational attainment by youth (however defined) in 2018 and the increase in welfare of youth headed households.

## 4.1. Definition of the Main Variables

All variables are taken from the HBS and associated reports (for the poverty line); for survey measures refer to URT (2002, 2019). Here we list the basic definitions of measures for each variable.

Household characteristics

- **CPL:** the ratio of household consumption per adult equivalent to the poverty line in logarithm form (to adjust for skewness of income)
- **poor:** = 1 for households below the basic needs poverty line (0 otherwise)
- **rural:** = 1 for households in rural area (0 otherwise)
- **hhsize:** total number of usual members in the household

Household head characteristics

- **education:** years of schooling of the household head
- **noeducation:** = 1 if household head completed less than three years of primary education (0 otherwise)
- **someprimary:** = 1 if head completed 4-6 years primary (0 otherwise)
- **primary:** = 1 if head completed the seven years of primary (0 otherwise)
- **somesecondary:** = 1 if head completed 2-3 years of secondary education (0 otherwise)
- **secondary:** = 1 if head completed the four years of lower secondary education (0 otherwise)
- **postsecondary:** = 1 if head has more than lower secondary education (0 otherwise)
- **age:** age of the household head in years
- **female:** = 1 if the head is female (0 otherwise)
- **married:** = 1 if the head is married (0 otherwise)
- **agric:** = 1 if the main economic activity of the head is agriculture/fishery (0 otherwise)
- **wage:** = 1 if the main economic activity of the head is wage employment (0 otherwise)
- **self:** = 1 if the main economic activity of the head is self-employment (out of agriculture) (0 otherwise)
- **Unemployed/inactive:** = 1 if the head is unemployed/inactive (0 otherwise)

## 4.2. Descriptive Statistics

Tables 1 and 2 show the means for the continuous variables and the percentages of the respective group's observations for the binary variables included in the analysis. Table 1 shows the characteristics of youth headed households (Youth) compared to adult headed households (Adult) for each year and compares the characteristics of young (aged 15-25) household heads (Head) to all other youth in the sample (Other). The share of households headed by a youth in the total sample decreased by a third from 7.5% in 2001 to 5% in 2018 (there were 473 youth headed households in 2018). This could be due to multiple factors (including changes in regional sampling as a larger share of the smaller sample is urban in 2018). In both years youth headed households had significantly higher welfare (CPL), were less likely to be poor, and were much smaller than adult headed households. The CPL increased and the proportion poor and size

decreased for both groups in 2018, reflecting the general reduction in poverty compared to 2001, although the differences remained.

**Table 1: Characteristics of Youth (15-25) Headed Households by Survey Year**

| *Characteristics* | | 2001 | | | 2018 | |
|---|---|---|---|---|---|---|
| *% unless stated* | Youth | Adult | Difference | Youth | Adult | Difference |
| *Households (by Head)* | | | | | | |
| CPL (ratio) | 1.86 | 1.44 | 0.42*** | 2.27 | 1.90 | 0.37*** |
| Poor | 0.25 | 0.38 | -0.13*** | 0.14 | 0.26 | -0.12*** |
| Rural | 0.79 | 0.77 | 0.02*** | 0.73 | 0.64 | 0.09*** |
| Size (number) | 4.26 | 6.92 | -2.66*** | 3.69 | 6.64 | -2.95*** |
| *Youth* | Head | Other | Difference | Head | Other | Difference |
| Education (years) | 5.98 | 5.45 | 0.53*** | 6.85 | 7.15 | -0.27*** |
| No education | 0.16 | 0.22 | -0.06*** | 0.15 | 0.13 | 0.02*** |
| Some primary | 0.05 | 0.19 | -0.14*** | 0.11 | 0.10 | 0.01*** |
| Primary | 0.76 | 0.54 | 0.22*** | 0.49 | 0.46 | 0.03*** |
| Some secondary | 0.01 | 0.02 | -0.01*** | 0.07 | 0.12 | -0.05*** |
| Secondary | 0.02 | 0.03 | -0.01*** | 0.19 | 0.17 | 0.02*** |
| Post-secondary | 0.00 | 0.01 | -0.01*** | 0.00 | 0.02 | -0.02*** |
| Age (years) | 23.21 | 19.46 | 3.75*** | 23.03 | 19.19 | 3.84*** |
| Female | 0.25 | 0.59 | -0.34*** | 0.19 | 0.55 | -0.36*** |
| Married | 0.72 | 0.30 | 0.42*** | 0.75 | 0.24 | 0.51*** |
| *Observations* | 1,647 | 22,481 | - | 473 | 8,121 | - |

*Notes:* Author's calculations from HBS 2001 and 2018 data weighted using survey weights; mean value for continuous variables and % share of sample for binary indicators. CPL is household consumption relative to the poverty line. 'Difference' for *households* is the value for youth headed households minus the corresponding value for adult headed households; 'Difference' for *youth* characteristics are based on the average for youth heads minus the mean for other youth (with significance $^{*}$ $p < 0.10$, $^{**}$ $p < 0.05$, $^{***}$ $p < 0.01$ based on Lincom's test of mean differences).

Compared to youths that were not heads, a much higher proportion of youth heads were married, although more than a quarter were not, and they were almost four years older on average than other youth. A quarter of youth heads were female in 2001, declining to about a fifth in 2018, and a significant share of other youth were married (many would be the spouses of youth heads). Although youth heads had slightly more years of education in 2001, this was reversed by 2018: education of youth heads increased by about one year on average compared to about two years on average for other youth. In 2001 youth heads were more likely to have completed primary but also more likely to have no education; by 2018 the proportion of youth heads with any level of post primary education was significantly lower than for other youth. This is consistent with a tendency for more educated individuals to delay forming households. However, many factors interact to motivate household formation and education is only one of these (see Section 5.2).

**Table 2: Characteristics for Adult Headed Households, 2001 and 2018**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | 2001 | 2018 | 2018 | Difference |
| *Characteristics* | Age >35 | 35<Age≤53 | Age>53 | (3)-(1) |
| *Household* | | | | |
| 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*** |
| Size | 6.89 | 6.56 | 6.64 | -0.25** |
| *Head* | | | | |
| Education (years) | 4.24 | 6.28 | 4.51 | 0.27** |
| No education | 0.42 | 0.18 | 0.40 | -0.02* |
| Some primary | 0.18 | 0.06 | 0.15 | -0.03** |
| Primary | 0.33 | 0.64 | 0.35 | 0.02 |
| Some secondary | 0.00 | 0.01 | 0.01 | 0.01* |
| Secondary | 0.05 | 0.08 | 0.07 | 0.02** |
| Post-secondary | 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 | |

*Notes:* As for Table 1 except difference is the value in column (3), 2018 adults that were adult (aged >35) also in 2001, minus the corresponding value in (1). Column 2 includes adults in 2018 who were youth (broad definition) in 2001.

**Figure 1: Household Consumption (CPL) by Head Age Group and Year**

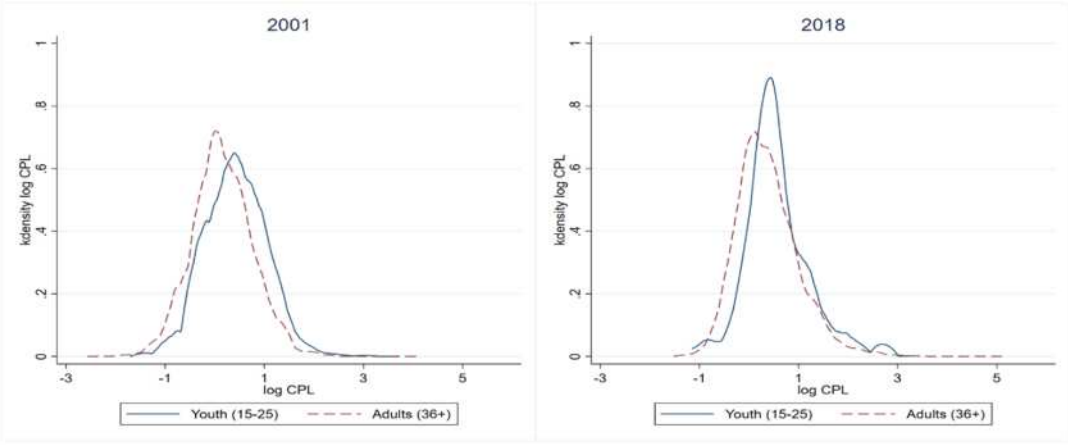
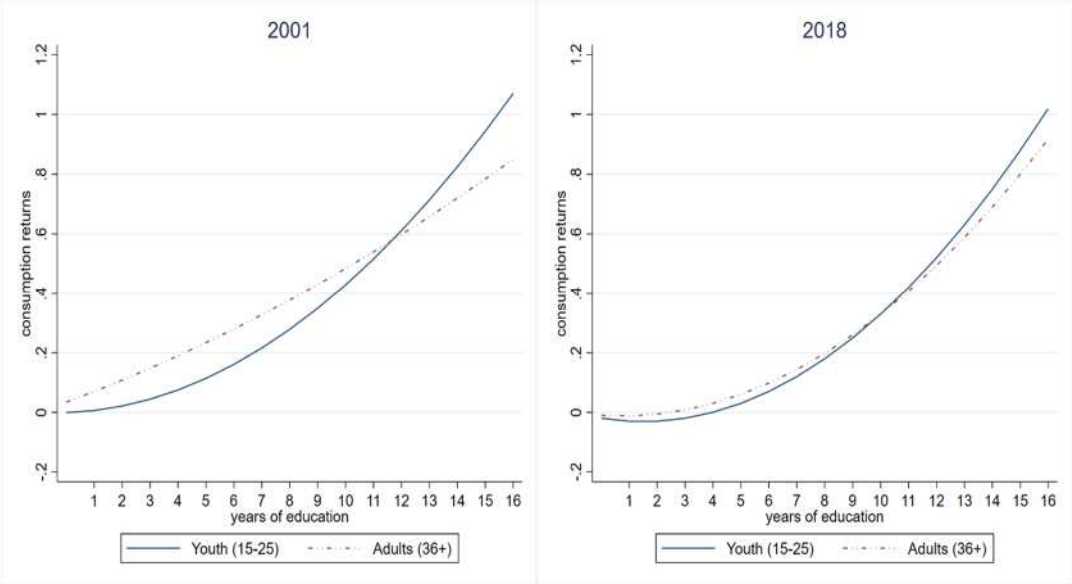



