FACULTY OF SOCIAL SCIENCES
UNIVERSITY OF COPENHAGEN

PhD Thesis

Ermias Gebru Weldesenbet

Empirical essays on the long-term outcomes of childhood experiences: poverty, weather shock and migration

Academic Advisors: Mette Ejrnæs
Henrik Hansen

Submitted: February 2022

Contents

Acknowledgements

First of all, I would like to thank God for helping me finalize my study and reach this milestone. Also, I acknowledge my Supervisor, Mette Ejræns, with heartfelt gratitude for the ongoing guidance, supervision, and insights throughout my study journey. She has been instrumental in coining the framework, reviewing and enriching the drafts for all of my theses. As I had a frail health condition during my study, she has been remarkably forbearing and relentless in supporting and guiding my effort to achieve my goal. I have also immensely benefited from the advice and supervision of my co-supervisor, Henrik Hansen. I am very grateful for his support. I would not have achieved this without my supervisors' valuable input, continuous follow-up, and committed support. To this, I am deeply grateful for and would like to say thanks so much. I am also grateful to the PhD program coordinator Jakob Roland Munch for his support during my study. I want to thank Helene Bie Lilleør for allowing me to use the KHDS data before it was available to the public domain. I am thankful to Biniam Bedasso and Nonso Obikili, who collaborated with me in writing one of the papers. I am particularly appreciative to Biniam Bedasso for his consolidating ideas and insightful comments on my two other papers. Lastly, I want to thank my mother, wife, family, and friends for being there when I needed them the most. I owe a special tribute to my mother, who had no opportunity to go to school herself but has taught me the importance of education since childhood. Her courage and persistence in life have been my source of inspiration.

Summary

This PhD thesis contains three self-contained chapters, focusing on the long-term outcome of different childhood experiences. As children are vulnerable to early life experiences, whether they have a good or bad start has long-term implications on their life. For that reason, understanding childhood conditions that affect later life outcomes is of paramount importance to policymaking. While there is a large body of literature on developed countries, the empirical investigation of early life conditions for developing countries, especially sub-Saharan Africa, has been constrained mainly due to the lack of longitudinal data following children for a long time. More recently, studies focusing on developing countries are beginning to emerge with the maturity of some longitudinal datasets tracking children, even if it remains much to discover in the context of sub-Saharan Africa. The thesis seeks to address this lacuna in African research by exploring three diverse childhood conditions affecting outcomes in adulthood.

The first chapter sets the scene for the dissertation by investigating the broad experience of growing up in poverty before examining more specific childhood conditions. The chapter seeks to understand the extent to which poverty is passed from parents to children in young adulthood. It also investigates one of the routes by which parental poverty affects economic status in adulthood. The second chapter explores the long-term impact of malnutrition due to weather shock. As education is a critical factor that helps poor children escape poverty, the third chapter examines whether having migrant family members affects children's education.

The first two chapters depend on a unique panel of household data and historical

rainfall data from Tanzania. On the other hand, the last chapter relies on household migration history data from Nigeria. Because a linked set of processes from childhood to adulthood are involved in an individual's socioeconomic achievement, it is a daunting task to establish a causal interpretation of the link between childhood experience and adulthood outcomes. When the data allows it, the study tries to provide causal analysis. The chapters are discussed in more detail below.

Chapter 1, **"Intergenerational Poverty Transmission in Tanzania: The role of parental resources"**, explores the extent of intergenerational poverty as well as one of the mechanisms of poverty transmission in Tanzania. I find that the risk of falling into poverty remained low, while the probability of escaping from poverty increased between childhood and adulthood. In adulthood, children from poor families have a poverty risk (0.38) three times higher than those from non-poor family backgrounds (0.13). Investigating the role of parental resources in the intergenerational transmission of poverty, I find that parental financial resource in childhood is strongly associated with an individual's poverty risk in early adulthood. The results further indicate that human capital investment in children mediates some of the effects of childhood parental resources on economic status in adulthood. The results imply that interventions supporting low-income families build their children's human capital are essential to break the intergenerational cycle of poverty.

Chapter 2, **"The Long-Term Effects of Early-Life Exposure to Weather Shocks: Evidence from Tanzania"**, examines whether early-life exposure to rainfall shocks has a long-term impact on individuals' health, education, and socioeconomic status in rural Tanzania, where livelihoods heavily depend on rain-fed agriculture. The main finding is that rainfall in the birth year affects an individual's education and socioeconomic status in adulthood. I also find that higher birth-year rainfall leads to significant improvement in childhood nutritional status. The results point to the importance of early childhood nutrition intervention.

Chapter 3, **"Emigration and education: the schooling of the left behind**

in Nigeria (joint with Biniam Bedasso and Nonso Obikili), investigates the impact of family migration on left-behind children educational attainment using household survey data from Nigeria. We find that being in a migrant household increases the probability of completing secondary school and attending some postsecondary education. We also find that belonging to a migrant household increases the probability of own future migration. We further explore channels through which the migration of family members affects education. We provide tentative evidence suggesting that anticipation of own future migration may be behind increased educational attainment.

Resumé (Summary in Danish)

Denne ph.d.-afhandling indeholder tre selvstændige kapitler, der fokuserer på det langsigtede resultat af forskellige barndomserfaringer. Da børn er sårbare over for tidlige livserfaringer, hvad end de har oplevet gode eller negative erfaringer, har det langsigtede konsekvenser for deres liv. Det er derfor af afgørende betydning for politikudformningen, at forstå barndomsforhold, der kan påvirke dem senere i livet. Der findes en stor mængde litteratur om udviklingslande, men kun begrænsede empiriske undersøgelser af tidlige livsbetingelser for ikke-udviklingslande, især Afrika syd for Sahara. Hvilket skyldes, mangel på longitudinelle data, der følger børn i lang tid. For nylig er undersøgelser, der fokuserer på ikke-udviklingslande, begyndt at dukke op med begyndelsen, af nogle longitudinelle datasæt, der sporer børn. Afhandlingen søger at adressere denne lakune i afrikansk forskning ved at udforske tre forskellige barndomsforhold, der påvirker resultater i voksenlivet.

Det første kapitel, handler om, at undersøge den generelle erfaring med, at vokse op i fattigdom, inden undersøgelsen af mere specifikke barndomsforhold. Kapitlet søger at forstå, i hvilket omfang i voksenlivet, fattigdom overføres fra forældre til børn. Den undersøger også en af de måder, hvorpå forældres fattigdom påvirker økonomisk status i voksenalderen. Det andet kapitel udforsker den langsigtede virkning af fejlernæring foårsaget af vejrchok. Da uddannelse er en kritisk faktor, der hjælper fattige børn med at slippe ud af fattigdom, undersøger det tredje kapitel, om det at have indvandrerfamiliemedlemmer påvirker børns uddannelse.

De første to kapitler tager udgangspunkt i, et unikt panel af husstandsdata og

historiske nedbørsdata fra Tanzania. Hvor det sidste kapitel omhandler husstands migrations historiedata fra Nigeria. Grundet et sammenhængende sæt af processer fra barndom til voksenliv er involveret i et individs socioøkonomiske præstation, er det en svær opgave at etablere en kausal fortolkning af sammenhængen mellem barndomserfaring og voksenlivsresultater. Når dataene tillader det, forsøger undersøgelsen at give en kausal analyse. Kapitlerne diskuteres mere detaljeret nedenfor.

Kapitel 1, **"Intergenerationel fattigdomstransmission i Tanzania: Forældreressourcernes rolle"**, udforsker omfanget af fattigdom mellem generationerne såvel som en af mekanismerne for fattigdomsoverførsel i Tanzania. Jeg oplever, at risikoen for at havne i fattigdom forblev lav, hvorimod sandsynligheden for at slippe ud af fattigdom steg mellem barndom og voksenliv. I voksenalderen har børn fra fattige familier en fattigdomsrisiko (0,38) tre gange højere end børn fra en ikke-fattig familiebaggrund (0,13). Ved at undersøge forældreressourcernes rolle i den intergenerationelle overførsel af fattigdom, finder jeg, at forældrenes økonomiske ressourcer i barndommen er stærkt forbundet med et individs fattigdomsrisiko i den tidlige voksenalder. Resultaterne indikerer endvidere, at menneskelig kapitalinvestering i børn medierer nogle af virkningerne af barndommens forældreressourcer på økonomisk status i voksenlivet. Resultaterne antyder, at interventioner, der støtter lavindkomstfamilier, og opbygger deres børns menneskelige kapital, er afgørende for at bryde fattigdomscyklussen mellem generationerne.

Kapitel 2, **"De langsigtede virkninger af tidlig livseksponering af vejrchock: Evidens fra Tanzania"**, undersøger, om tidlig livseksponering af regnchok har en langsigtet indvirkning på individers sundhed, uddannelse og socioøkonomiske status i landdistrikterne i Tanzania, hvor levebrød i høj grad afhænger af regnfodret landbrug. Hovedfundet er, at nedbør i fødselsåret påvirker en persons uddannelse og socioøkonomiske status i voksenalderen. Jeg finder også, at højere fødselsårsregn fører til en betydelig forbedring af barndommens ernæringsstatus. Resultaterne peger på vigtigheden af en tidlig ernæringsintervention i barndom.

Kapitel 3, **"Emigration og uddannelse: de tilbageståendes skolegang i Nigeria"** (sammen med Biniam Bedasso og Nonso Obikili), undersøgers virkningen af familiemigration på efterladte børns uddannelsesniveau ved hjælp af husstandsundersøgelsesdata fra Nigeria. Vi oplever, at det at være i en migranthusstand øger sandsynligheden for at fuldføre gymnasiet og deltage i nogle videregående uddannelser. Vi finder også, at tilhørsforhold til en migranthusstand øger sandsynligheden for egen fremtidig migration. Vi undersøger yderligere kanaler, hvorigennem migration af familiemedlemmer påvirker uddannelse. Vi giver foreløbige beviser, der tyder på, at forventning om egen fremtidig migration kan ligge bag øget uddannelsesniveau.

Chapter 1

Intergenerational Poverty Transmission in Tanzania: *The role of parental resources*

Intergenerational Poverty Transmission in Tanzania: The role of parental resources

Ermias Gebru Weldesenbet*

Department of Economics, University of Copenhagen

February 23, 2022

Abstract

This paper uses long-running household panel data to explore the extent of intergenerational poverty as well as one of the mechanisms of poverty transmission in Tanzania. We use the mixed latent Markov model to account for measurement error and heterogeneity in poverty experience. We find that the risk of falling into poverty remained low, while the probability of escaping from poverty increased between childhood and adulthood. In adulthood, children from poor families have a poverty risk (0.38) three times higher than those from non-poor family backgrounds (0.13). The analysis also indicates that transitorily poor have a lower risk in adulthood than those who sustained poverty during childhood. Investigating the role of parental resources in the intergenerational transmission of poverty, we find that parental financial resource in childhood is strongly associated with an individual's poverty risk in early adulthood. The results further indicate that human capital investment in children mediates some of the effects of childhood parental resources on economic status in adulthood. We find a strong association between parental financial resources in childhood and children's human capital in young adulthood. Based on the result, we suggest that interventions supporting low-income families build their children's human capital are essential to break the intergenerational cycle of poverty.

Keywords: Intergenerational Poverty; Parental Resources; Human Capital; Tanzania. **JEL Code:** C52, I32, O15, O55

*Email:ermias.weldesenbet@econ.ku.dk

1 Introduction

Intergenerational transmission of economic status has been a topic of interest to researchers and policymakers alike. While there is a large body of literature on developed countries focusing on intergenerational income mobility, the intergenerational aspect of poverty has received much interest, mainly in the United States and some European countries with long-running household panel data (Rodgers, 1995; Moore, 2001; Harper et al., 2003; Gibbons and Blanden, 2006; Duarte et al., 2018).Previous research from these countries suggests children growing up in low-income families are more likely than those from more advantaged families to experience poor human capital outcomes later in life, consequently remaining in poverty during adulthood (Rodgers, 1995; Corcoran, 2001; Airio et al., 2005; Corak, 2006; Papanastasiou and Papatheodorou, 2010; Duarte et al., 2018). The related literature on developing countries remains scarce mainly due to the lack of longitudinal data tracking households for a long time (Baulch, 2011; Bird, 2013).