Table 2 provides the data for adult headed households, separated into two groups in 2018 – those in the broad youth category in 2001 (aged 35 to 53 years in 2018) and the adult category in 2001 (aged 54 years and above in 2018). It is clear that household welfare rose by 2018;

household size, the proportion rural and in poverty all fell, and there were few differences in household characteristics of the two 2018 groups. Educational attainment rose, especially for the younger adult group. Adults aged 35 to 53 years in 2018 have significantly more education than those aged over 53 (who are quite similar to adults in 2001); the proportion with no education is much lower and the proportion with completed primary is much higher (other differences are small).

It is clear comparing Tables 1 and 2 that youth headed households have higher CPL and lower poverty rates than adult headed households in both years. This is shown in the density plots for household CPL in Figure 1. In both years, the distribution for the youth (15-25) is to the right of that for adults; the proportion of youth households above the poverty line (logCPL>0) is greater, especially in 2018.[7] The peaks around logCPL=0 (CPL=1) are consistent with a large number of households clustered close to the poverty line (Belghith *et al.*, 2020). As such households are likely to be vulnerable it should be acknowledged that the estimates of improvements in welfare should be considered as modest (and potentially fragile).

**Figure 2: Implied Returns to Education by Age Group and Year**




## 5. Results and Discussion

Section 5.1 reports whether returns to education for youth (15-25), in terms of household welfare, are different from those of adults; and whether they have changed between 2001 and 2018. Comparable analysis for broader youth (15-35) is in Appendix B. To partially address potential endogeneity, Section 5.2 shows that although education influences the likelihood for a youth being a head of household, this is unlikely to alter our main results.

7 The pattern is similar for broader youth (15-35) except that the shape in 2018 is closer to that in 2001 (Appendix Figure A1; Donath *et al.*, 2021b).

**Table 3: OLS Estimates of Returns to Education by Age and Survey Year**

| | 2001 | | 2018 | |
|---|---|---|---|---|
| | Youth (15-25) | Adult | Youth (15-25) | Adult |
| *Sch* | -0.001 | 0.034*** | -0.020 | -0.010*** |
| | (0.013) | (0.003) | (0.021) | (0.004) |
| $Sch^2$ | 0.004*** | 0.001*** | 0.005*** | 0.004*** |
| | (0.001) | (0.000) | (0.002) | (0.000) |
| *Age* | 0.072 | -0.011*** | 0.394* | 0.002 |
| | (0.099) | (0.003) | (0.210) | (0.003) |
| $Age^2$ | -0.125 | 0.008*** | -0.870* | -0.002 |
| | (0.227) | (0.002) | (0.475) | (0.003) |
| *Female* | 0.173*** | 0.010 | 0.144** | 0.054*** |
| | (0.031) | (0.014) | (0.061) | (0.018) |
| *Rural* | -0.145*** | -0.147*** | -0.215*** | -0.183*** |
| | (0.032) | (0.011) | (0.064) | (0.016) |
| *Married* | 0.076** | 0.071*** | -0.009 | 0.052*** |
| | (0.034) | (0.014) | (0.063) | (0.019) |
| *lnSize* | -0.550*** | -0.451*** | -0.493*** | -0.473*** |
| | (0.026) | (0.008) | (0.054) | (0.011) |
| _cons | -0.088 | 1.153*** | -3.577 | 0.656*** |
| | (1.068) | (0.082) | (2.304) | (0.099) |
| Controls | Yes | Yes | Yes | Yes |
| AME(*Sch*) | 0.048*** | 0.043*** | 0.047*** | 0.027*** |
| | (0.006) | (0.001) | (0.009) | (0.002) |
| *N* | 1,647 | 13,983 | 473 | 6,945 |
| $R^2$ | 0.40 | 0.40 | 0.488 | 0.44 |

*Notes:* AME(*Sch*) is the average mean effect of a year of schooling; Household size is in logs (*lnSize*). Other controls included (not reported) are livestock per capita, region of residence and ownership of assets. Standard errors in parentheses (significance indicated by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

### 5.1 Education and Youth Welfare

Table 3 presents the results for CPL for each year based on OLS regression estimates of equation (1). Years of schooling (*Sch*) of the household head is positively and significantly correlated with welfare for adult headed households only, but $Sch^2$ is positive and significant for youth and adults, in both years. The significance of schooling squared implies a strong convex relationship between education and welfare – each extra year of schooling is associated with higher welfare than the previous year – and this is stronger for adults but slightly weaker for youth in 2018. The presence of $Sch^2$ complicates the interpretation of the coefficients of schooling variables so we plot the implied welfare returns to education from Table 3 in Figure

2. In 2001, youth headed households had lower returns up to 12 years education (completing lower secondary) and beyond this adult headed households had lower returns. By 2018 returns are only positive for youth (adults) beyond about four (three) years of school and decline throughout, especially for adults, although the extent of decline is less beyond 14 years of education.[8] Most other included regressors have the expected sign; coefficients on *Age* and $Age^2$ are only clearly significant for adults in 2001(implying a threshold age before earnings rise) and *Married* is insignificant for youth in 2018. In contrast, coefficients are positive and significant for *Female* (except adults in 2001), but negative and significant for rural and household size (both associated with lower welfare, more pronounced in 2018).

*Education and Welfare Changes*

The factors contributing to differences in mean household welfare between youth and adult headed households for each year are investigated using RIF regressions to decompose into two parts as explained in Section 3: the part due to differences in characteristics or endowment (the explained part) and that due to differences in returns to these characteristics (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 reweighting error.

Table 4 presents the decomposition results by year. To simplify interpretation, the coefficients of the education variables (*sch* and $sch^2$) are aggregated into one variable 'education'; the coefficient of *age* and $age^2$ into 'headage'; and ownership of assets, livestock per capita, and dummies for regions of residence into 'other controls'.[9] The top panel of Table 4 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 (for the reweighted model) implying that it is only the difference in endowment that explains differences in welfare between the two age groups. Both the specification and reweighting errors are insignificant, suggesting that the model is correctly specified and reweighed (Firpo *et al.*, 2018; Rios-Avila, 2020). Although the reweighting error is insignificant, it is large relative to the raw difference and the 2001 counterfactual is not significantly different from zero. To assess if the reweighting has an undue effect on the estimates, unweighted results are also reported (in this case there is no counterfactual and the explained part is insignificant). The coefficients for the explained and unexplained parts are much larger, consistent with the unweighted model not accounting for the effects of specification and reweight errors, and although the size and in some cases significance of variables alters, the education variable remains significant.

8 The pattern is different for broader youth (15-35): in 2001 returns to adults fell below youth after 8 years of education, whereas in 2018 returns fell for both groups, notably for broad youth, and were slightly higher for adults (Appendix Figure B1).

9 The Stata command *oaxaca_rif* is calibrated for this common approach in decomposition (and the aggregations fit with the specification).

**Table 4: Reweighted and Unweighted RIF Decomposition Within Survey Years**

| | 2001 | | 2018 | |
|---|---|---|---|---|
| | Weighted | Unweighted | Weighted | Unweighted |
| *Overall* | | | | |
| Youth (15-25) | 0.420*** | 0.420*** | 0.571*** | 0.571*** |
| Counterfactual | 0.067 | . | 0.411*** | . |
| Adult | 0.149*** | 0.149*** | 0.374*** | 0.374*** |
| Difference | 0.271*** | 0.271*** | 0.197*** | 0.197*** |
| Explained | 0.352*** | 4.386 | 0.160*** | 0.980 |
| Unexplained | -0.082 | -4.115 | 0.037 | -0.783 |
| *Explained* | 0.292*** | 4.386 | 0.248** | 0.980 |
| education | 0.101*** | 0.059** | 0.085*** | 0.032* |
| headage | 0.018** | 4.065 | 0.006 | 0.837 |
| female | 0.000 | 0.000 | -0.006 | -0.004 |
| rural | -0.002 | 0.001 | 0.007 | -0.003 |
| married | -0.000 | -0.001 | 0.001 | 0.001 |
| lnSize | 0.200*** | 0.281*** | 0.292*** | 0.230*** |
| Other controls | -0.025 | -0.022 | -0.138*** | -0.105*** |
| *Unexplained* | -2.298 | -4.115 | -8.037 | -0.783 |
| education | 0.037 | -0.007 | 0.016 | 0.015 |
| age | -1.681 | -1.573 | -4.953 | -0.865 |
| female | 0.008 | 0.002 | 0.017 | 0.017* |
| rural | -0.013 | -0.018 | 0.014 | 0.006 |
| married | -0.015 | 0.016 | 0.015 | -0.021 |
| lnSize | -0.568 | -0.247** | 0.344 | 0.063 |
| Other controls | 0.347 | 0.072 | 0.068 | 0.056 |
| constant | -0.388 | -2.353 | -3.499 | -0.005 |
| *Specification error* | 0.061 | | -0.088 | |
| *Reweight error* | 2.216 | | 8.075 | |
| N1 (youth) | 1,647 | 1,647 | 473 | 473 |
| N2 (adult) | 13,983 | 13,983 | 6,945 | 6,945 |
| N (all observations) | 15,630 | 15,630 | 7,418 | 7,418 |

*Notes:* 'Other controls' is the aggregate effect of livestock per capita, region of residence and ownership of assets. Binary variables are normalised; significance * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The breakdown of the 'pure explained' component reveals that the coefficient on education is positive and significant (although smaller and weakly significant in the unweighted model), confirming that the youths heading households have significantly better education attainment than adults heading households. About a third of the pure explained welfare difference is attributable to differences in educational attainment between youths and adults (the remainder is largely attributable to differences in household size). 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 35% (0.101/0.292) higher in 2001 and 34% higher in 2018.[10] Household size is the only other characteristic that has a consistently significant contribution to the explained difference. These results are consistent with Table 3. The difference in returns to education (unexplained),

10 The benefit is higher (40%) for broad youth in 2001 but slightly smaller (32%) in 2018 (Table B2).

however, does not have a significant effect on welfare (also consistent with Table 3 and Figure 2). No variables are significant for the unexplained difference.

**Table 5a: RIF Oaxaca-Blinder Decomposition for Youth (15-25), Pooled and Males**

| | Pooled | | Male | |
|---|---|---|---|---|
| | Weighted | Unweighted | Weighted | Unweighted |
| *Overall* | | | | |
| 2018 | 0.571*** | 0.571*** | 0.535*** | 0.535*** |
| Counterfactual | 0.263*** | . | 0.228*** | . |
| 2001 | 0.420*** | 0.420*** | 0.438*** | 0.438*** |
| Difference | 0.152* | 0.152* | 0.097 | 0.097 |
| Explained | 0.309*** | 0.310*** | 0.307*** | 0.304*** |
| Unexplained | -0.157 | -0.158** | -0.210* | -0.207** |
| *Explained* | 0.396*** | 0.310*** | 0.391*** | 0.304*** |
| education | 0.097*** | 0.070*** | 0.091** | 0.067*** |
| age | -0.001 | 0.002 | 0.001 | 0.003 |
| female | 0.004 | -0.004 | | |
| rural | 0.013* | 0.005 | 0.018* | 0.011* |
| married | -0.003 | -0.001 | -0.007 | 0.001 |
| lnhhsize | 0.043* | 0.052* | 0.050** | 0.028 |
| Other controls | 0.229*** | 0.186*** | 0.227*** | 0.182** |
| *Unexplained* | 0.073 | -0.158** | -0.062 | -0.207** |
| education | 0.134 | -0.147 | 0.123 | -0.189 |
| age | -4.083 | -2.553 | -0.353 | -2.257 |
| female | 0.031 | 0.015 | | |
| rural | -0.132 | 0.005 | -0.138 | -0.029 |
| married | -0.074 | -0.024 | -0.122 | -0.061 |
| lnhhsize | 0.272 | 0.194* | 0.268 | 0.199* |
| Other controls | 0.033 | -0.005 | -0.076 | -0.006 |
| constant | 3.924 | 2.394 | 0.184 | 2.118 |
| *Specification error* | -0.087 | | -0.084 | |
| *Reweight error* | -0.230 | | -0.148 | |
| N1 (2018) | 473 | 473 | 377 | 377 |
| N2 (2001) | 1,647 | 1,647 | 1,181 | 1,181 |
| N (all observations) | 2,120 | 2,120 | 1,558 | 1,558 |

*Notes:* No separate estimates for female heads due to too few observations. 'Other controls' is the aggregate effect of livestock per capita, region of residence and ownership of assets. Binary variables are normalised; significance * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5 provides the RIF decomposition for the youth heading households between 2001 and 2018. To assess heterogeneity, the decomposition is performed by gender (Table 5a) and rural/urban residence (Table 5b). Again, both the specification and reweighting errors are insignificant, and in this case the reweighting error is not so large relative to the raw difference while the counterfactuals are significantly different from zero. Nevertheless, unweighted results are also reported and the coefficients for the explained and unexplained parts are similar in magnitude and significance to those for the reweighted model. We can have more confidence

in the reweighted model for decomposing effects over time for youth headed households than for the comparison with adult headed households.

**Table 5b: Decomposition for Youth (15-25), Rural versus Urban**

| | Rural | | Urban | |
|---|---|---|---|---|
| | Weighted | Unweighted | Weighted | Unweighted |
| *Overall* | | | | |
| 2018 | 0.415*** | 0.415*** | 0.995*** | 0.995*** |
| Counterfactual | 0.211** | . | 0.728*** | . |
| 2001 | 0.340*** | 0.340*** | 0.722*** | 0.722*** |
| Difference | 0.075 | 0.075 | 0.272 | 0.272 |
| Explained | 0.204** | 0.203*** | 0.267* | 0.534*** |
| Unexplained | -0.128 | -0.127 | 0.005 | -0.262 |
| *Explained* | 0.273*** | 0.203*** | 0.416** | 0.534*** |
| education | 0.061** | 0.025** | 0.155** | 0.207*** |
| age | -0.001 | -0.003 | 0.008 | 0.019 |
| female | 0.008 | -0.002 | -0.013 | 0.020 |
| rural | | | | |
| married | -0.000 | -0.000 | 0.008 | -0.005 |
| lnhhsize | 0.027 | 0.029 | -0.023 | 0.084** |
| Other controls | 0.169*** | 0.155** | 0.286* | 0.194 |
| *Unexplained* | 0.095 | -0.127 | -0.504 | -0.262 |
| education | 0.155 | -0.075 | -0.162 | 0.000 |
| age | -5.960 | -3.552 | -5.172 | 4.923 |
| female | 0.032 | 0.022* | -0.062 | -0.057* |
| rural | | | | |
| married | -0.139 | 0.011 | 0.092 | -0.049 |
| lnhhsize | 0.497 | 0.270** | -0.831 | 0.112 |
| Other controls | -0.133 | -0.066 | 0.293 | 0.046 |
| constant | 5.727 | 3.350 | 5.336 | -5.373 |
| *Specification error* | -0.069 | | -0.149 | |
| *Reweight error* | -0.224 | | 0.509 | |
| N1 (2018) | 333 | 333 | 140 | 140 |
| N2 (2001) | 608 | 608 | 1,039 | 1,039 |
| N (all observations) | 941 | 941 | 1,179 | 1,179 |

*Notes:* As for Table 5a.

The results for the pooled model in Table 5a suggest that the difference in welfare between the two periods is attributed to differences in characteristics (the explained part) – other controls (such as assets) are the main factor but education accounts for about 25% of the explained difference while household size is slightly more important for male heads – and coefficients are insignificant. Although the raw difference is at best weakly significant, the counterfactual is relatively large. The share of education in the explained part implies that if the youth in 2001 had the same education endowment as their 2018 counterparts, their welfare would have been about 25% higher (slightly lower for males).

There is evidence of heterogeneity by place of residence (Table 5b): education (and other controls to a lesser extent) is less important for rural than urban households; whereas education accounts for about 37% of the explained difference in urban areas it only accounts for 22% of the difference in rural areas. This is consistent with education endowments being higher in urban areas and the greater association between education and welfare for urban households is consistent with the greater availability of wage employment. In rural areas, in contrast, education endowments are lower and there is less wage employment – a greater share of employment is in lower skilled agriculture. No variables are significant for the unexplained part (in the reweighted model), implying that the decline in returns to education had no detrimental effect on welfare.

### 5.2 Robustness Checks

Endogeneity of education due to unobserved ability and of education for youth selection into heading a household are concerns as factors associated with higher welfare, such as education, are also associated with a higher likelihood of youth to form a household. Whereas methods to address these issues are well documented in the literature, how to combine these methods in a RIF decomposition has not been established (Firpo *et al.*, 2018). 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 influences the likelihood for a youth aged 15-25 being a head of household. Although by no means a panacea, LPM has the advantage of allowing the inclusion of household fixed effects.