In Sub-Saharan Africa (SSA), because most of the existing household panel surveys either span for a short period or do not track children in the original sample households after leaving their parents' homes, most household data are not suitable for analyzing poverty's intergenerational dimension (Jenkins and Siedler, 2007; Bird, 2007; Baulch and Hoddinott, 2000). As a result, the early research on intergenerational transmission of poverty (ITP) in SSA focused mainly on qualitative work (Cooper, 2010). With the gradual arrival of new longitudinal datasets, some recent quantitative studies are trying to understand the nature of ITP in SSA (Behrman et al., 2017). Nevertheless, the issue remains not adequately studied in the context of SSA. Therefore, this study attempts to contribute to our understanding of the ITP in Africa using long-run panel data from Tanzania covering repeated observations of parents and children from childhood to adulthood. The study has two objectives. First, it explores the extent to which poverty is passed from parents to children in young adulthood in Tanzania. Second, it investigates the role of parental resources in the ITP and identifies mechanisms linking childhood parental resources and economic status in adulthood.

Several explanations are provided in the literature for the ITP. They are broadly categorized into four sets: economic, cultural-behavioural, policy-related and structural factors (Corcoran et al., 1997; Stenberg, 2000). The first explanation attaches importance to the role of economic resources; the second to inherited beliefs, values, attitudes and behaviours; the third to welfare dependency; and the fourth to economic, social and political structures in the ITP. Moreover, new perspectives provide additional explanations. For instance, one argument suggests that low parental IQ causes children to be poor in adulthood. Another view related to the first one explains ITP based on the neurobiological impacts of poverty on children (Gatzke-Kopp and Creavey, 2017). This paper focuses on the first explanation because there is a significant emphasis in the literature on this approach to explain ITP in developing countries. The micro-level factors in general and household level factors that affect economic resources avilable to the household in particular underlined as the main domain of ITP process in developing countries.

The first approach explains ITP in terms of lack of adequate parental resource transfer and investment in children's human capital. Children who grow up in a more affluent family are more likely to acquire higher human capital, which increases their socioeconomic achievement in adulthood. In contrast, children who grow up in poverty are less likely to attain a high level of human capital, making it difficult to escape poverty in adulthood. According to this view, household-level factors such as parental education, family size, and family structure affect children's human capital, thus accounting for their poverty as adults. In addition to affecting the availability of economic resources, the factors mentioned above also influence children's chances of escaping poverty as adults through parenting practices and role models. As a secondary objective, the study attempts to provide suggestive evidence on the role of parental resources in light of this argument.

The analysis has two phases. The first phase estimates poverty transitions between childhood and adulthood using Markov chain models. We apply various discrete-time Markov chain models and select the model that best suits our data. Using the best model, we then examine poverty risk in adulthood for those who experienced poverty during childhood and those who did not experience poverty. In the second phase, we follow a two-stage regression approach to investigate one of the routes by which parental poverty affects economic status in adulthood. First, we try to establish the role of parental resources in intergenerational transmission using a sustainability-livelihood analysis framework (Sharp, 2003). This is followed by analyzing the relationship between parental resources in childhood and children's human capital outcomes in adulthood. In rural areas of SSA where livelihood mainly depends on agriculture, the role of parental asset transfer, especially land, in the inter-generational transmission of economic status is overemphasized. Thus, the second stage of the analysis helps us explore to what extent human capital mediates parental resource effect on children's economic status in rural settings in SSA.

We use unique panel data from Kagera Health and Development Survey (KHDS) collected by the World Bank and the Rockwool Foundation Unit to conduct the empirical analysis. The survey followed children from childhood (1991) to young adulthood (2010). We use children who were below 15 years during 1991-94 for the study. This study applies the national basic needs poverty line to determine children's poverty status across different waves. We measure parental financial resources using household per capita consumption and parental human capital using parental schooling and parental height. Similarly, we measure children's human capital using years of education and height for age (HAZ) scores and height measurement.

We find that the latent Markov model, which allows for measurement error and the latent mover stayer model, which accounts for heterogeneity and measurement error, best fit our data. Error corrected estimates show about 57 percent of the children were raised in poor childhood families . There was significant movement out of poverty while the risk of falling into poverty remained relatively low between childhood and adulthood. We also find that children from poor families have poverty risk (0.38) in young adulthood, three times higher than those from non-poor family backgrounds (0.13). Children who were transitorily poor during childhood have a lower risk in adulthood than those who sustained poverty. The sensitivity tests show that our main results are reasonably robust to various potential sample errors and changes in specification.

Moreover, the regression analysis strongly supports the explanation based on parental resources. After controlling for different forms of capital, parental financial resources in childhood is strongly associated with an individual's poverty risk in early adulthood. Our result further indicates that human capital investment in children mediates some of the effects of childhood parental resources on economic status in adulthood. We find a strong association between parental financial resources in childhood and children's human capital in young adulthood. In particular, we find that a child's years of education and height in young adulthood increase with per capita household consumption in the childhood family. The result points to the importance of interventions supporting low-income families to build their children human capital in reducing poverty and increasing social mobility.

The remainder of the paper proceeds as follows. The next section describes the study's data. Section 3 presents the methodology used to estimate poverty transitions. Section 4 unveils a summary of the estimation results and provides robustness checks. Finally, the last section offers a conclusion.

2 Data

The study uses long-running household panel data from the Kagera Health and Development Survey (KHDS)[1] , conducted in the northwestern region of Tanzania. The survey was initially designed and implemented by the World Bank and the Muhimbili University College of Health Sciences. The KHDS has six waves of data collected

[1]The KHDS was initially adapted from the World Bank's Living Standards Measurement Study (LSMS) questionnaires

between 1991 and 2010. The first four waves of the data were collected from 1991 to 1994 at six months intervals. Overall, 915 households selected from 51 clusters were interviewed during the baseline survey. Excluding families for which all members were deceased, the survey re-interviewed 93 percent of the baseline households in 2004 and 92 percent in 2010. As the objective in the last two rounds of the survey was to re-interview all baseline household members even if they moved out of their original households, the sample increased to 2700 households in 2004 and 3200 households in 2010 (see Beegle et al. 2006 and De Weerdt et al. 2012, for sampling details). Overall, the re-contact rate is well above most known household panel surveys in developing countries summarized in Alderman et al. (2001)[2].

We restrict the study sample to children aged below 15 years who lived in the survey households during 1991-1994. In the baseline survey, there were 2050 children of head aged less than 15 years living with their parents in 875 households. Out of these, 163 were deceased before 2010 and 243 were not traced during the last round of the survey, leaving us with 1644 children interviewed in 1991-94 and 2010. Because not all households were interviewed an equal number of times during the baseline waves (1991-94), the analysis focuses on four of the six waves (1991, 1994, 2004 and 2010). The final sample consists of 1442 children who have complete information on poverty status for the four waves.

Girls make up half of the analysis sample. The sample average age in childhood (1991) was eight years, with no significant difference between girls and boys. A higher proportion (65 percent) of the children were in the younger age category (below ten years). In the fourth round observed as a child (in 1994), only 9 percent of the sample were above 15 years old. When observed as adult in 2010, the sample average age was 26 years, with 75 percent between 19 and 30. About 54 percent of the children were married or lived with a partner in the same year. Girls tend to marry earlier than boys (61 vs 47 percent).

[2]They reported an attrition rate ranging between 1.5 percent and 17.5 percent, with most of the surveys included spanning a short period.

Most (about 80 percent) of the sample children lived in large family (more than five family members) households during childhood. The average household size in 1991 was eight persons per household, which is higher than the national average (5.3), reported in the Demographic and Health Survey (DHS). A large majority (82 percent) of the children grew up in male-headed homes.

About 40 percent of the sample were below seven years and were not enrolled in school in the first year observed as a child (1991). The average years of schooling among those enrolled in school was 2.6 years. During the year observed as adults (in 2010), average years of education in the sample increased to 7 years, with only 17 percent having above primary level education. Only a small proportion of boys (3 percent) and girls (0.5 percent) received a university education.

Half of the sample stayed in the childhood villages in young adulthood. About 9 percent moved to nearby villages, 24 percent to elsewhere in the Kagera region, 16 percent elsewhere in Tanzania, and less than 2 percent of the children migrated to other countries. The majority of boys (60 percent) stayed in the childhood village compared to girls (37 percent). About 60 percent of the sample were involved in agriculture, 25 percent in paid employment (formal and informal), and 5 percent were still enrolled in school when observed as adults in 2010.

2.1 Poverty measurement and sample characteristics by poverty status

Information on national poverty lines is drawn from Tanzania Household Budget Survey (HBS) conducted in 1991/92, 2001 and 2007. HBS constructed two poverty lines based on each survey: food poverty and the basic needs poverty line. The food poverty line represents the cost of a basket of food consumption necessary to get sufficient calories and was calculated using the food consumption pattern of the poorest 50 percent of the population. The basic needs poverty line, on the other hand, represents the expenditure necessary to meet basic human needs and includes the cost of basic food and non-food consumption. HBS calculated it by adjusting the food poverty line using the non-food share of expenditure for the poorest 25 percent of the population. This study applies the 1991/92 basic needs poverty line to determine children's poverty status across different waves. The fisher price index constructed using KHDS data was used to adjust the poverty line. The resulting poverty lines were compared with household per capita consumption in different waves to determine household poverty status.

In 1991, about 36 percent of the sample children lived in poor families. This is lower than 41 percent for rural areas reported in HBS. While about 55 percent of the sample children experienced poverty, 45 percent were never in poverty during childhood. Male and female poverty rates are similar in childhood (37 vs 35 percent). Poor and non-poor children differ considerably in their families' access to livelihood factors during childhood. Poor children grew up in families with lower financial resources, farmland per capita, assets, and labour force than non-poor children. For instance, mean monthly consumption per capita for poor households was less than half the average for non-poor families (18087 vs 38816 Tanzanian Shilling). Poor children's families reported 0.4 acres less as farmland per capita and nearly one person less as labour force, as did parents of non-poor children. There was also a significant difference in assets measured by the index between poor and non-poor children's families. However, there is little difference between poor and non-poor children's families in social capital measured by household members' participation in non-paid work helping neighbours or relatives. Poor children were also disadvantaged relative to non-poor children with respect to parental human capital and other household characteristics. Parents of poor children had less schooling and a shorter height than parents of non-poor children. Average father and mother schooling for poor children were 5.4 and 5.1 years compared to 5.9 and 5.5 for non-poor children. Poor children were two times more likely to live in female-headed households during childhood relative to non-poor children.

Poor children had poor nutritional status (measured by their HAZ and WAZ scores) in childhood compared to non-poor children. Poor children's height and weight were on average 1.5 standard deviations below the median child of a healthy and well-nourished reference population relative to 1.1 standard deviations below for non-poor children. The difference in years of schooling during childhood did not parallel the difference in nutritional status. However, the slight difference in years of education during childhood grew wide into young adulthood. Average years of schooling for children who grew up in poor households were 4.9 in 2004 and 6.5 in 2010 compared to 5.7 in 2004 and 7.4 in 2010 for children who grew up non-poor. Only 10 percent of poor children attained above primary level education compared to 21 percent of non-poor children. Moreover, children raised in low-income families had shorter height and significantly lower per capita consumption as young adults than children raised in non-poor families.

There was substantial difference in adult outcomes between boys and girls, particularly for those who grew up poor. Among those raised in poor households, boys were better off than girls in terms of height, attaining above primary-level education, and monthly per capita expenditure. On the other hand, among children who grew up non-poor, there were no significant differences in adult outcomes between boys and girls except for height. In addition, a more significant proportion of boys (45 percent) were involved in paid employment and self-employment out of agriculture than girls (23 percent) regardless of childhood poverty status.

In 2010, the proportion of children living in poverty decreased to 17 percent. A higher level of poverty in young adulthood existed among females (20 percent), those working in agriculture (24 percent), and children who stayed in the baseline village (21 percent). Furthermore, the poverty headcount ratio decreases with children's education level. The majority of individuals (95 percent) who attained above primary level education were not poor in young adulthood.

3 Empirical strategy

The study uses the mixed latent Markov chain model to investigate poverty transitions between childhood and young adulthood. In the context of poverty analysis, the

Table 1: Characteristics of Sample Children by Childhood Poverty Status

Variables	Non-poor	Poor	All sample
Childhood family measures			
Expenditure per capita	38816	18087	31370
Farm land per capita	1.075	0.675	0.931
Asset index	0.770	-0.331	0.374
Father's education	5.946	5.401	5.757
Mother's education	5.517	5.132	5.395
Mother's height	158.3	156.0	157.4
Father's height	168.3	167.3	168.0
Labor force	3.735	3.021	3.479
Social capital indicator	0.108	0.0946	0.103
Female headed household	0.125	0.208	0.155
Own childhood measures			
Years of schooling	2.732	2.594	2.686
Height-for -age z score	-1.195	-1.513	-1.309
Weight-for-age z score	-1.184	-1.487	-1.293
Height (centimeters)	120.7	118.0	119.7
Own adulthood measures			
Years of schooling	7.4	6.5	7.1
Height (centimeters)	162.7	161.2	162.1
Expenditure per capita	65569	43302	57549
Attained above primary school	0.206	0.108	0.171
Proportion of married	0.54	0.54	0.54
Proportion of working in paid employment out of agriculture	0.37	0.30	0.34
Boys			
Years of schooling	7.4	6.6	7.1
Height (centimeters)	167.8	165.7	167.0
Expenditure per capita	70894	47441	62028
Attained above primary school	0.22	0.13	0.18
Proportion of married	0.47	0.45	0.47
Proportion of working in paid employment out of agriculture	0.47	0.41	0.45
Girls			
Years of schooling	7.3	6.4	7.1
Height (centimeters)	158.1	156.4	157.5
Expenditure per capita	60443	38641	52988
Attained above primary school	0.20	0.09	0.16
Proportion of married	0.59	0.62	0.61
Proportion of working in paid employment out of agriculture	0.26	0.18	0.23

All childhood variables were measured in 1991-94 and all adulthood variables were measured in 2010.

Markov process starts with a certain probability of occurrence of each poverty category (poor and non-poor) at time ($t = 1$). The mixed latent Markov chain model assumes that poverty status at time ($t+1$) depends not only on the poverty state at the previous period (t) but also on chain (latent class) membership. The model accounts for heterogeneity in the population concerning transition behaviour and measurement error. It is obtained by combining the mixed Markov model that allows for heterogeneity with the measurement model that makes error allowance.

The mixed latent Markov model is appropriate in this study for two reasons. Firstly, children whose per capita consumption is closer to the poverty line are expected to have a different probability of changing state than those with per capita consumption distant above or below the poverty line. Additionally, due to measurement errors at various stages of the data collection process, there is a chance that households, especially those with per capita expenditure near the poverty line, are represented in poverty states different from their actual conditions.

The observed poverty status in the four waves (1991, 1994, 2004 and 2010) is represented by I, J, K and L. The true poverty state for the corresponding waves is not observed and represented by the latent variables A, B, C, and D. The mixed latent Markov model assumes the transition between true poverty states operates at the latent level and follows the Markov chain process. The model links the true poverty states with the observed states using response probabilities, reflecting the likelihood of observing the manifest poverty states for different true poverty states. It further assumes that the population under study comprises S unobserved (latent) subgroups with unique transitions over time.

The latent mixed Markov model expresses the expected frequency F in each cell of the poverty transition table as the function of distribution of the latent variable δ_a, response probabilities, ρ, and latent transitions,τ and sample proportion in latent classes π_s. The structure of the mixed latent Markov model is as follows:

$$F_{ijkl} = N \sum_{s=1}^{S} \sum_{a=1}^{A} \sum_{b=1}^{B} \sum_{c=1}^{C} \sum_{d=1}^{D} \pi_s \delta_{sa} \tau_{sb|a} \tau_{sc|b} \tau_{sd|c} \rho_{si|a} \rho_{sj|b} \rho_{sk|c} \rho_{sl|d} \tag{1}$$

Where subscript ($i = (0,1)$) indexes poverty states in 1991, ($j = (0,1)$) in 1994, ($k = (0,1)$) in 2004, and ($l = (0,1)$) in 2010. Whereas the years 1991, 1994 and 2004 indicate the poverty status of childhood families, 2010 is the poverty status of households children lived as young adults. δ_{sa} refers to the probability of being poor or not at time t (in 1991) conditional on latent class membership, $\tau_{sb|a}$ refers to the probability of transition from poverty category a at time t to category b at time $t+1$, and the rest are defined similarly. The parameter π_s refers to the sample proportions in each S chain.

It is possible to estimate only some versions of the mixed latent Markov model with four waves data. Neither the stationary nor the non-stationary latent mixed model is identified. In this study, we employ the latent mover stayer model that assumes two chains (setting $s = 2$) and constant latent transitions in one of the chains. One chain (the movers) follows a normal Markov process, and the other does not have change over time (the stayers)[3].

The parameter ρ refers to the response probability matrices. The diagonal response probabilities of the ρ matrix are considered reliabilities, and the off-diagonal probabilities as error rates at the various waves. To make the model identifiable with four waves data, we restrict that response probabilities, which show the proportion of misclassified cases, are homogenous across time. The assumption seems reasonable for this study since information on household expenditure, based on which poverty indicator was constructed, was collected mainly in the same way over the various waves[4].

The model specifies constant measurement error for movers whereas perfect measurement for stayers. This assumption seems realistic for the following explanation. Children whose per capita expenditure is continuously above or below the poverty line

[3] That is, transition probabilities are restricted to be $\tau_{j|si} = 1$ for $i = j$ and $\tau_{j|si} = 0$ otherwise

[4] The basic structure of the questionnaire concerning the expenditure components remained largely the same across the various waves.

and hence experience stability have likely placed some distance above or below. Therefore, their poverty status tends to be measured correctly, even if per capita expenditure is measured with error. On the other hand, children switching states are likely closer to the poverty line, at least in some waves. Hence, even a slight measurement error in expenditure probably has a direct bearing on their poverty status.

We compare the latent mover stayer model with various alternative Markov chain models to check if it best suits our data. We estimate four groups of Markov chain models. The first is the simple model that does not account for heterogeneity and measurement error. We estimate mixed models that allow for heterogeneity but make no error allowance in the second group. The third group contains latent models that allow for measurement error but not heterogeneity. Finally, we estimate the latent mover stayer model that accounts for both heterogeneity and measurement error.

Estimation involves defining log-linear models for the different Markov models. Given multinomial sampling distribution, log-linear parameters and expected frequency of the various Markov type models are estimated using maximum likelihood estimation. To evaluate the goodness of fit of the models, we use a likelihood ratio test (L^2). The test assesses how estimated cell frequencies differ from the corresponding observed cell frequencies. Additionally, we apply the dissimilarity index (Δ), which indicates the proportion of observations that should be moved to make the observed and expected transition tables identical.

4 Empirical Results

4.1 Evaluation of goodness of fit of the different Markov models

The various alternative Markov chain models were estimated using LEM software, a program that uses the EM algorithm to obtain Maximum Likelihood estimation of the models' parameters Diebolt and Ip (1995). A set of results, including the expected frequencies, degrees of freedom, likelihood-ratio chi-squared statistic (L^2)), and dissimilarity index (Δ) for the various models, is summarized in Table 2. The first four columns of the table cross classify children based on their poverty status at four measurement points from childhood to adulthood. We also report the observed frequencies for each cell of the poverty transition table in the fifth column for comparison purposes. The expected frequencies were obtained using sample size, initial distribution and transition probabilities estimated under the relevant models. The frequencies add up to the total sample size in all the estimated models.

Each model in Table 2 makes different assumptions about the stochastic process underlying poverty transitions. Model A is a simple Markov model with stationary transition probabilities, while Model B removes the stationarity restriction. Model C is two- chains mixed Markov model with time homogenous transition probabilities in both chains. Model D is the mover-stayer specification with stationary transition probabilities for the movers, and model E relaxes the stationarity restriction. Model F is a latent class model with time homogenous response probabilities. Model G is obtained from model F by relaxing the homogeneity assumption. Model H is a latent Markov model with stationary transition probabilities and time homogenous response probabilities while allowing the latent transition probabilities to vary over time produces model I. Finally, model J is a partially latent mover stayer model with non-stationary transition and homogenous response probabilities. Model A and B do not allow population heterogeneity and measurement error; C, D and E allow population heterogeneity but not measurement error; F, G, H, and I allow measurement error but not heterogeneity; model J allows both.

The analysis begins with assessing the overall fit of the various models. We start with the simplest model and proceed successively to the more plausible models by relaxing one or more restrictions to see if the fit is improved. As Table 2 shows, the stationary Markov model (A) has an (L^2) value of 117 on 12 degrees of freedom and fits the data poorly, with 11 percent of the cases wrongly classified. Removing the stationarity restriction (model B) improves fit, with only 6 percent of cases misclassified under the non-stationary model. The LR difference between the two models (=74 on 4 degrees of freedom) is highly significant, suggesting that allowing transition probabilities to vary over time is essential to represent the data. Even if it misclassifies fewer cases than the stationary model, the non-stationary model is still unsatisfactory (LR=43 with 8 degrees of freedom). Both simple models (A and B) provide a poor approximation to the poverty transition table, indicating that at least one of the assumptions underlying the simple Markov model is incorrect.

As a first step, we try to improve the model fit by allowing heterogeneity in the population. We have various alternative specifications of the mixed Markov model. A stationary mixed Markov model (C) with two chains, each experiencing a distinct poverty trajectory, does not provide adequate fit ($L^2 = 82$) on 8 degrees of freedom). The LR difference between the stationary mixed and simple Markov model is significant (LR=35 on 4 degrees of freedom), signifying that the former model provides a better fit. We also find fewer (around 8 percent) wrongly classified cases with stationary mixed Markov model in line with the test. This result seems to favour population heterogeneity over homogeneity concerning poverty trajectory. Since the non-stationary simple Markov model is not nested within the stationary mixed Markov model, direct comparison using the LR difference test is not applicable. However, comparison based on BIC favours the non-stationary simple Markov model.

We next consider models that make a specific assumption about the nature of heterogeneity prevailing in the population. As can be seen from Table 2, about 41 percent of the children were in the same poverty state throughout the observation period. Additionally, we have a more considerable difference between the observed and expected frequencies under the simple Markov model for cells representing children who remained in the same poverty state throughout the observation period and cells designating children who changed states only once. This feature suggests that allowing the existence of stayers would be more appropriate. Accordingly, we consider the mover-stayer model that assumes no transition for children in one of the chains. The stationary mover-stayer model (D) that restricts the movers' transition probability to be stationary does not provide adequate fit (with ($L^2 = 90$) on 10 degrees of freedom). Of course, the misclassified cases decrease to 8 percent from 11 percent for the stationary simple Markov model. This improvement equals the fit obtained with the stationary mixed Markov model.

Allowing transition probability for movers to vary over time (E) significantly improves the fit compared to both the simple Markov and the stationary mixed Markov models and provides an acceptable fit to the data. There is no difference at the 0.05 significance level between the observed and the estimated frequencies under the non-stationary mover stayer model. The mismatch between the observed and estimated frequencies accounts for only 2.5 percent of the cases. This shows, as opposed to the simple Markov model assumption, children seem to follow two distinct poverty trajectories: one experiencing perfect stability and the other segment is characterised by changing poverty states.

Before considering models with measurement error, we estimate a latent class model assuming the same latent variable underline the observed poverty transition at each wave. When fitted to the data, the model tests the hypothesis that all observed changes are errors. The model proposes perfect stability and transition from a poor childhood to non-poor adulthood, non-poor childhood to poor adulthood due to measurement error, which is very restrictive. We consider first a latent class model with time homogenous reliabilities (F). Not surprisingly, the fit of the stationary latent class model appears to be very bad. With around 13 percent of the cases wrongly classified, the model yields the most significant mismatch.