**Table 6: LPM Regression Results by Gender and Place of Residence 2001 (Age 15-25)**

| | (1) All | (2) Female | (3) Male | (4) Rural | (5) Urban |
|---|---|---|---|---|---|
| *Sch* | 0.003*** | 0.001 | 0.001 | 0.004** | 0.002* |
| | (0.001) | (0.001) | (0.002) | (0.002) | (0.001) |
| *female* | -0.114*** | | | -0.185*** | -0.074*** |
| | (0.005) | | | (0.010) | (0.005) |
| *married* | 0.088*** | 0.013 | 0.194*** | 0.101*** | 0.088*** |
| | (0.008) | (0.008) | (0.032) | (0.012) | (0.010) |
| *AGR* | -0.004 | 0.003 | 0.011 | 0.005 | -0.019*** |
| | (0.004) | (0.006) | (0.008) | (0.008) | (0.005) |
| *WAGE* | 0.098*** | 0.053*** | 0.034 | 0.092*** | 0.097*** |
| | (0.011) | (0.018) | (0.022) | (0.031) | (0.012) |
| *SELF* | 0.107*** | 0.046** | 0.061*** | 0.032 | 0.125*** |
| | (0.011) | (0.023) | (0.020) | (0.026) | (0.012) |
| _cons | 0.077*** | 0.019** | 0.065*** | 0.111*** | 0.063*** |
| | (0.006) | (0.007) | (0.011) | (0.011) | (0.008) |
| *N* | 24,128 | 13,369 | 10,759 | 8,171 | 15,957 |
| $R^2$ | 0.727 | 0.923 | 0.928 | 0.681 | 0.762 |

*Note:* Dummies for sectors of main employment are AGR (agriculture), *WAGE* (wage employment) and *SELF* (non-agricultural self-employment); Robust standard errors in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

The LPM estimates in Table 6 for 2001 show that, after controlling for household fixed effects, more educated youth were significantly more likely to be a household head – an extra

year of education is associated with about 0.003 increase on average in the probability of the youth being a head (no difference between males and females and higher for rural than urban). Females are less likely to be a head, especially in rural areas. Being married is the single most important factor for males (insignificant for females). Youth in self or wage employment, both associated with higher household welfare (Khan & Morrissey, 2020), are more likely to be heads. Table 7 shows that the strength of education increased in 2018, albeit only to 0.005 (lower in urban areas). Other results are similar except that agricultural employment is also significant although wage employment has the highest coefficient.

Despite the significance of the coefficients of education, the sizes are negligible. A coefficient on education of 0.005 (most coefficients are lower, or insignificant) implies 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 unlikely to significantly affect the estimated coefficients of education in our main results, especially as mean years are relatively low (six years in 2001 and seven in 2018). While this signals a concern, the low coefficient gives us cautious confidence in the main results. Appendix Tables B4 and B5 show that this also holds for broad youth.

**Table 7: LPM Regression Results by Gender and Place of Residence 2018 (Age 15-25)**

| | (1) All | (2) Female | (3) Male | (4) Rural | (5) Urban |
|---|---|---|---|---|---|
| *Sch* | 0.005*** | -0.000 | 0.002 | 0.005*** | 0.003* |
| | (0.001) | (0.001) | (0.002) | (0.002) | (0.002) |
| *female* | -0.137*** | | | -0.156*** | -0.089*** |
| | (0.010) | | | (0.012) | (0.015) |
| *married* | 0.060*** | 0.009 | 0.038 | 0.065*** | 0.054* |
| | (0.011) | (0.011) | (0.027) | (0.011) | (0.032) |
| *AGR* | 0.022** | -0.002 | 0.012 | 0.018* | 0.013 |
| | (0.010) | (0.009) | (0.012) | (0.010) | (0.030) |
| *WAGE* | 0.106*** | 0.028 | 0.026 | 0.115*** | 0.107*** |
| | (0.024) | (0.053) | (0.035) | (0.039) | (0.031) |
| *SELF* | 0.061*** | 0.028 | 0.038 | 0.017 | 0.111*** |
| | (0.019) | (0.025) | (0.024) | (0.026) | (0.028) |
| _cons | 0.055*** | 0.016* | 0.060*** | 0.064*** | 0.035* |
| | (0.012) | (0.008) | (0.016) | (0.014) | (0.020) |
| *N* | 8,594 | 4,425 | 4,169 | 6,095 | 2,499 |
| $R^2$ | 0.726 | 0.951 | 0.967 | 0.698 | 0.802 |

*Notes:* As for Table 6.

## 6 Conclusions

This paper focused on investigating how much of the welfare difference between households headed by youth (aged 15-25) in 2001 and 2018 and the difference between youth and adult headed households in each of the years can be attributed to changes 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 15–25 years in 2018. In contrast to previous studies examining the association between education and welfare at any given point in time, this study

examines 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. Samples of household heads from the 2001 and 2018 HBS were investigated using RIF decomposition of the mean.

The analysis decomposed the welfare differences between cohorts into the part attributable to differences in characteristics (education is the focus) and the part attributable to returns to these characteristics (again the focus is on returns to education). The decomposition of welfare between youth in 2001 (pre reform) and in 2018 (post reform) sheds light on the effect 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, there is evidence that UPE improved welfare through increasing educational attainment.

Proxying welfare by household (per adult equivalent) consumption expenditure relative to the national poverty line, the study used a reweighted RIF to decompose the welfare differences. The analysis shows that youths, having more education, enjoy higher welfare levels than adults in both years; if adults had the same education as youth their welfare would be about a third higher. The difference in the returns to education (which decline) is not significant. Comparing youth cohorts across years, the youth in 2018 have higher education and welfare levels than their 2001 counterparts; increased educational attainment between 2001 and 2018 can explain 24% of the increase in youth welfare. Despite the decline in returns to education, we find no evidence that this reduced welfare. Although we did not control for endogeneity of education or endogenous selection of youth to be heads of households in our welfare decomposition, given that absence of established methods in RIF decomposition, we show that education is a minor factor determining whether a young person is a head of household.

The analysis is based on comparisons of youth and adult headed households, given the difficulty in measuring education and consumption of individuals at the household level. As a result, we could not analyse the welfare effects for the youth who live in households headed by an adult. Nevertheless, we add to the evidence in Delesalle (2021) that the early UPE in the 1970s increased incomes in 2002 by showing the the extensive UPE and expansion of secondary education in the 2000s contributed to increased income (at the household level) in 2018. The benefit of increased levels of education (endowments) offset the decline in returns to schooling. The limited impact on returns is consistent with literature showing that UPE has little impact on educational attainment (Al-Samarrai, 2006; Kan & Klasen (2021). Addressing the low returns requires increasing the quality of education and training and expanding labour market opportunities.

## Data Statement

The cleaned and constructed data used in the analysis, and the do files, are available on request.

## REFERENCES

Agyire-Tettey, F., Ackah, C. G., & Asuman, D. (2018). An Unconditional Quantile Regression Based Decomposition of Spatial Welfare Inequalities in Ghana. *Journal of Development Studies,* **54**(3), 537–556. https://doi.org/10.1080/00220388.2017.1296571

Alem, Y., & Söderbom, M. (2012). Household-Level Consumption in Urban Ethiopia: The Effects of a Large Food Price Shock. *World Development,* **40**(1), 146–162. https://doi.org/10.1016/j.worlddev.2011.04.020

Al-Samarrai, S. (2006), Achieving education for all: How much does money matter? *Journal of International Development*, 18 (2), 179–206.

Al-Samarrai, S., & 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. https://doi.org/10.1080/00220380701789950

Arouri, M., Nguyen, C., & Youssef, A. Ben. (2015). Natural Disasters, Household Welfare, and Resilience: Evidence from Rural Vietnam. *World Development*, **70**, 59–77. https://doi.org/10.1016/j.worlddev.2014.12.017

Arsalan, A., Tschirley, D., & Egger, E.-M. (2021). Rural Youth Welfare along the Rural-urban Gradient: An Empirical Analysis across Developing Countries. *Journal of Development Studies*, 57 (4), 544-570. https://doi.org/10.1080/00220388.2020.1808197

Belghith, N. B. H., Karamba, W., Talbert, E., & de Boisseson, P. (2020). *Tanzania Mainland Poverty Assessment: Tanzania's Path to Poverty Reduction and Pro-Poor Growth*. The World Bank. http://documents.worldbank.org/curated/en/254411585030305188/pdf/Part-1-Path-to-Poverty-Reduction-and-Pro-Poor-Growth.pdf

Blinder, A. S. (1973). Wage Discrimination: Reduced Form and Structural Estimates. *The Journal of Human Resources,* **8**(4), 436–455.

Colclough, C., Al-Samarrai, S., Rose, P. & Tembon, M. (2003). *Achieving Schooling For All in Africa: Costs, Commitment and Gender*, Aldershot: Ashgate.

Davis, B., Di Guiseppe, S. & Zezza, A. (2017). Are African Households (Not) Leaving Agriculture? Patterns of Households' Income Sources in Rural Sub-Saharan Africa. *Food Policy*, 67: 153–74

De Janvry, A., & Sadoulet, E. (2016). *Development Economics: Theory and Practice*. Taylor and Francis, London and New York. https://doi.org/10.4337/9781785365065.00013

Deaton, A. (2018). *The Analysis of Household Surveys* (Reissue), The World Bank, Washington D.C. https://doi.org/10.1596/0-8018-5254-4

Delesalle, E. (2021). The effect of the Universal Primary Education program on consumption and on the employment sector: Evidence from Tanzania. *World Development,* **142**, 105345. https://doi.org/10.1016/j.worlddev.2020.105345

Donath, L., Morrissey, O., & Owens, T. (2021a). Does the Pay Period Matter in Estimating Returns to Schooling? Evidence from East Africa*,* School of Economics, University of Nottingham: *CREDIT Research Paper 21/01*.

Donath, L., Morrissey, O., & Owens, T. (2021b). Youth Education and Household Welfare in Tanzania*,* School of Economics, University of Nottingham: *CREDIT Research Paper 21/04*.