Table 2: Summary of Estimated Frequencies and Goodness of Fit Test Statistics for Different Markov Type Models

Poverty status at time				Observed frequency	Estimated frequency for model									
t=1	t=2	t=3	t=4		A	B	C	D	E	F	G	H	I	J
0	0	0	0	557	518.1	517.0	557.0	557.0	557.0	556.2	554.4	546.8	553.5	557.0
0	0	0	1	62	106.5	74.4	79.3	87.3	57.3	98.5	66.4	77.9	64.6	61.1
0	0	1	0	113	74.8	119.9	70.3	67.0	106.9	98.5	117.5	74.5	101.3	104.5
0	0	1	1	31	53.6	51.5	42.0	47.1	41.6	49.6	20.3	27.2	43.8	40.4
0	1	0	0	64	74.8	72.0	70.3	67.0	70.0	98.5	62.4	95.1	69.3	69.8
0	1	0	1	10	15.3	10.3	29.3	20.6	14.7	49.6	22.1	22.6	14.5	14.9
0	1	1	0	47	37.6	43.7	37.3	36.2	43.3	49.6	47.2	36.0	45.3	43.3
0	1	1	1	24	27.0	18.8	22.3	25.4	16.9	39.8	17.3	27.9	19.6	16.7
1	0	0	0	152	213.9	182.2	182.8	181.9	153.4	98.5	147.5	192.9	157.1	153.4
1	0	0	1	29	43.9	26.2	50.3	56.0	32.2	49.6	34.1	34.2	24.1	28.4
1	0	1	0	61	30.8	42.2	44.6	43.0	59.9	49.6	69.3	43.4	58.3	62.4
1	0	1	1	27	22.1	18.1	26.6	30.2	23.4	39.8	22.2	26.9	25.2	24.1
1	1	0	0	130	107.7	131.6	99.8	98.2	122.5	49.6	106.5	96.0	123.0	122.4
1	1	0	1	29	22.1	18.9	26.6	30.2	25.7	39.8	39.2	31.4	25.9	26.2
1	1	1	0	65	54.2	79.9	54.5	53.0	75.7	39.8	83.9	58.2	80.8	75.9
1	1	1	1	41	38.8	34.4	48.3	41.0	41.0	34.5	31.0	50.2	35.0	41.0
Likelihood ratio chi Square					117.1	43.1	82.4	90.4	12.2	243.7	34.3	75.3	14.9	11.2
Degrees of freedom					12	8	8	10	6	12	6	10	6	4
Dissimilarity index					0.1133	0.0661	0.0788	0.0820	0.0256	0.1332	0.0445	0.0843	0.0321	0.0236

A, stationary Markov chain, B, non-stationary Markov chain, C stationary mixed Markov chain, D, stationary mover stayer,E, non -stationary mover stayer F-stationary latent class model G non -stationary latent class model H stationary and time homogenous latent Markov chain, I, non-stationary and time homogenous latent Markov chain J, latent mover stayer (non-stationary and time homogenous).

On the other hand, allowing the reliabilities to vary between measurement points (G) brings considerable improvement over the simple Markov model. The non-stationary latent class model misclassifies only 4 percent of the cases, 32 percent improvement in the goodness of fit over the non-stationary simple Markov model. This result suggests a considerable part of the observed mobility could be attributed to measurement error.

Unlike the latent class model, the latent Markov model allows true change and measurement error. As discussed earlier, the response probabilities have to be homogenous across time to identify the model with four measurement points. We first assume time homogenous transition probabilities. The stationary latent Markov model (H) provides a better fit to the data (with 8 percent misclassified cases) compared to the stationary simple Markov model (11 percent) and stationary latent class model (13 percent). Allowing the latent transition probabilities to vary over time (I) fits the data better. With an (L^2) value of 14 on 6 degrees of freedom, there are no significant differences at 0.01 level between the estimated and observed frequencies. The model misclassifies only 3 percent of the cases.

Thus far, we find that accounting for either heterogeneity or measurement error improves the model fit. We next consider the partially latent mover stayer model (J) that simultaneously allows for heterogeneity and measurement error. The model fits the data reasonably well (at the 0.05 significance level) with an (L^2) value of 11 on 4 degrees of freedom. The model also produces the least misclassified cases (2 percent). However, the model is only as good a fit as the mover stayer and the latent Markov models. It is found that the improvement in the goodness of fit as measured by LR difference over the mover stayer (equivalent to LR= 1 on 2 degrees of freedom) and the latent Markov models (equivalent to LR= 3 on 2 degrees of freedom) is not statistically significant at 5 percent level.

In summary, the simple Markov model, as expected, provides a poor approximation to the poverty transition table. The stationary simple model fits the data poorly; the non-stationary model fits significantly better but is inadequate. On the other hand, the partially latent mover stayer, non-stationary mover stayer and latent Markov models provide an equally good fit. We choose the latent Markov and the latent mover stayer models for the remaining analysis because they better reflect the poverty transition process in the data.

4.2 Estimate of poverty dynamics

The previous section considered the goodness of fit of various Markov models and identified the partially latent mover stayer and the non-stationary latent Markov models fit well our data. This section presents the estimated parameters of the two models. Before looking at the extent of intergenerational transmission of poverty among children, it is essential to study the overall poverty dynamics during the observation period. We analyse the dynamics by classifying the observation period into three: from 1991 to 1994, from 1994 to 2004, and from 2004 to 2010. Transition probabilities in both the latent Markov and partially latent mover stayer models are corrected for measurement error. As a first step, we describe the poverty dynamics using the non-stationary latent Markov model (Table 3), which is later contrasted with the result implied by the partially latent mover stayer model (Table 4).

The result based on the latent Markov model shows that after correcting for measurement error, the majority (about 57 percent) of the children were raised in poor childhood families relative to only 37 percent observed in the data. Between 1991 and 1994, there was almost no risk of falling into poverty; all non-poor children remained non-poor. Besides, there was a relatively low probability of escaping poverty. For children who were poor in 1991, 78 percent remained poor in 1994. From 1994 to 2004, both the risk of falling into (0.25) and moving out of poverty (0.32) increased compared to 1991-94. In 2004 – 2010, the risk of slipping into poverty decreased (0.12) while moving out of poverty sharply increased (0.53) compared to 1994-2004. Overall, the risk of falling into poverty remained relatively low. In contrast, the probability of escaping out of poverty increased between childhood and adulthood. As a result, the poverty rate based on the latent poverty distribution in 2010 decreased to 32 percent from 57 percent in 1991.

Turning to the latent mover stayer model, the majority (86 percent) of the children belong to the movers' chain, representing those moving between poverty states. Stayers account for only 14 percent of the sample, which is significantly lower than 20 percent estimated using the mover stayer specification without accounting for measurement error. Among the stayers, the majority (94) of the children were non-poor during childhood. On the other hand, among the movers, the majority (65 percent) of the children were poor during childhood. Given the poverty distribution during childhood and the proportion of stayers, the model indicates that about 13 percent of the children were never observed in poverty while about 1 percent were always in poverty.

Based on the latent mover stayer model, the movers' poverty dynamics are mainly similar to the pattern obtained for the whole sample using the latent Markov model. During 1991-94, there was very low poverty mobility concerning slipping into poverty (0.04). While transition out of poverty (0.26) was higher than obtained using the whole sample, most children remained in the same poverty state during the period. From 1994 to 2004 there was relatively higher mobility compared to 1991-94. The risk of moving into and out of poverty significantly increased to 0.35 and 0.37, respectively. Between 2004 and 2010, there was high mobility out of poverty (0.57), whereas the probability of slipping into poverty significantly decreased (0.18). When taken together, the result shows that there was significant movement out of poverty while falling into poverty remained relatively low between 1991 and 2010. As a result, in 2010, 29 percent of the movers were poor compared to 65 percent in 1991.

Moreover, in most of the observation period, the risk of poverty for poor children in the previous period is significantly higher than for non-poor. The observed trend partly reflects improving economic conditions at the macro level. There has been sustained and robust economic growth in Tanzania since 2000, which lifted many people out of poverty.

Table 3: Summary of Estimated Parameter Values for Stationary and Non-Stationary Latent Markov Models

Model	Initial Probabilities		Latent transition Probabilities								Response Probabilities	
	t=1(1991)		t to t+1 (from 1991 to 1994)		t+1 to t+2 (from 1994 to 2004)		t+2 to t+3 (from 2004 to 2010)		t to t+3 (from 1991 to 2010)			
	Class		Poor	Non-poor	Poor	Non-poor	Poor	Non-poor	Poor	Non-poor	Poor	Non-poor
Stationary Latent Markov	Poor	0.4659	0.6687	0.3313	0.6687	0.3313	0.6687	0.3313	0.5804	0.4196	0.6906	0.3094
		(0.0371)	(0.0368)	(0.0368)	(0.0368)	(0.0368)	(0.0368)	(0.0368)	(0.0691)	(0.0691)	(0.0409)	(0.0409)
	Non-poor	0.5341	0.0382	0.9618	0.0382	0.9618	0.0382	0.9618	0.1749	0.8251	0.0905	0.9095
		(0.0371)	(0.0216)	(0.0216)	(0.0216)	(0.0216)	(0.0216)	(0.0216)	(0.0170)	(0.0170)	(0.0152)	(0.0152)
Non-Stationary Latent Markov	Poor	0.5731	0.7806	0.2194	0.6825	0.3175	0.4714	0.5286	0.3785	0.6215	0.6406	0.3594
		(0.0197)	(0.0314)	(0.0314)	(0.0383)	(0.0383)	(0.0370)	(0.0370)	(0.0377)	(0.0377)	(0.0225)	(0.0225)
	Non-poor	0.4269	0.0000	1.0000	0.2489	0.7511	0.1157	0.8843	0.1316	0.8684	0.0005	0.9995
		(0.0197)	(0.0000)	(0.0000)	(0.0252)	(0.0252)	(0.0220)	(0.0220)	(0.0415)	(0.0415)	(0.0001)	(0.0001)

Table 4: Summary of Estimated Parameter Values for Partially Latent Mover Stayer Model (Non-Stationary)

Chain proportion		Initial Probabilities		Transition Probabilities										Response Probabilities	
		t=1(=1991)		t to t+1 (from 1991 to 1994)		t+1 to t+2(from 1994 to 2004)		t+2 to t+3(from 2004 to 2010)		t to t+3 (from 1991 to 2010)		t+1 to t+3 (from 1994 to 2010)			
Chain		Class		Poor	Non-poor	Poor	Non-poor	Poor	Non-poor	Poor	Non-poor	Poor	Non-poor	Poor	Non-poor
Stayers	0.144	Poor	0.056	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
		Non poor	0.944	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
Movers	0.855	Poor	0.645	0.741	0.258	0.632	0.367	0.425	0.574	0.215	0.784	0.280	0.720	0.655	0.344
		Non poor	0.354	0.039	0.960	0.352	0.647	0.176	0.823	0.233	0.766	0.259	0.740	0.000	1.000

4.3 Implications for the intergenerational transmission of poverty

Considering the transition from 1991 to 2010 based on the latent Markov model, parental economic status during childhood strongly predicts children's economic situation in adulthood. We find that children from poor families have poverty risk (0.38) in young adulthood, three times higher than those from non-poor family backgrounds (0.13). About 67 percent of children raised in poor households remained poor in young adulthood. Conversely, children from non-poor family backgrounds were much more likely to be non-poor in young adulthood. The majority of the children who were non-poor during childhood (87 percent) remained non-poor in young adulthood. The result suggests that as much as childhood poverty significantly harms, affluence helps children's adult economic status. Besides, there is a higher intergenerational transmission rate among children from non-poor families than those from low-income families.

The latent mover stayer model depicts a poverty trajectory for the movers somewhat different from that obtained for the whole sample using the latent Markov model. Similar to the earlier finding, there is a low probability of slipping into poverty during adulthood for children from non-poor family backgrounds. Of children raised in non-poor families, 77 percent remained non-poor during adulthood. However, the latent mover model estimates considerable mobility out of poverty for children in the mover chain compared to the finding for the whole sample. Of those children raised in low-income families, 78 percent escaped poverty. This trend remains the same when we switch the base year from 1991 to 1994. The transition from 1994 to 2010 also shows change over time towards moving out of poverty and staying non-poor. Poor children in the stayer group were chronically poor while impoverished children in the mover chain moved between poverty states during childhood. The result suggests children from transitorily poor family have a lower risk in adulthood than those who sustained poverty during childhood.

Another interesting difference is that in contrast to the result obtained for the whole sample, the risk of poverty in adulthood is not significantly different between children from poor and non-poor childhood families in the movers' chain. The result reflects a small difference in per capita consumption in childhood families between the two groups in the movers' chain. The difference between the two groups in the movers' chain is significantly smaller than the stayers' chain. Therefore, we may infer that the tendency to inherit parental economic status is high when parental family per capita consumption is far below and far above the poverty line.

4.4 Implications of Measurement Error

This section investigates measurement error, error-corrected stability and changes implied by the two preferred Markov models. The last columns of Table 3 and Table 4 report estimated response probabilities. As already mentioned earlier, the response probability is interpreted as the reliability of the poverty indicator. The response probabilities on the diagonal of the matrix, where both the observed and the latent poverty states are the same, measure reliability, and the off-diagonal probabilities show the magnitude of measurement error. Based on the latent Markov model, the level of reliability for the non-poor state is almost 1, indicating that nearly all truly non-poor are observed to be non-poor. On the contrary, the poor state's reliability is 0.64, implying that about 36 percent of poor children at the latent are observed to be non-poor.

The latent mover stayer model's result shows that the response probabilities for stayers are equal to the identity matrix since we assumed that stayers are perfectly measured. The reliability matrix for the movers confirms the pattern observed based on the latent Markov model. The non-poor state is perfectly measured among the movers, while the poor state has the same magnitude of error rate (0.34). Finding the same magnitude of reliability and error rate even after blaming only the movers for measurement error corroborates the assumption of perfect measurement for stayers. Based on the result, a measurement error occurs mainly because of difficulty identifying the poor.

Using the response probabilities, we divide change and stability implied by the two

Table 5: Estimated Proportion of Change and Stability

	Simple Markov	Latent Markov	Latent Mover Stayer
Perfect stability			0.144
Stability (TOS)	0.386	0.427	0.264
True stability (TRS)		0.307	0.175
Error (EPS)		0.120	0.090
Change (TOC)	0.614	0.573	0.592
True change (TRC)		0.230	0.249
Error (EPC)		0.342	0.342
Total error		0.463	0.432
OBS	0.415		
OBC	0.585		

models into different quantities of interest. Table 5 provides estimated proportions of change and stability for the two models together with the observed proportion of change (OBC) and stability (OBS). The observed proportion of change and stability were computed based on the observed patterns in the data. TOS, the total proportion of stability, is defined as the proportion of children who remained in the initial poverty state during the observation period. For the latent mover stayer model, the proportion of children in the stayers' chain is considered perfectly stable. The total proportion of change (TOC) and TOS in principle should add up to 1 means TOC amounts to $1-TOS-perfectstability$. Total change and stability may be further divided into true and error components. True stability is TOS corrected for measurement error, and the error proportion of stability is equal to $EPS = TOS - TRS$. Similarly, subtracting the true proportion of change (TRC) from the total proportion of change (TOC) yields the error proportion of change (EPC).

In the latent Markov model, TOS and TOC are 0.43 and 0.57. Similarly, stability (total stability plus perfect stability) is 41 percent, and change is 59 percent based on the latent mover stayer model. By comparing these figures with the OBS (0.42) and OBC (0.59), one may consider that the effect of measurement error on the percentage of total change and stability is minimal. However, when looking closer at the results, we can see that the proportion of true stability (TRS plus perfect stability=0.32) and true change ($TRC = 0.25$), which is observed, is much smaller than OBS and OBC. This means that considerable error in poverty measurement appears as both a change and stability. It is also found that EPC is significantly higher than the EPS in both the latent Markov and partially latent mover stayer models, indicating that error appears mainly as a change. This finding partly has to do with the large proportion of children in the mover class and partly with the high reliability of measuring non-poor, which constitutes most stayers.

As the non-poor state is found perfectly measured (with reliability almost equal to 1), misclassifying a poor respondent as non-poor appears to be the source of error for both stability and change. For stability, this implies that the observed proportion of children who remained non-poor is overestimated while the observed proportion of children who remained poor is underestimated. Likewise, by comparing the latent transition with the observed transition, we can see that the latent probability of remaining in poverty is underestimated. Thus, the observed data underestimates intergenerational transmission of poverty while overestimating intergenerational transmission of economic status among non-poor children. On the other hand, the implication for change is that escaping poverty will be overestimated if we ignore measurement error.

4.5 Robustness checks

In the previous section, we investigated the implications of measurement error. This section examines the robustness of our main results to potential sample errors and various changes in specification. First, we check the robustness to sample selection due to non-response. The main estimation was made using children with complete information on poverty status for the four waves. Excluding observations with missing data on poverty status may lead to biased parameter estimates if they are not missing at random. To assess the role of missing data on poverty status in influencing our result, we specify and test different models for missing data mechanisms. Data on poverty status may be missing completely at random (MCAR), missing at random (MAR), or not missing at random (NMAR). The first case occurs if the probability of missing data is independent of poverty status both in the observed and missing waves. Data on children's poverty status is considered MAR if the likelihood of missing data may depend on the observed waves' poverty status, but not on the missing waves. Suppose the probability of missing data depends on poverty status in one or more of the missing waves. In that case, we say that the data is not missing at random (NMAR). The missing data mechanism is ignorable only if data is either MCAR or MAR.

The missing data pattern in the dataset is reported in Table A1 in the Appendix. Among the sample children, information on poverty status is mainly missing for 1994 and 2004. Because only a few cases are missing information on poverty status for 1991 and 2010, we ignore the last two missing data patterns and consider poverty status in 1991 and 2010 as fully observed. The test results for the different models in Table A2 in the Appendix show that the missing data mechanism is not MCAR. The saturated MAR and the third NMAR model, where the missing pattern depends on poverty status in all missing waves, fit the data better.

We further examine whether the latent mover stayer model's parameters are sensitive to MAR and NMAR specifications. Table A3 and A4 in the Appendix respectively report parameter estimates for ignorable saturated MAR and non-ignorable NMAR response mechanisms. In both cases, parameter estimates of the latent mover stayer model using incomplete data are the same as those obtained using complete data. Only the proportion of poor in childhood, both in the stayer and mover groups, is slightly increased compared to the estimates obtained using complete data. Thus excluding observations with missing information on poverty status does not significantly affect the estimation.

Second, we check whether the sample selection due to panel attrition confounds our result. When we selected the sample for the study, we excluded children who were not observed in 2010. These children dropped out of the panel because they could not be traced or due to death. This selection can also bias parameter estimates if dropping children vary systematically from those staying in the panel. The various non-response mechanisms were specified and tested using incomplete data for all children initially interviewed in the baseline survey satisfying the sample selection criteria. As shown from the test result in Table A2 in the Appendix, both MCAR and MAR models performed poorly. It appears that the missing data mechanism is NMAR and hence not ignorable. Nevertheless, the estimates obtained using incomplete data are not significantly different from those obtained using complete data. The only noticeable difference with the estimates obtained using complete data is that the proportion of stayers slightly increased, and the probability of remaining in poverty for movers slightly decreased. Therefore, even if the missing mechanism due to attrition is non-random and non-ignorable, it does not significantly affect the parameter estimates.

Third, we test the sensitivity of our results to the choice of the poverty line. Although there is no significant difference in wellbeing between individuals slightly below and slightly above the poverty line, they are placed in two different poverty states due to the choice of threshold. To check the possible effect of changing the threshold, we alter the poverty line by 5 percent. A 5 percent increase in the national poverty line generates an average 3.7 percent increase in the poverty rate while the same decrease generates 3.3 percent decrease in the poverty rate in the different waves. Similarly, a 10 percent change in the threshold also generates less than proportionate change in poverty rate. The result shows that the poverty rate is only slightly sensitive to the choice of the poverty line. Additionally, the change does not affect the overall mobility pattern significantly.

Apart from changing the national poverty threshold, we also assess how sensitive the result is to setting alternative poverty lines. Instead of the national poverty line used in the primary analysis, we measure poverty using international (higher) and relative (lower) poverty lines. We find that measuring poverty using the $1.25 per day international poverty line generates a significantly higher proportion of poor. About 96 percent of children are classified as poor in childhood by this standard, which decreased to 63 percent in adulthood. Similar to the main result, the overall poverty dynamics using international poverty line also shows significant movement out of poverty and a low risk of falling into poverty between childhood and adulthood.

On the other hand, using a 60 percent share of mean per capita consumption to measure poverty generates a significantly lower poverty rate (18 percent in childhood and 44 percent in adulthood). It also tells a different story in terms of mobility compared to the primary analysis. As opposed to absolute poverty, relative poverty has increased over time between childhood and adulthood. In sum, the above analysis shows that except to some extent being sensitive to the choice of the poverty line, our main result is reasonably robust to various potential sample errors and changes in specification.

4.6 Mechanisms of the Intergenerational Transmission of Poverty

We found that growing up poor increases the likelihood of poverty in adulthood from the first part. The study's second aim is to provide evidence for one of the many ITP explanations. This section explores the role of parental resources in ITP and mechanisms linking childhood parental resources and poverty status in adulthood. We follow a two-stage approach to do this. First, we estimate the logit specification of the equation in 2 to establish the role of parental resources in ITP. A range of household and extra household factors influence the intergenerational transmission of poverty. Estimating the causal effect of these factors on children's poverty status in adulthood is daunting. We instead follow DFID's sustainable analysis livelihood framework formulation, based on Sharp (2003) research in Ethiopia, to provide suggestive evidence similar to Wu et al. (2019). We follow this approach because parents transfer poverty to children through various factors that influence children's chances of being poor in adulthood. The livelihood framework considers poverty status (Y) as a function of livelihood factors measuring financial (F), human (H), natural (N), physical (P) and social capital (S). We focus on the relationship between poverty status in young adulthood and livelihood factors in the parental family during childhood. Parents' access to and use of the different capitals affects children's wellbeing through the life course. It influences their socioeconomic outcomes in adulthood by shaping their capability and earning potential.

$$Y_i = Y(F_i, H_i, N_i, P_i, S_i) \tag{2}$$

We include monthly per capita household consumption to measure financial capital and household labour force and parental education to measure household human capital. Per capita farmland measures the natural capital, and assets index measures the physical wealth of the parental family. Finally, indicators for participation in social activities measure social capital. To capture the long-term availability of livelihood factors, we use the average of the variables over all the years a respondent was observed in childhood. Table A6 in the appendix provides a more detailed description of the variables.

Table 6 presents results from logistic regression of equation 2. The first column shows a baseline association between poverty status in adulthood and parental financial resources during childhood. In the second column, we control other livelihood factors in parental families affecting poverty status. The third column provides the odd ratio of the covariates. In all specifications, average monthly per capita consumption has a statistically negative relationship with children's poverty status in young adulthood. In our preferred specification (column 2), for a three-fold increase in monthly per capita consumption in parental family, the odds of poverty in adulthood decreases by 65 percent. All the other livelihood factors, except the indicator for social capital, also have the expected negative sign. However, only the coefficients on asset index, father schooling and mother schooling are statistically significant. Therefore, the result shows that parental financial resources and human capital are strongly associated with a child's poverty risk in adulthood.