Firpo, S. P., Fortin, N. M., & Lemieux, T. (2009). Unconditional Quantile Regressions. *Econometrica*, **77**(3), 953–973. https://doi.org/10.3982/ecta6822

Firpo, S. P., Fortin, N. M., & Lemieux, T. (2018). Decomposing wage distributions using recentered influence function regressions. *Econometrics,* **6**(2), 1–41. https://doi.org/10.3390/econometrics6020028

Fortin, N., Lemieux, T., & Firpo, S. (2011). Decomposition Methods in Economics. In O. Ashenfelter & D. Card (Eds.), *Handbook of Labor Economics Volume 4* (pp. 1–102). Amsterdam: Elsevier https://doi.org/10.1016/S0169-7218(11)00407-2

Girma, S., & Kedir, A. (2005). Heterogeneity in returns to schooling : Econometric evidence from Ethiopia. *The Journal of Development Studies*, *41*(8), 1405–1416. https://doi.org/10.1080/00220380500187026

Himaz, R., & Aturupana, H. (2018). Schooling and household welfare: The case of Sri Lanka from 1990 to 2006. *Review of Development Economics,* **22**(2), 592–609. https://doi.org/10.1111/rode.12355

Kan, S. & Klasen , S. (2021), Evaluating universal primary education in Uganda: School fee abolition and educational outcomes, *Review of Development Economics*, 25: 116–147, DOI: 10.1111/rode.12725

Khan, R., & Morrissey, O. (2020). Income diversification and household welfare in Tanzania 2008 – 13, UNU-WIDER, Helsinki: *WIDER Working Papers 2020/110*.

Lassibille, G., & Tan, J. P. (2005). The Returns to Education in Rwanda. *Journal of African Economies*, *14*(1), 92–116. https://doi.org/10.1093/jae/ejh035

Leyaro, V & Joseph, C. (2019), Employment Mobility and Returns to Technical and Vocational Training: Empirical Evidence for Tanzania, School of Economics, University of Nottingham: *CREDIT Research Paper 19/03*

Psacharopoulos, G., & Patrinos, H. A. (2018). Returns to investment in education: a decennial review of the global literature. *Education Economics*, *26*(5), 445–458. https://doi.org/10.1080/09645292.2018.1484426

Oaxaca, R. (1973). Male-Female Wage Differentials in Urban Labor Markets. *International Economic Review,* **14**(3), 693–709.

Ramadan, R., Hlasny, V., & 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. https://doi.org/10.1111/roiw.12396

Rios-Avila, F. (2020). Recentered influence functions (RIFs) in Stata: RIF regression and RIF decomposition. *Stata Journal,* **20**(1), 51–94. https://doi.org/10.1177/1536867X20909690

Serneels, P., Beegle, K., & Dillon, A. (2017). Do returns to education depend on how and whom you ask? *Economics of Education Review*, *60*, 5–19. https://doi.org/10.1016/j.econedurev.2017.07.010

Skoufias, E., & 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. https://doi.org/10.1093/jeg/lbq030

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. http://www.tanzania.go.tz/economicsurvey1/2006/index.html (Accessed 10 September 2011).

URT. (2015). *Tanzania Integrated Labour Force Survey 2014*. National Bureau of Statistics, Dar es Salaam.

URT. (2016). *National Basic Education Statistics in Tanzania (BEST) 2012 - 2016*. Ministry of Education, Science and Technology, Dar es Salaam.

URT. (2018). *National population projections*. National Bureau of Statistics and Office of the Chief Government Statistician, Dar es Salaam and 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. https://doi.org/10.1016/j.worlddev.2013.05.005

Van den Broeck, G., and T. Kilic (2019). Dynamics of Off-Farm Employment in Sub-Saharan Africa: A Gender Perspective. *World Development* 119: 81–99

## Appendices

In the main paper youth are defined as those aged 15-25 in 2018 as they should have attended primary school after UPE was implemented and would therefore be the principal beneficiaries of the expansion in access to education. As the official definition of youth in Tanzania is 15-35, some of those aged 26-35 could have benefited (either from repeating some primary school or the reforms in secondary school) this appendix reports the same analysis for broader youth (15-35), with some separate analysis for the 26-35 age group. Appendix A provides descriptive statistics and Appendix B the econometric analysis.

### Appendix A Data and Descriptive Statistics

As noted in the main paper we use data from the Tanzanian Household Budget Surveys (HBS) for 2001 (collected from May 2000 to June 2001) and 2018 (December 2017 to November 2018), with usable samples of 22,022 and 9,552 households respectively. The HBS is used because Tanzanian surveys with data on individual earnings have limited samples. The 2014 Tanzanian Integrated Labour Force Survey, for example, has a sample of just over 19,000 (with 22% in formal employment, 31% in agriculture, 36% classed as informal employment and 11% classed as unemployed) and data on income for some 14,300 (Leyaro & Joseph, 2019, p23 and Table 7). However, samples are much smaller to compare youth in 2000 (pre-UPE) with youth in 2014 who benefitted from UPE (Leyaro & Joseph, 2019, do not report the age distribution) and there is no labour force survey with large samples giving earnings for early 2000s. The four waves of the Tanzania National Panel Surveys (TNPS) from 2008 to 2016 have limited coverage of agricultural and non-wage (informal or own account) employment and samples are small for 'treated youth' (who benefitted from UPE). The youth in the TNPS (pooling the four waves from 2008 to 2016) who would have at least partially benefited from the 2001 UPE reform (born after 1988) amount to 2,677 observations out of the 11,215 observations but only 487 observations would have fully benefited (7 years or younger in 2001, or born after 1995). These surveys do not go back to the early 2000s so do not permit comparing youth before UPE to youth who benefitted.

Households are categorised according to the head's age group. Youth households include all households headed by a youth aged between 15 and 35, and adult households include all households headed by an adult aged over 35 (who won't have benefitted from the expansion of education). The welfare indicator is measured at the household level as the ratio of the per adult equivalent household consumption to the national poverty line (CPL) and is assigned to the head of household. Education is the household head's level of education measured in years. Variables are as defined in the main paper.

Table A1 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. In 2001, there is no difference in terms of rural location or share with female heads; by 2018, the share of rural households and female heads is slightly lower for youth. A slightly greater proportion of youth heads are married. 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. The CPL increased for both groups in 2018 reflecting the general reduction in poverty compared to 2001.

Youth heads have more years of schooling in both years. It is notable that the proportion of youth heads with no education remained stable around a fifth (higher than for 15-35 youth so mostly in the 26-35 age) whereas the proportion of adult households with no education fell considerably to 27%, although it remained higher. The proportion of youth households with

completed primary was twice that for adults in 2001, with similar small levels completing secondary and a greater (but very small) proportion of adults with post-secondary. By 2018, 24% of youth had post-primary education compared to 11% of adults.

**Table A1: Summary Statistics by Age Group (Youth 15-35) and Year**

| *Characteristics* | 2001 | | | 2018 | | |
|---|---|---|---|---|---|---|
| | Youth | Adult | Difference | Youth | Adult | Difference |
| *Household* | | | | | | |
| 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** |
| Size | 5.18 | 6.89 | 1.70*** | 4.75 | 6.59 | 1.84*** |
| *Head* | | | | | | |
| Education (yrs) | 6.04 | 4.24 | -1.80*** | 6.73 | 5.59 | -1.14*** |
| No education | 0.18 | 0.42 | 0.24*** | 0.19 | 0.27 | 0.08*** |
| Some primary | 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* |
| Some secondary | 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*** |
| Post-secondary | 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 | 003* | 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 | - |

*Notes:* Observations refer to number of households headed by youth (15-35) or adults (over 35); data weighted using survey weights; mean value for continuous variables and % share of sample for binary indicators. CPL is household consumption relative to the poverty line. 'Difference' is the value for adult headed households minus the corresponding value for youth headed household (with significance * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ based on Lincom's test of mean differences).

*Source*: Author's calculations from HBS 2001 and 2018.

Table A2 compares the same characteristics of youth and adult headed households but considers the difference for each between the years (youth means in 2001 compared to 2018 and similarly for adults). Youth's average household consumption increased by 24% and relative poverty rates declined significantly. Youth education also increased significantly, with the most pronounced increase at post-primary education levels (the share with only completed primary fell to below half). 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 from 1% to 4%. A large proportion of secondary school students drop out (6% of the youth in 2018) so never complete the level. The national qualifying exam in the second year of secondary school, which requires students who fail to repeat the year, may be one of the reasons. For adult headed households, the increase in CPL was even greater, as was the decline in the share in poverty, although welfare remained lower than for youth households. Educational attainment also increased significantly. The comparison for adult headed households is complicated by the fact that some adults in 2018 will have been in the youth category in 2001 (see Table 2 in the main paper).

**Table A2: Summary Statistics by Age (Youth 15-35) Between Years**

| *Characteristics* | Youth (15-35) | | | Adult | | |
|---|---|---|---|---|---|---|
| | 2001 | 2018 | Difference | 2001 | 2018 | Difference |
| *Household* | | | | | | |
| 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*** |
| Size | 5.18 | 4.75 | 0.43*** | 6.89 | 6.59 | -0.30*** |
| *Head* | | | | | | |
| Education (yrs) | 6.04 | 6.73 | 0.69*** | 4.24 | 5.59 | 1.35*** |
| No education | 0.18 | 0.19 | 0.01 | 0.42 | 0.27 | -0.15*** |
| Some primary | 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*** |
| Some secondary | 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*** |
| Post-secondary | 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 | - |

*Notes:* As for Table A1 except difference is the 2018 value minus the corresponding value for 2001.

*Source*: Author's calculations from HBS 2001 and 2018.