Next, we focus on the relationship after controlling individual covariates in young adulthood to see whether the strong association between parental financial resources and poverty status in adulthood found above still survives after controlling individual

Table 6: Childhood Parental Resources and Poverty Risk in Adulthood

	1	2	3	4
Ln(Expenditure per capita in parental household)	-1.364	-1.037	0.355	-0.872
	(0.141)***	(0.235)***	(0.083)***	(0.292)***
Farm land percapita in parental family		-0.227	0.797	-0.280
		(0.167)	(0.133)	(0.184)
Asset index in parental family		-0.143	0.867	-0.115
		(0.070)**	(0.061)**	(0.078)
Labor force in parental family		-0.020	0.980	-0.106
		(0.074)	(0.072)	(0.086)
Father's education		-0.244	0.783	-0.175
		(0.054)***	(0.042)***	(0.058)***
Mother's education		-0.100	0.904	-0.109
		(0.051)**	(0.046)**	(0.060)*
Social capital indicator in parental family		0.179	1.196	0.379
		(0.340)	(0.407)	(0.374)
Years of schooling in adulthood				-0.243
				(0.046)***
Ln(Height (centimeters) in adulthood)				0.137
				(2.480)
Ln(Age in adulthood)				0.244
				(0.551)
Female(indicator)				0.709
				(0.263)***

* p¡0.1; ** p¡0.05; *** p¡0.01 Regression in column (1) is a baseline specification. Regression in column (2) controls livelihood factors in parental families. Regression in column (3) provides the odd ratio of the covariates. Regression in column (4) controls individual covariates in young adulthood.

characteristics. The fourth column reports results after controlling individuals' years of education, height in adulthood, age and sex in addition to the livelihood factors. The result shows that parental resource remains strongly associated with an individual's poverty status in adulthood. The notable difference compared to the previous estimation is that the coefficient on parental resources is significantly reduced. This finding suggests that the individual covariates such as years of schooling might mediate some of the effects of parental resources on poverty status. Another result worth noting is that the indicator for females has a statistically positive association with poverty status, indicating that girls are more likely to be poor in early adulthood than boys.

To explore the mediating effect of child human capital, we estimate the relationship between parental resources in childhood and child human capital outcomes in young adulthood. The reduced-form equation in 3 follows the equation specified in many papers estimating human capital outcomes. Child human capital (H) is dependent on parental (P), individual (I) and unobserved community level (ε_c) and individual (ε_i) characteristics. Parental characteristics included in the regression are household per capita consumption, father's and mother's years of schooling, maternal age, and maternal height. Household per capita consumption measures parental resource availability during childhood. The mother's age is included in the regression to account for the mother's experience and knowledge, which is vital in household decisions such as children schooling and health outcomes.

$$H_i = \beta_0 + \beta \times P + \gamma \times I + \varepsilon_c + \varepsilon_i \tag{3}$$

We control a child's sex and age at the individual level to capture age and gender-specific differences in human capital accumulation. Village-specific fixed-effects capture community infrastructures, such as school and health facility availability, access to safe water, and electricity access in the community. It also controls other community-level observable and unobservable characteristics affecting child human capital outcomes. We examine three indicators of individual human capital. We use height-for-age z score

Table 7: Parental Financial Resource and Human Capital in Adulthood

	Years of schooling	HAZ	Years of schooling	Height
Ln(Expenditure per capita in parental household)	0.657	0.334	0.843	1.945
	(0.227)***	(0.158)**	(0.475)*	(0.986)*
Father's education	0.022	-0.008	0.090	-0.323
	(0.052)	(0.033)	(0.081)	(0.181)*
Mother's education	-0.092	0.043	0.087	0.048
	(0.082)	(0.048)	(0.089)	(0.264)
Mother's height	0.020	0.034	0.033	0.354
	(0.019)	(0.014)**	(0.026)	(0.075)***
Father's height	0.017	0.037	0.013	0.285
	(0.011)	(0.009)***	(0.019)	(0.052)***
Ln(Mother's age)	0.036	0.380	1.139	-0.223
	(0.502)	(0.323)	(0.612)*	(1.921)

* p¡0.1; ** p¡0.05; *** p¡0.01 Coefficients of individual variables and cluster fixed effects are not reported in the table.

(HAZ) in childhood and height in young adulthood to measure the long-run health outcome and years of schooling completed as a measure of educational outcome.

Table 7 presents OLS estimation results of the effect of parental financial resources on children's human capital. The first two columns show results of childhood outcomes, while the following two columns present that of adulthood outcomes. In all regressions, household per capita consumption in the childhood family is strongly associated with the child's human capital. In childhood, higher parental financial resources, measured by per capita household consumption, increase years of schooling and improve children's HAZ score. Similarly, children's years of education and height in young adulthood increase with per capita household consumption in the childhood family. A 10 percent increase in per capita consumption in childhood families is associated with 0.8 years increase in schooling, 0.3 improvements in HAZ score and a 2 cm increase in height. The result suggests that low parental resources increase the risk of poverty in adulthood by reducing human capital investment in children.

Moreover, the result also indicates that parental human capital plays a vital role in children's human capital accumulation. In particular, we find that parental height is positively related to children's HAZ score in childhood and height in adulthood. The result could be due to genetic links or due to the role of parental resources since parental human capital, and parental resources are significantly correlated. In the second case, this finding also strengths the strong association between parental resources and child human capital we find above. Therefore, the result suggests that child human capital is the primary mechanism linking parental financial resources in childhood and poverty status in adulthood.

5 Conclusion

This study explored the extent of intergenerational poverty as well as one of the mechanisms of poverty transmission in Tanzania. We used long-running panel data covering repeated observations of parents and children from childhood to adulthood. The first part of the analysis used the latent Markov model and the latent mover stayer model to investigate the poverty transition between childhood and adulthood. The first model corrects for measurement error, while the second one accounts for heterogeneity in the population and measurement error in the data. After correcting for measurement error, about 57 percent of the sample children were raised in poor childhood families. We found that the risk of falling into poverty remained low, while the probability of escaping out of poverty increased between childhood and adulthood; as a result, the poverty rate decreased to 32 percent in adulthood. The analysis also revealed that children from poor families have poverty risk (0.38) in adulthood, three times higher than those from non-poor family backgrounds (0.13).

A further investigation of the poverty risk in adulthood after accounting heterogeneity in poverty experience provided us with additional insights not apparent from analyzing the whole sample. The results indicated that poverty risk in adulthood in the movers' chain is significantly lower than that in the stayers chain, suggesting that transitorily poor have a lower risk in adulthood than those who sustained poverty during childhood. We also found that the likelihood of poverty in adulthood in the movers' chain is not significantly different for children from poor and non-poor childhood families. The result reflects a small difference in per capita consumption in childhood families between the two groups in the movers' chain. When taken together, the findings implied that parental economic status during childhood strongly predicts children's economic situation in adulthood when per capita consumption is far below or above the poverty line. This agrees with the intergenerational mobility literature finding that mobility is lower at the bottom and top of the income distribution than in the middle.

Additional analysis on the implications of measurement error revealed that the observed data underestimates ITP and overestimates intergenerational transmission of economic status among the non-poor. We also found that difficulty identifying the poor is the primary source of measurement error in the data. The non-poor state was found to be accurately measured. Moreover, our analysis demonstrated that the main results are reasonably robust to potential sampling errors and various specification changes. In our attempt to provide evidence on the role of parental resources in ITP, we found that parental financial resource in childhood is strongly associated with an individual's poverty status in early adulthood. Our result indicated that human capital investment in children mediates some of the effects of childhood parental resources on economic status in adulthood. We found a strong association between parental financial resources in childhood and children's human capital in young adulthood. In particular, we found that a child's years of education and height in young adulthood increase with per capita household consumption in the childhood family. Based on the result, we suggest that interventions supporting low-income families to build their children's human capital are essential to break the intergenerational cycle of poverty.

The study has produced quantitative evidence on ITP in Tanzania that may also help understand the nature of ITP in other parts of Africa. However, several caveats need to be considered while interpreting the study results. First, our analyses used one-year data to measure outcomes in adulthood, unlike childhood outcomes measured using multiple waves data. A one-year data may not be sufficient to accurately measure economic conditions in adulthood. Second, the study did not include all variables related to the different ITP explanations when establishing the role of parental resources due to a lack of data. Third, the study identified mechanisms of ITP based on association because the data did not permit establishing a causal link. Nevertheless, causal and non-causal mechanisms underpinning ITP may have different implications for policy design. Therefore, addressing these limitations may be a focus area for future ITP studies on Tanzania.

References

Ilpo Airio, Pasi Moisio, and Mikko Niemelä. Intergenerational transmission of poverty in finland in the 1990s. *European Journal of social security*, 7(3):253–269, 2005.

Harold Alderman, Jere R Behrman, Hans-Peter Kohler, John A Maluccio, and Susan Cotts Watkins. Attrition in longitudinal household survey data: some tests for three developing-country samples. *Demographic research*, 5:79–124, 2001.

Bob Baulch. *Why poverty persists: Poverty dynamics in Asia and Africa*. Edward Elgar Publishing, 2011.

Bob Baulch and John Hoddinott. Economic mobility and poverty dynamics in developing countries. *The Journal of Development Studies*, 36(6):1–24, 2000.

Kathleen Beegle, Joachim De Weerdt, and Stefan Dercon. Kagera health and development survey 2004 basic information document. *The World Bank. www. worldbank. com/lsms/country/kagera2/docs/KHDS2004% 20BID% 20feb06. pdf[accessed March 13, 2007]*, 2006.

Jere R Behrman, Whitney Schott, Subha Mani, Benjamin T Crookston, Kirk Dearden, Le Thuc Duc, Lia CH Fernald, and Aryeh D Stein. Intergenerational transmission of poverty and inequality: parental resources and schooling attainment and children's human capital in ethiopia, india, peru, and vietnam. *Economic development and cultural change*, 65(4):657–697, 2017.

K Bird. The intergenerational transmission of poverty: an overview. chronic poverty research centre working paper no. 99, 2007.

Kate Bird. The intergenerational transmission of poverty: An overview. *Chronic poverty*, pages 60–84, 2013.

Elizabeth Cooper. Inheritance and the intergenerational transmission of poverty in sub-saharan africa: policy considerations. *Chronic Poverty Research Centre Working Paper*, (159), 2010.

Miles Corak. *Do poor children become poor adults? Lessons from a cross-country comparison of generational earnings mobility*. Emerald Group Publishing Limited, 2006.

Mary Corcoran. Mobility, persistence, and the consequences of poverty for children: Child and adult outcomes. *Understanding poverty*, pages 127–161, 2001.

Mary Corcoran, Terry Adams, et al. Race, sex, and the intergenerational transmission of poverty. *Consequences of growing up poor*, pages 461–517, 1997.

J. De Weerdt, K. Beegle, HB. Lilleør, S. Dercon, K. Hirvonen, M. Kirchberger, and S. Krutikova. Kagera health and development survey 2010: Basic information document. rockwool foundation working paper series 46, 2012.

J Diebolt and E HS Ip. A stochastic em algorithm for approximating the maximum likelihood estimate. Technical report, Sandia National Lab.(SNL-CA), Livermore, CA (United States), 1995.

Rosa Duarte, Sandra Ferrando-Latorre, and José Alberto Molina. How to escape poverty through education?: Intergenerational evidence in spain. *Applied Economics Letters*, 25(9):624–627, 2018.

Lisa M Gatzke-Kopp and Kristine L Creavey. Unsealing fate: Policy practices aimed at reducing the intergenerational transmission of poverty. *Policy Insights from the Behavioral and Brain Sciences*, 4(2):115–122, 2017.

Stephen Gibbons and Jo Blanden. *The persistence of poverty across generations: A view from two British cohorts*. The Policy Press on behalf of the Joseph Rowntree Foundation, 2006.

Caroline Harper, Rachel Marcus, and Karen Moore. Enduring poverty and the conditions of childhood: lifecourse and intergenerational poverty transmissions. *World development*, 31(3):535–554, 2003.

Stephen P Jenkins and Thomas Siedler. The intergenerational transmission of poverty in industrialized countries. *Chronic poverty research centre working paper*, (75), 2007.

Karen Moore. Frameworks for understanding the inter-generational transmission of poverty and well-being in developing countries. *Chronic Poverty Research Centre Working Paper*, (8), 2001.

Stefanos A Papanastasiou and Christos Papatheodorou. Intergenerational transmission of poverty in the eu: An empirical analysis. In *1st International Conference in Political Economy*, 2010.

Joan R Rodgers. An empirical study of intergenerational transmission of poverty in the united states. *Social Science Quarterly*, pages 178–194, 1995.

Kay Sharp. Measuring destitution: integrating qualitative and quantitative approaches in the analysis of survey data. 2003.

Sten-Åke Stenberg. Inheritance of welfare recipiency: an intergenerational study of social assistance recipiency in postwar sweden. *Journal of Marriage and Family*, 62 (1):228–239, 2000.