**Figure A1: CPL by Age (Youth 15-35) and Year**

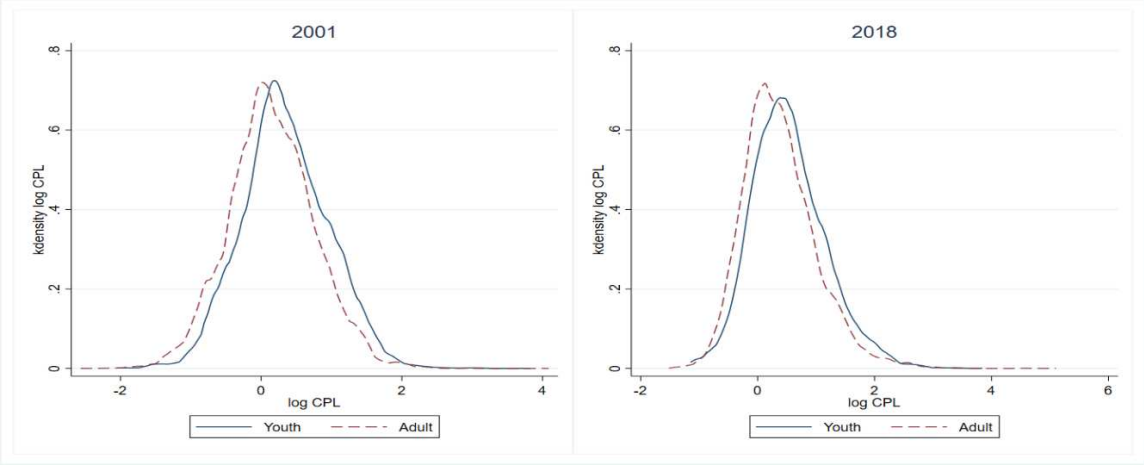
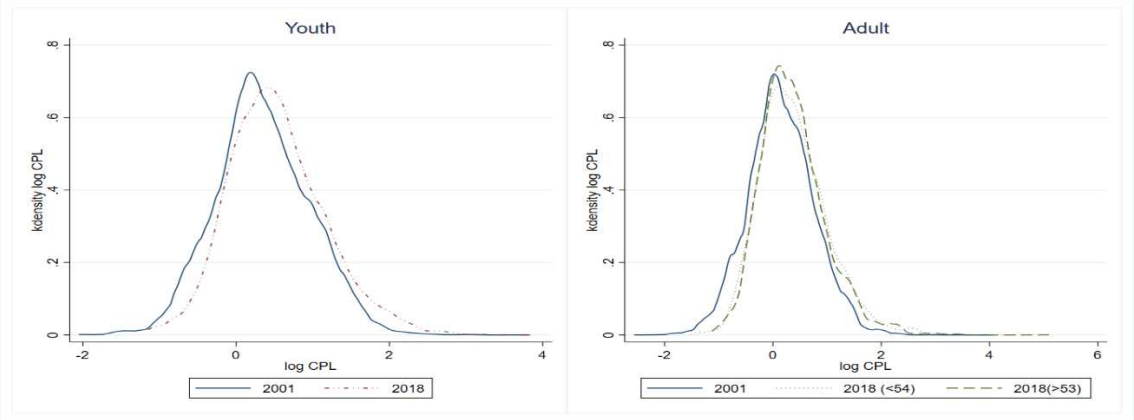


**Figure A2: CPL between Years by Age (Youth 15-35)**




Density plots in Figure A1 shows that CPL was higher than adults in both years and the peak of the distribution shifted more (to right) for youth although the gap between youth and adults narrowed to the right of 0 (national average poverty). The increase in youth consumption is shown in the left panel of Figure A2 (and appears greater for households with relatively lower CPL). The right panel shows the distribution for adults, separating two groups (as in Table 2). Both groups of households have significantly higher consumption in 2018, but the 35-53 age group enjoys slightly higher consumption than the 54 years and above.

**Table A3: Characteristics of Youth (15-35) in Households by Year**

| *Characteristics* | 2001 (15-35) | | | 2018 | | |
|---|---|---|---|---|---|---|
| | Head | Other | Difference | Head | Other | Difference |
| *Household* | | | | | | |
| 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*** |
| Size | 5.18 | 6.83 | -1.65*** | 4.75 | 6.51 | -1.76*** |
| *Youth* | | | | | | |
| Education (years) | 6.04 | 5.48 | 0.56*** | 6.73 | 7.00 | -0.27*** |
| No education | 0.18 | 0.22 | -0.04*** | 0.19 | 0.16 | 0.03*** |
| Some primary | 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*** |
| Some secondary | 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*** |
| Post-secondary | 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 | - |

*Notes:* As for Table A1 except difference is between youth who head households and other youth in 2001 and 2018.

*Source*: Author's calculations from HBS 2001 and 2018.

Table A3 compares household and youth characteristics of youth who are heads of households to youth who are not (other). Youth heads are more likely to be married and in rural areas; fewer than 20% are female and female heads are more likely to be married.[11] Youth who are heads are significantly older. Youth who do not head a household live in households that have significantly lower consumption and are more likely to be poor, suggesting that relatively higher earning youth are more likely to form their own household (consistent with being older). Youth who are not household heads have more years of schooling than household heads in 2018 but this was not the case in 2001. This shift is most pronounced for secondary education although youth heads are more likely to have post-secondary education. Surprisingly, almost a fifth of youth heads have no education (mostly rural), higher than for those aged 15-25; the proportion for other was higher in 2001 but lower in 2018.[12] This implies that a large number of youth did not benefit from UPE and these are likely to be poor youth headed households.[13]

**Table A4: Summary Statistics for Youth Groups in 2018**

| *Characteristics* | Youth (15-25) | | | Youth (26-32) | | |
|---|---|---|---|---|---|---|
| | Head | Other | Difference | Head | Other | Difference |
| *Household* | | | | | | |
| CPL | 2.27 | 1.90 | 0.37*** | 2.10 | 2.12 | -0.02 |
| Poor | 0.14 | 0.26 | -0.12*** | 0.18 | 0.22 | -0.04*** |
| Rural | 0.73 | 0.64 | 0.09*** | 0.67 | 0.60 | 0.07*** |
| Size | 3.69 | 6.64 | -2.95*** | 4.65 | 6.15 | -2.95*** |
| *Head* | | | | | | |
| Education (yrs) | 6.85 | 7.15 | -0.27*** | 6.72 | 6.83 | -0.11 |
| No education | 0.15 | 0.13 | 0.02*** | 0.19 | 0.19 | 0.00 |
| Some primary | 0.11 | 0.10 | 0.01*** | 0.07 | 0.06 | 0.01** |
| Primary | 0.49 | 0.46 | 0.03*** | 0.48 | 0.49 | -0.01 |
| Some secondary | 0.07 | 0.12 | -0.05*** | 0.07 | 0.04 | -0.03 |
| Secondary | 0.19 | 0.17 | 0.02*** | 0.16 | 0.17 | -0.01*** |
| Post-secondary | 0.00 | 0.02 | -0.02*** | 0.03 | 0.05 | -0.02*** |
| Age | 23.03 | 19.19 | 3.84*** | 29.52 | 28.79 | 0.73*** |
| Female | 0.19 | 0.55 | -0.36*** | 0.19 | 0.76 | -0.57*** |
| Married | 0.75 | 0.24 | 0.51*** | 0.84 | 0.72 | 0.12*** |
| *Observations* | 473 | 8,121 | - | 1,353 | 2,531 | - |

*Notes:* As for Table A3 except for different age groups in 2018.

*Source*: Author's calculations from HBS 2001 and 2018.

11 Of the 2,393 female heads, 830 (34%) are married whereas 15% (6,546) of the 42,971 other youths are married to a youth head.

12 Youth heads in rural areas are more likely to have no schooling (less than 3 years of primary education), 23% compared to 5% of urban youth heads; 24% of female heads have no schooling compared to 17% of males.

13 Poor households are more likely to be headed by a youth with no education – 31% compared to 14% of the non-poor households having a youth head with no education.

As the share of households headed by youth decreased by ten percentage points between 2001 and 2018, it is possible 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. While six per cent of youth heads had completed at least secondary education in 2001 this rose to 18% in 2018, consistent with observing that youth heads are older, and youth headed households tend to have higher consumption.

Table A4 compares characteristics of heads and other for younger (15-25) and 'older' (26-32, who may have had some benefit from reforms) youth in 2018. Compared to younger other, households of younger heads have higher CPL, much lower size and are far less likely to be poor, although a higher share are rural. Younger heads are also older, more likely to be married and less likely to be female; and have slightly less education (evident for some secondary and post-secondary). Compared to older other, households of older heads have similar CPL, much lower size and are less likely to be poor, although a higher share are rural. Older heads are only slightly older or more likely to be married, far less likely to be female; and have slightly less education (notably post-secondary). Comparing younger and older heads, households of younger heads have higher CPL, lower size, and are less likely to be poor or rural. Younger heads are slightly less likely to be married, have slightly more education and are more likely to have some primary and completed secondary (but less likely to have post-secondary, perhaps because they are younger); identical shares are female. This suggests that the benefits of expanding education increased the proportion of the treated age group getting some primary and continuing to complete secondary.