Xiaoying Wu, Xinhua Qi, Shan Yang, Chao Ye, and Biao Sun. Research on the intergenerational transmission of poverty in rural china based on sustainable livelihood analysis framework: A case study of six poverty-stricken counties. *Sustainability*, 11 (8):2341, 2019.

Appendices

Appendix A:

Table A1: Description of Missing Pattern in the Poverty Status Data

Response indicator in 1994	Response indicator in 2004	Missing pattern	Poverty status in 1991	1994	2004	2010	Frequency	Missing for	Remark
1	1	ABCD	1	1	1	1	1442	Complete data	
1	2	ABD	1	1		1	173	2004	
2	1	ACD	1		1	1	15	1994	
2	2	AD	1			1	8	1994 and 2004	
		ABC	1	1	1		1	2010	ignored
		AB	1	1			3	2004 and 2010	ignored
		A	1				2	1994, 2004 and 2010	ignored

Table A2: Test Result for the Latent Mover Stayer and Latent Markov Models under Different Assumptions about the Response Mechanisms

	Using incomplete data for sample children				Using all children originally interviewed in the baseline wave	
	Latent mover stayer		Latent Markov		Latent mover stayer	
Model	LR	Df	LR	df	LR	df
MCAR	51.21	21	63.87	24	65.93	35
Saturated MAR	31.02	15	36.64	17		
Second-order MAR	38.15	17	43.76	19	53.93	32
NMAR1	37.04	15	42.30	17	45.20	29
NMAR2	33.83	15	42.30	17	42.54	29
NMAR3	33.72	13	37.57	15	38.6475	26
NMAR4					46.18	29
NMAR5					35.37	23

Notes. The probability of missing data depends on poverty status in 1994 in NMAR1, in 2004 in NMAR2, in both 1994 and 2004 in NMAR3, in 2010 in NMAR4, and on poverty status in all three years in NMAR5.

Table A3: Summary of Estimated Parameter Values for Partially Latent Mover Stayer Model using Incomplete Data (Ignorable Missing Mechanism)

Chain proportion		Initial Probabilities		Transition Probabilities								Response Probabilities	
		t=1(=1991)		t to t+1 (from 1991 to1994)		t+1 to t+2(from 1994 to 2004)		t+2 to t+3(from 2004 to 2010)		t to t+3(from 1991 to 2010)			
Chain		Class		Poor	Non-poor	Poor	Non-poor	Poor	Non-poor	Poor	Non-poor	Poor	Non-poor
Stayers	0.149	Poor	0.078	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
		Non poor	0.921	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
Movers	0.850	Poor	0.645	0.723	0.276	0.632	0.367	0.391	0.608	0.309	0.690	0.653	0.346
		Non poor	0.354	0.058	0.941	0.374	0.625	0.174	0.825	0.250	0.749	0.000	1.000

Table A4: Summary of Estimated Parameter Values for Partially Latent Mover Stayer Model using Incomplete Data (Non-Ignorable Missing Mechanism)

Chain proportion		Initial Probabilities		Transition Probabilities								Response Probabilities	
		t=1(=1991)		t to t+1(from 1991 to 1994)		t+1 to t+2(from 1994 to 2004)		t+2 to t+3(from 2004 to 2010)		t to t+3(from 1991 to 2010)			
Chain		Class		Poor	Non-poor	Poor	Non-poor	Poor	Non-poor	Poor	Non-poor	Poor	Non-poor
Stayers	0.153	Poor	0.062	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
		Non poor	0.937	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
Movers	0.847	Poor	0.680	0.762	0.237	0.606	0.394	0.425	0.574	0.300	0.699	0.626	0.373
		Non poor	0.319	0.000	0.999	0.335	0.664	0.187	0.812	0.271	0.728	0.000	0.999

Table A5: Summary of Estimated Parameter Values for Partially Latent Mover Stayer Model using Incomplete Data for All Children (Non-Ignorable Missing Mechanism)

Chain proportion		Initial Probabilities		Transition Probabilities								Response Probabilities	
		t=1(=1991)		t to t+1 (from 1991 to 1994)		t+1 to t+2(from 1994 to 2004)		t+2 to t+3(from 2004 to 2010)		t to t+3(from 1991 to 2010)			
Chain		Class		Poor	Non-poor	Poor	Non-poor	Poor	Non-poor	Poor	Non-poor	Poor	Non-poor
Stayers	0.182	Poor	0.038	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
		Non poor	0.961	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
Movers	0.817	Poor	0.626	0.736	0.263	0.558	0.442	0.405	0.594	0.411	0.588	0.667	0.332
		Non poor	0.373	0.058	0.941	0.319	0.680	0.151	0.849	0.286	0.713	0.000	1.000

Table A6: Measurment of key Variables

Variable	Description of measures
Financial capital (F)	Average monthly expenditure per capita in the parental household over the years a respondent was observed in childhood
Natural capital (N)	Percapita farmland in the parental family
Physical capital (P)	Asset index constructed from different household assets using principal component analysis.
Human capital (H)	Household labour force and parental educational years measure household's human capital . When computing the household labour capacity, members under 12 were excluded, and members between 12 and 15 and above 61 were considered half labour.
Social capital (S)	Social participation is measured by indicator for whether any household member spent time helping neighbours or relatives outside the household with work on their farm or business without pay during the seven days preceding the survey, at least in one of the waves the respondent observed as a child.

Chapter 2

The Long-Term Effects of Early-Life Exposure to Weather Shocks: *Evidence from Tanzania*

The Long-Term Effects of Early-Life Exposure to Weather Shocks: Evidence from Tanzania

Ermias Gebru Weldesenbet *

Department of Economics, University of Copenhagen

February 23, 2022

Abstract

The paper examines whether early-life exposure to rainfall shocks has a long-term impact on health, education, and the socioeconomic statuses of individuals in rural Tanzania, where livelihoods heavily depend on rain-fed agriculture. We use a unique panel of data from a Kagera Health and Development Survey (KHDS) in which children were followed from childhood (1991) to adulthood (2010) together with historical rainfall data. We apply a sibling fixed-effect estimator to overcome potential endogeneity issues. We find that rainfall in birth year affects the education and socioeconomic statuses of children in adulthood. A 15 percent increase in rainfall in one's birth year and birth village (relative to average village rainfall) leads children to have 0.21 more years of schooling and live in a household in 2010 that scores 0.19 higher on an asset index. We then explore the relationship between early-life rainfall and childhood nutritional status to identify early-life rainfall's initial effect. We find that higher birth-year rainfall leads to significant decreases in height and weight deficits in children. A 15 percent increase in rainfall in one's birth year and birth village (relative to average village rainfall) improves height-for-age z score by 0.20 and weight-for-age z score by 0.26. When taken together, our results point to the importance of early childhood nutrition intervention. Sensitivity checks show that the results are robust to sample selection.

Keywords: rainfall shocks; malnutrition; long-term outcomes; children; Tanzania.
JEL Code: D12, I12, O15, O55, Q18, Q55

*Email:ermias.weldesenbet@econ.ku.dk

1 Introduction

A growing body of literature in development economics documents the long-term effect of exposure to shocks in early life. Recent evidence suggests that exposure to early-life shocks during gestation and early childhood is associated with adverse later-life outcomes, such as poor health and lower educational and socioeconomic status (Glewwe et al., 2001; Case et al., 2005; Alderman et al., 2006; Almond, 2006; Maccini and Yang, 2009). Most of the studies in the area come from developed countries, and there is relatively little work on developing countries due to a lack of data.[1] Given that the nature and frequency of early-life shocks and the availability and effectiveness of mitigation strategies are very different between developed and developing countries, much remains to be discovered about the long-term effects of early-life shocks in the context of developing countries. This paper examines whether early-life exposure to rainfall shocks has a long-term impact on individuals' health, education, and socioeconomic status in Tanzania.

The majority of previous research on developing countries examines the impact of more extreme types of early-life shocks such as civil war, famine, and pandemics (see, for example, (Alderman et al., 2006; Chen and Zhou, 2007; Meng and Qian, 2009; Umana-Aponte et al., 2011; Ampaabeng and Tan, 2013; Dercon and Porter, 2014). Nevertheless, rural households in developing countries commonly face less extreme shocks like rainfall shocks. Because a large proportion of the agricultural area in developing countries (more than 95 percent in sub-Saharan Africa) is rainfed, fluctuations in rainfall levels result in significant variations in agricultural output and farm incomes (Wani et al., 2009). Since there are limited social security arrangements and incomplete insurance markets in these countries, the reduction of agricultural output following an adverse rainfall shock generates a decrease in consumption. It could therefore have long-term implications for the welfare of rural households. To design policies that have high relevance for the rural population in developing countries, understanding rainfall shock's long-term effect is of

[1]Only a few datasets in Africa track individuals from childhood into adulthood.

prime importance.

While a growing number of recent studies look at the association between early-life weather shocks and child outcomes, they have focused on short-term effects. On the other hand, most studies analyzing the long-term effects of exposure to early-life weather shocks (Godoy et al., 2008; Maccini and Yang, 2009; Thai and Falaris, 2014; Cornwell and Inder, 2015; Zamand and Hyder, 2016; Chi et al., 2018; Fitz and League, 2020; Lin et al., 2021; Nübler et al., 2021; Yamashita and Trinh, 2021; Chang et al., 2022) also have some limitations. In nearly all studies, the outcome variables are observed only during adulthood and lack measurement of early life variables except for weather shocks. As a result, the studies do not indicate a link between exposure to weather shocks in early life and childhood health and nutritional status and, consequently, cannot establish the initial effect. Moreover, only a few studies compare how robust the results are to the problems caused by selective mortality due to a lack of data[2]

This paper provides new evidence of the long-term effect of early-life exposure to weather shocks using data fitting to address the aforementioned drawbacks. More specifically, the study examines the impact of early-life rainfall shock on adult health, education, and socioeconomic outcomes of individuals in Tanzania. We then explore the relationship between early-life rainfall and childhood nutritional status, measured by height-for-age and weight-for-age, to establish the initial effect. Given that previous researches provide mixed evidence regarding the timing of exposure to shocks, we also identify the stage in the early-life cycle during which children are most vulnerable to rainfall shocks.

We use a unique panel data from the Kagera Health and Development Survey (KHDS) in which children were followed from childhood (1991) to adulthood (2010). We also link historical rainfall data with the longitudinal individual data. Rainfall shock is measured using deviations from the long-run village average. Since crop production in Tanzania is predominantly rain-fed, higher rainfall is linked to higher income and

[2]For instance, Maccini and Yang (2009) depend on the relationship between birth year rainfall and cohort size in the district to test how selective mortality affects their result. They lack longitudinal data to check whether birth year rainfall affects the likelihood of survival to adulthood.

better nutrition through its effect on harvests. We apply a sibling fixed-effect estimator to control for any unobserved family and locality characteristics that may affect long-term outcomes. We allow for year of birth and season of birth fixed effects to control for any unobservables that vary with year and season of birth. We also have controls for birth order and gender effects.

We find that rainfall in birth year affects the education and socioeconomic status of children in adulthood. A 15 percent increase in rainfall in one's birth year and birth village (relative to average village rainfall) leads children to have 0.21 more years of schooling and live in a household in 2010 that scores 0.19 higher on an asset index. Such an increase in birth-year rainfall is also associated with a 30 percent increased likelihood to report rich economic status. By contrast, we find no significant relationship between birth-year rainfall and health outcomes in adulthood.

We then explore the relationship between early-life rainfall and childhood nutritional status to identify early-life rainfall's initial effect. We find that childhood nutritional status varies with birth-year rainfall. Higher birth-year rainfall leads to significant decreases in height and weight deficit in children relative to children from a well-nourished population. A 15 percent increase in rainfall in one's birth year and birth village (relative to average village rainfall) improves height-for-age z score by 0.20 and weight-for-age z score by 0.26. Compared with rainfall in the birth year, rainfall in the years before and after the birth year has no statistically significant relationship with childhood nutritional status and adult outcomes. The result suggests that rainfall shock in the year of birth matters the most. Because rainfall shock translates into income shock and food shortage with delay, nutritional deprivation resulting from adverse rainfall shock in a birth year likely occurs during the second year of life. This implies that nutritional intake during the second year of life is vital in influencing childhood health, adult education, and socioeconomic status. Overall, our findings suggest that anti-poverty interventions that promote the provision of nutritional supplements during the postweaning period have a significant long-run payoff. In addition, the findings underscore the importance of policies that help rural households smooth consumption.