**Table A5: Characteristics of Youth (15-25) Heads of Household by Year**

| *Characteristics* | 2001 | | | 2018 | | |
|---|---|---|---|---|---|---|
| | Head | Other | Difference | Head | Other | Difference |
| *Household* | | | | | | |
| CPL | 1.86 | 1.44 | 0.42*** | 2.27 | 1.90 | 0.37*** |
| Poor | 0.25 | 0.38 | -0.13*** | 0.14 | 0.26 | -0.12*** |
| Rural | 0.79 | 0.77 | 0.02*** | 0.73 | 0.64 | 0.09*** |
| Size | 4.26 | 6.92 | -2.66*** | 3.69 | 6.64 | -2.95*** |
| *Youth* | | | | | | |
| Education (years) | 5.98 | 5.45 | 0.53*** | 6.85 | 7.15 | -0.27*** |
| No education | 0.16 | 0.22 | -0.06*** | 0.15 | 0.13 | 0.02*** |
| Some primary | 0.05 | 0.19 | -0.14*** | 0.11 | 0.10 | 0.01*** |
| Primary | 0.76 | 0.54 | 0.22*** | 0.49 | 0.46 | 0.03*** |
| Some secondary | 0.01 | 0.02 | -0.01*** | 0.07 | 0.12 | -0.05*** |
| Secondary | 0.02 | 0.03 | -0.01*** | 0.19 | 0.17 | 0.02*** |
| Post-secondary | 0.00 | 0.01 | -0.01*** | 0.00 | 0.02 | -0.02*** |
| Age | 23.21 | 19.46 | 3.75*** | 23.03 | 19.19 | 3.84*** |
| Female | 0.25 | 0.59 | -0.34*** | 0.19 | 0.55 | -0.36*** |
| Married | 0.72 | 0.30 | 0.42*** | 0.75 | 0.24 | 0.51*** |
| *Observations* | 1,647 | 22,481 | - | 473 | 8,121 | - |

*Notes:* As for Table A1 except difference is between youth (15-25) who head households and other youth.

*Source*: Author's calculations from HBS 2001 and 2018.

Table A5 can be compared to Table 1 in the paper except that here (as in Table A3 for youth aged 15-35) the comparison is between youth aged 15-25 who are heads of households to youth who are not (other). In both years youth headed households had significantly higher welfare (CPL) and were less likely to be poor, than adult households with youth (15-25). The CPL increased for both groups in 2018 reflecting the general reduction in poverty compared to 2001, although the difference declined. Compared to other, a slightly higher proportion of youth headed households were rural, fewer had female heads, were far more likely to be married, and had smaller household size – all of the differences were greater in 2018.

In terms of education endowment, there is notable change: years of education increased (by more than 1 year on average) but by less for youth headed households, which were more educated in 2001 but less educated than youth in other households by 2018. Youth have more years of schooling in both years. Proportions with no education and completed primary fell, whereas proportions with some or completed secondary rose, especially for youth in other households. This is consistent with a tendency for more educated individuals to delay forming households.

## Appendix B: Additional Analysis

**Table B1: Returns to Education by Age (Youth 15-35) 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) |
| $Sch^2$ | 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) |
| $Age^2$ | -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) |
| *lnSize* | -0.516*** | -0.451*** | -0.523*** | -0.473*** |
| | (0.011) | (0.008) | (0.021) | (0.011) |
| _cons | -0.005 | 1.153*** | 0.331 | 0.656*** |
| | (0.219) | (0.082) | (0.396) | (0.099) |
| AME(*Sch*) | 0.056*** | 0.043*** | 0.042*** | 0.027*** |
| | (0.002) | (0.001) | (0.003) | (0.002) |
| *N* | 8,039 | 13,983 | 2,507 | 6,945 |
| $R^2$ | 0.41 | 0.40 | 0.50 | 0.44 |

*Notes:* AME(*Sch*) is the average mean effect of a year of schooling; Household size is in logs (*lnSize*). Other controls included (not reported) are livestock per capita, region of residence and ownership of assets. Standard errors in parentheses (significance indicated by * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

Table B1 presents the estimates of returns (in terms of CPL) to head's education, where youth are the 15-35 age group. In 2001, years of schooling (*Sch*) is positively and significantly correlated with welfare for both youth and adult headed households, but negatively correlated for 2018 (up to some threshold given the positive and significant $Sch^2$). The significance of $Sch^2$ implies a strong convex relationship — each extra year of schooling is associated with higher welfare than the previous year (at least beyond some threshold if the coefficient on schooling is negative). All other included regressors have the expected sign. Coefficients on *Sch*, *Age*, $Age^2$ and *Married* are smaller and sometimes insignificant in 2018 compared to 2001. In contrast, coefficients are larger and positive for females, suggesting female headed households benefitted relative to males, and for $Sch^2$ whereas they are larger but negative for rural and household size (both associated with lower welfare).

**Figure B1: Implied Returns to Education by Age (Youth 15-35) and Year**

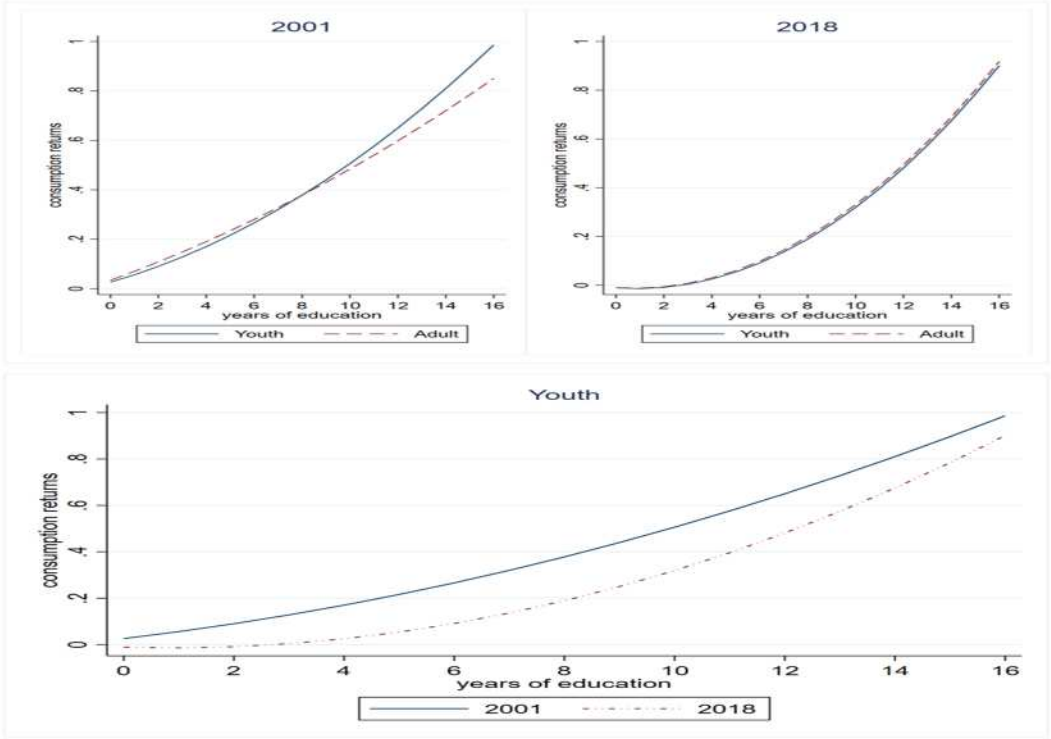



**Figure B2: Implied Returns to Education by Youth Groups, 2018**

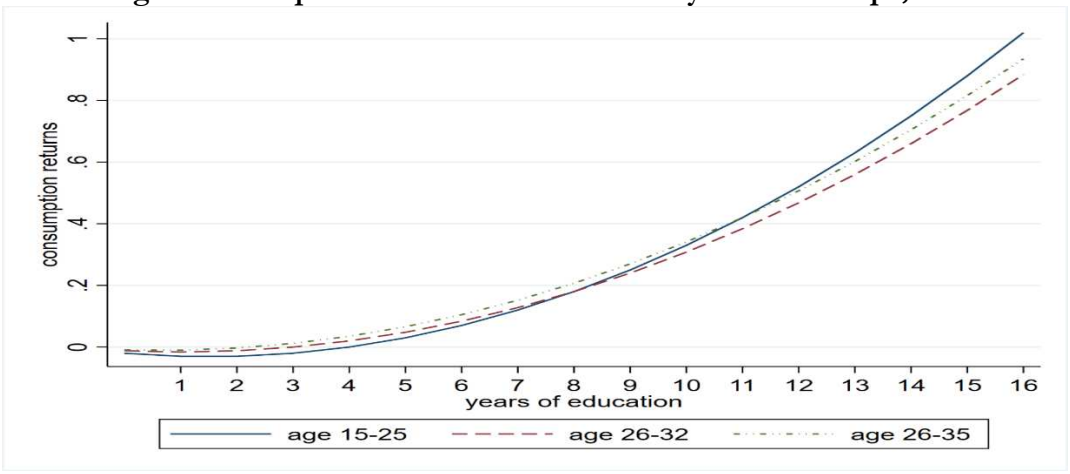


Figure B1 plots the implied returns combining coefficients on *Sch* and $Sch^2$, showing that in 2018 returns are positive beyond about three years of school. The top panel shows that youth headed households in 2001 had higher returns to post-primary education than adult headed households and this advantage disappeared in 2018; although both groups had more years of education returns fell (lower for mean years of education in 2018 compared to 2001). The bottom panel shows how 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 —returns to education to decline as education attainment increases in the population.

Figure B2 plots the implied returns for youth divided into three age groups in 2018 (results corresponding to Table B1 are available on request, showing that returns are higher for the 15-25 age group – only coefficients on $Sch^2$ are significant – and increase with age for that group only). The younger the group the more years of education are required before returns turn positive: 4 years for 15-25, 3 years for 26-32 and 2 years for 26-35 (the 33-35 group is too small to separate). Returns for the 15-25 group rise above the 26-32 group after 8 years (corresponding to completed primary) and above the 26-35 group after 11 years, corresponding to completed secondary. This indicates that significant gains in schooling (higher returns compared to older groups) only arise for younger youth able to continue through lower secondary. Although none of the younger youth heads had post-secondary education (Table A4) the implication of Figure B2 is those who are undertaking it will benefit through higher returns.

The differences in mean CPL between youth and adult headed households are investigated using RIF regressions to decompose into the part due to differences in characteristics (explained) and the part due to differences in returns to these characteristics (unexplained). Each of the two parts are broken down into two subparts: the explained into pure explained and the specification error; and the unexplained into pure unexplained and reweighting error. If both the specification and the reweighting error are small and insignificant this implies the model is correctly specified and reweighted, which transpires to be the case (as for the analysis in the paper).

Table B2 presents the decomposition results by year. The coefficients of the education variables (*sch* and $sch^2$) are aggregated into one variable 'education'; the coefficient of *age* and $age^2$ into 'headage'; and ownership of assets, livestock per capita, and dummies for regions of residence into 'other controls'.[14] Only the explained component is significant in both years implying that it is only the difference in endowment that explains differences in welfare between the two age groups. The breakdown of the 'pure explained' component reveals that education is positive and significant – youths heading households have significantly better education attainment than adults heading households. A significant portion of the welfare difference between youth and adult headed households is attributable to differences in educational attainment: of the 'pure explained' welfare differences of 0.251 and 0.151 in 2001 and 2018 respectively, approximately 40% and 32% are attributed to differences in educational attainment between youths and adults. 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. Household size is the only other characteristic that has a large contribution to the explained difference and rural location is also significant in both years, although female and married are also significant. The difference in returns to education does not have a significant effect on welfare. Household size (2001) and married (2018) are the only significant factors for returns.

14 The Stata command *oaxaca_rif* is calibrated for this common approach in decomposition.

**Table B2: Reweighted RIF Decomposition Within Years (Youth 15-35)**

| | (1)<br>2001 | (2)<br>2018 |
|---|---|---|
| *Overall* | | |
| Youth (15-35) | 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*** |
| lnSize | 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* |
| lnSize | -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 (youth) | 8,039 | 2,507 |
| N2 (adult) | 13,983 | 6,945 |
| N (all observations) | 22,022 | 9,452 |

*Notes:* 'Other controls' is the aggregate effect of livestock per capita, region of residence and ownership of assets. Binary variables are normalised; significance * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The RIF decomposition by gender and rural/urban residence is in Table B3 (results comparing different youth age groups are similar and available on request). The results suggest that the difference in welfare between the two periods is mainly attributed to differences in characteristics with evidence of heterogeneity of the effects of education (the only consistently significant endowment) across groups (covariates are included in the baseline regression). Other controls are a significant endowment for males, rural and urban, presumably reflecting an effect of assets (accumulation). Education is significant for the unexplained part for rural and female, the negative value indicating the relative decline in returns.

The results in column (1) 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 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.

**Table B3: Reweighted RIF Decomposition for Youth (15-35)**

| | (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*** | -0.180 |
| 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 (2018) | 2,507 | 473 | 2,034 | 1,653 | 854 |
| N2 (2001) | 8,039 | 1,920 | 6,119 | 2,687 | 5,352 |
| N (all observations) | 10,546 | 2,393 | 8,153 | 4,340 | 6,206 |

*Notes:* As for Table B2.

**Robustness Checks**

Endogeneity of education due to unobserved ability is a concern, as is the potential endogeneity of education of youth selection into heading a household, as factors associated with higher welfare, such as education, are also associated with a higher likelihood of youth to form a household. We use a linear probability model (LPM) to assess whether education influences the likelihood for a youth head of household. Although by no means a panacea, LPM has the advantage of allowing the inclusion of household fixed effects.

**Table B4: LPM Results by Gender and Place of Residence 2001 (Youth 15-35)**

| | (1) All | (2) Female | (3) Male | (4) Rural | (5) Urban |
|---|---|---|---|---|---|
| *Sch* | 0.006*** | 0.003** | 0.003 | 0.007*** | 0.005*** |
| | (0.001) | (0.001) | (0.002) | (0.002) | (0.001) |
| *youth<26* | -0.244*** | -0.118*** | -0.196*** | -0.243*** | -0.240*** |
| | (0.006) | (0.010) | (0.017) | (0.010) | (0.008) |
| *female* | -0.302*** | | | -0.374*** | -0.257*** |
| | (0.005) | | | (0.008) | (0.006) |
| *married* | 0.082*** | -0.033*** | 0.351*** | 0.117*** | 0.069*** |
| | (0.007) | (0.009) | (0.022) | (0.012) | (0.009) |
| *AGR* | -0.019*** | -0.001 | -0.001 | -0.017* | -0.034*** |
| | (0.006) | (0.008) | (0.012) | (0.011) | (0.007) |
| *WAGE* | 0.182*** | 0.109*** | 0.070*** | 0.172*** | 0.189*** |
| | (0.010) | (0.018) | (0.025) | (0.026) | (0.011) |
| *SELF* | 0.156*** | 0.109*** | 0.097*** | 0.045** | 0.184*** |
| | (0.010) | (0.020) | (0.021) | (0.023) | (0.011) |
| _cons | 0.419*** | 0.139*** | 0.296*** | 0.454*** | 0.397*** |
| | (0.010) | (0.014) | (0.023) | (0.016) | (0.012) |
| *N* | 39,542 | 21,873 | 17,669 | 13,486 | 26,056 |
| $R^2$ | 0.72 | 0.89 | 0.92 | 0.69 | 0.73 |

*Note:* Dummies for sectors of main employment are AGR (agriculture), *WAGE* (wage employment) and *SELF* (non-agricultural self-employment), 'inactive' is the omitted category; *<26* is an age dummy =1 if aged less than 26 years and 0 otherwise. Robust standard errors in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$).

The LPM estimates are in Table B4 for 2001 and Table B5 for 2018. Table B4 shows that, after controlling for household fixed effects, more educated youth were significantly more likely to be a household head – an extra year of education is associated with about 0.006 increase on average in the probability of the youth being a head – although it is insignificant for males (and highest for rural). Younger youth, those working in agriculture and females are less likely to be a head, whereas married are more likely. Youth in self or wage employment, both associated with higher household welfare (Khan & Morrissey, 2020), are more likely to be heads (especially for females). Table B5 shows that the strength of effects declined in 2018 for all variables. On average, an extra year of education increases the probability of being a head by 0.003 but is only significant overall and for rural (at 0.004). The significant association between education and headship in 2018 is generally driven by youths residing in rural areas. Younger youth and females are less likely to be a head, whereas married are more likely (except for females, consistent with only 7% of youth heads being married females). Type of employment is insignificant for males, only self-employment is significant for females; any employment is associated with increased likelihood of being a head (overall, rural and urban) but the effect is least for agriculture.

Despite the significance of the coefficients of education in Tables B4 and B5, the sizes are negligible. A coefficient on education of 0.005 (most coefficients are lower, especially in 2018) implies 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 unlikely to significantly affect the estimated coefficients of education in our main results, especially as mean years are relatively low (six years in 2001 and seven in 2018). As in the paper, the low coefficient gives us cautious confidence in the main results.

**Table B5: LPM Results by Gender and Place of Residence 2018 (Youth 15-35)**

| | (1) All | (2) Female | (3) Male | (4) Rural | (5) Urban |
|---|---|---|---|---|---|
| *Sch* | 0.003** | 0.001 | 0.001 | 0.004** | 0.003 |
| | (0.002) | (0.002) | (0.003) | (0.002) | (0.003) |
| *youth<26* | -0.220*** | -0.093*** | -0.152*** | -0.220*** | -0.215*** |
| | (0.011) | (0.021) | (0.027) | (0.013) | (0.021) |
| *female* | -0.348*** | | | -0.356*** | -0.332*** |
| | (0.009) | | | (0.011) | (0.016) |
| *married* | 0.091*** | -0.021 | 0.220*** | 0.095*** | 0.086*** |
| | (0.013) | (0.018) | (0.032) | (0.015) | (0.027) |
| *AGR* | 0.042*** | 0.012 | 0.026 | 0.029** | 0.081** |
| | (0.012) | (0.018) | (0.021) | (0.012) | (0.035) |
| *WAGE* | 0.163*** | 0.051 | 0.062 | 0.152*** | 0.179*** |
| | (0.020) | (0.044) | (0.044) | (0.032) | (0.025) |
| *SELF* | 0.134*** | 0.094*** | 0.023 | 0.097*** | 0.173*** |
| | (0.016) | (0.035) | (0.034) | (0.022) | (0.023) |
| _cons | 0.382*** | 0.103*** | 0.294*** | 0.387*** | 0.371*** |
| | (0.017) | (0.025) | (0.036) | (0.019) | (0.033) |
| *N* | 13,975 | 7,358 | 6,617 | 9,720 | 4,255 |
| $R^2$ | 0.72 | 0.90 | 0.95 | 0.71 | 0.75 |

*Notes:* As for Table B6.