Our results add to the growing body of research highlighting the period after weaning from breast milk as a stage during which children are most vulnerable to shocks. Glewwe and King (2001) find that malnutrition in the second year of life (compared to malnutrition in utero and the first year of life) has a more significant negative impact on cognitive development in the Philippines. Hoddinott and Kinsey (2001) also report that children aged 12 to 24 months exposed to drought in rural Zimbabwe lose 1.5–2 cm of growth after the drought. In a similar work, Alderman et al. (2006) show that children who were malnourished from 12 to 24 months of age are shorter and attain less schooling by adulthood than well-nourished children. In a study from Ethiopia, Dercon and Porter (2014) find that children exposed to famine from 12 months to 36 months are shorter, less likely to have finished primary school, and more likely to be ill by adulthood than older cohorts who were not exposed. Furthermore, Ampaabeng and Tan (2013) report that children who were newborn to 2 years of age at the time of famine in Ghana score lower on intelligence tests compared to older cohorts.

In related work, Alderman et al. (2009) also examine whether childhood nutrition affects future education in the Kagera region in Tanzania using data from the 2004 Kagera Health and Development Survey (KHDS). Their principal findings are that malnutrition in childhood delays school entry and lowers years of schooling attained in Tanzania. While there are some similarities between Alderman et al. (2009) and our work, there are three major differences that make our work unique. First, the authors use crop loss reported by a household as well as flood and drought weather shocks at the community level as instruments in estimating the effect of childhood health and nutrition status on educational outcomes, while we study the reduced form effect of early life rainfall shock directly using historical rainfall data. The direct consideration of the reduced form equation in our case, as opposed to the structural equation, is less likely to suffer from omitted variable bias. Moreover, since there are potentially many weaknesses associated with self-reported variables such as response bias, it is also advantageous to use an objective measure of rainfall shock.

Second, we focus on a wide range of outcomes instead of only on education outcomes, as in Alderman et al. (2009). Finally, outcome variables are observed in individuals' teenage years in Alderman et al. (2009), while we observe outcomes in both childhood and young adulthood. The average age of sample children in Alderman et al. (2009) is 16 years, while the average age in our data, as will be shown, is 23 years, with 80 percent of the sample older than 20 years of age.

The remainder of the paper is organized as follows: The next section presents a theoretical framework for the study, and Section 3 introduces the empirical model. Section 4 provides a brief description of the study area, the Kagera region of Tanzania. Section 5 discusses the data drawn from the Kagera Health and Development Survey (KHDS). Section 6 presents the empirical results, along with robustness checks. In the final section of the paper, Section 7, we summarize the findings and offer concluding remarks.

2 Theoretical Framework

There are several welfare indicators in adulthood that can potentially be explained by environmental conditions in the period around infanthood. For this paper, we chose to focus on adult economic wellbeing, health, and education. Accordingly, the welfare of a given adult at time (t) is given by the following vector of indicators:

$$W_t = [Y_t, H_t, E_t] \tag{1}$$

where Y_t is economic wellbeing as represented by consumption and wealth, H_t is health capital, and E_t is educational attainment.

In line with human capital theory, economic wellbeing at time (t) is determined, among other things, by the health and education capital of an individual:

$$Y_t = y(H_t, E_t, X_t, Z) \tag{2}$$

where X_t is a vector of other time-variant factors affecting economic wellbeing, and Z is a vector of time-invariant variables affecting economic wellbeing.

Adult health and education outcomes, in turn, can be defined as functions of childhood human capital. According to Grossman (1972), individual health at time t is given by a health production function consisting of initial health endowment, H_0, as well as a series of health inputs, N, acquired over time. The health production function may also consist of the demographic characteristics of the individual, D, and a vector of community-level infrastructure, disease, and environmental variables, C.

$$H_t = h(H_0, N_1, \ldots, N_t, D, C_1, \ldots, C_t) \tag{3}$$

Adult educational outcome is given by an educational production function. Bowles (1970) defines an educational production function as "the relationship between school and student inputs and a measure of school output" (p. 11). Hence, educational attainment in adulthood is a function of a series of individual educational and community- or school-level inputs acquired since early childhood. Apart from a flow of inputs, the initial endowment of cognitive capacity holds a cascading effect on education through adulthood. We assume that childhood health is a good proxy of initial cognitive capacity. Therefore, the educational production function can be written as follows,

$$E_t = e(H_0, I_1, \ldots, I_t, S_1, \ldots, S_t) \tag{4}$$

where I represents individual educational inputs, and S represents community- or school-level inputs. Equations 2–4 show that initial health endowment holds a direct or indirect effect on all of the adulthood welfare indicators specified in 1. Therefore, it is vital to have a good grasp of factors that are expected to affect initial health endowment. In this regard, initial health endowment is partly determined by genetic characteristics passed on from parents, G. Moreover, childhood health is a function of environmental conditions experienced in early life including at the fetal stage or during infancy, R_0. Finally, community-level health infrastructure and disease environment in early-life, C_0, influence childhood health.

$$H_0 = k(G, R_0, C_0) \tag{5}$$

Combining 2–5, adulthood welfare can be written as a function of environmental conditions in early life, individual demographic variables, and a series of household socioeconomic and genetic indicators and measures of community-level infrastructure and disease environment. Econometrically, this relationship can be estimated as a reduced-form function of early-life rainfall, individual demographic variables, and family fixed effects.

3 Estimation Strategy

This study aims to determine whether exposure to early-life rainfall shock affects outcomes later in life. We estimate the following reduced-form relationship in which an individual's adult outcome is posited to depend on early-life rainfall, year of birth, season of birth, birth order, gender, and family fixed effects.

$$Y_{ijt} = R\beta + X\alpha + \tau_t + \delta_t + \mu_j + \varepsilon_{ijt} \tag{6}$$

Where Y_{ijt} represents an outcome for individual i in family j at time t. The vector R includes rainfall variables from three years before birth to three years after birth. There are mixed findings in the literature on child nutrition. The "fetal origins hypothesis" emphasizes that nutritional deprivation in utero permanently reduces height in adulthood (Barker, 1998). On the other hand, more recent studies stress that nutritional deprivation during the period after weaning from breast milk has long-term irreversible effects (Glewwe and King, 2001; Hoddinott and Kinsey, 2001). We include rainfall at various points in early life to compare its impact on outcomes in adulthood.

X is a vector of child-level control variables, including birth order and gender. We control for birth order because competition with siblings for resources likely affects a child's nutritional status. τ is a year-of-birth fixed effect to account for any differences due to the child's year of birth. δ is a season of birth fixed effect to account for the fact that parents may time children to be born in particular seasons. Evidence shows that parents in sub-Saharan Africa time births across seasons in response to seasonality in labour demand and disease prevalence (Artadi, 2005). μ_j is a family-specific fixed effect that controls for any unobserved family characteristics that may affect adult outcomes, and ε_{ijt} is a mean zero error term. The coefficient β is the parameter of primary interest and represents the effect of rainfall shock during various stages of early life on outcomes in adulthood. Because the fixed effects absorb any unobservables that vary with the year, season, or household, the effect we pick up will be that of rainfall shock at various early-life stages.

Empirical studies of the long-term effect of early-life conditions are haunted by the fact that unobserved factors may influence early-life health and subsequent adult outcomes. This study identifies the links between early-life health and long-term outcomes indirectly through rainfall variations. Although causality is less difficult to ascertain in rainfall data, one may question the exogeneity of early-life rainfall because households could choose settlements based on rainfall distribution. For instance, Frankema et al. (2017) show that rainfall levels and variability are associated with district-level population density in tropical Africa. Another possible concern is that rural families could change fertility patterns in response to rainfall shock. For instance, Shah and Steinberg (2017) argue that rural mothers may delay fertility during drought because fathers could migrate in search of work. We apply a sibling fixed-effect estimator to control for any unobserved family and locality characteristics that may affect long-term outcomes. By using within sibling estimation, in addition to addressing the concerns above, we can rule out potential attrition bias resulting from household and locality characteristics.

We use a sibling fixed-effect model for continuous outcomes and a fixed-effects logit model or Chamberlain's conditional logit model (Chamberlain, 1980) for dichotomous outcomes. We need multiple observations (siblings) to estimate a fixed-effect model. In the simplest case of two siblings per household, differencing the observations within a family eliminates the family fixed effect. The model then estimates the rainfall effect by regressing within-family differences in the outcome on the differences in the independent variables. Identifying the effect of early-life rainfall on adulthood outcomes relies on differences in rainfall at early life between or among siblings. Given that the birth village is the same among siblings, within-family variation in rainfall exists due to differences in the year of birth and season of birth. In the fixed-effects logit model, only families with a difference in siblings' outcomes contribute to the likelihood function. Families in which all siblings have the same outcome are not used in estimating the parameters of the model, which reduces the effective sample size.

4 Background

Kagera is a region in the northwestern part of Tanzania. It is one of Tanzania's most remote areas, situated about 1400 kilometres from the capital Dar es Salaam. The region contains a large part of Lake Victoria and shares borders with Uganda in the north and Burundi and Rwanda in the west. Kagera consists of about 30,000 square kilometres of land surface and approximately 10,000 square kilometres of water surface and is divided into five districts: Biharamulo, Bukoba, Karagwe, Muleba, and Ngara.

According to the 2012 census, the population of Kagera is 2.5 million (URT, 2012b). It is made up of diverse ethnic groups with Haya and Nyambo tribes dominating in the north, and Subi, Sukuma, Zinza, and Hangaza in the south. Kagera is overwhelmingly rural, with more than 80 percent of the region's economically active population engaged in agriculture (URT, 2012a). The agricultural production system in the region is primarily rain-fed. Only 2 percent of the households involved in agriculture reported having access to irrigation in 2007/08 (URT, 2012a). The farming system is also characterized by traditional cultivation methods and low input low yield subsistence agriculture. Furthermore, agriculture in the region is mostly dominated by smallholder farmers.

Kagera climate consists of two rainy seasons and two dry seasons. The short rainy season (Vuli) occurs between October and December, and the long rainy season (Masika) occurs between March and May. Kagera experiences a long dry season between June and September and a short dry season between January and February. The long rains and the short rains are associated with the northward and southward movement of the Inter-Tropical Convergence Zone (ITCZ), respectively. The long rains are more intense and less variable compared to the short rains. The data used in this paper shows that Kagera receives, on average, 37 percent of total annual rainfall during the long rainy season and 33 percent of total annual rainfall during the short rainy season.

Most of the Kagera region has two cropping seasons a year under rainfed conditions. Crop production covers a wide range of crops, including cash crops – such as coffee, tea, and cotton – and food crops, such as bananas, beans, maize, and cassava. The region is one of the largest coffee-producing areas of Tanzania. Perennial crops (such as banana and coffee) mainly grow in the northern part. In contrast, annual crops (such as beans, maize sorghum, and cassava) grow in Kagera's southern region. Planting for annual crops typically occurs in March, while harvesting occurs in July and August.

Because agricultural production systems are mostly rain-fed, the impact of rainfall variability is highly pronounced in Tanzania. Studies have shown the importance of rainfall variation in explaining crop yield. Rowhani et al. (2011) examined the relationship between rainfall variability and crop yields in Tanzania by focusing on maize, sorghum, and rice. They found that higher rainfall is associated with increased yields. The result also showed that increased precipitation variability during the growing season harms average yields. A 20% increase in intra-seasonal rainfall variability reduces agricultural yields by 4.2%, 7.2%, and 7.6%, respectively, for maize, sorghum, and rice.

Lema and Majule (2009) find that shortening of the rainy season reduces yields. Moreover, Bengtsson (2010) reports that households in the Kagera region are exposed to significant income fluctuation resulting from rainfall shocks.

Apart from its effect on harvest and farm income, rainfall may affect child nutrition and health through other channels. For instance, by changing the relative cost of time, rainfall may affect parental time between farm work and child-rearing. Positive rainfall shock may imply more work on the farm and less time for child-rearing, while the opposite may hold true for negative rainfall shock. Additionally, high rainfall may lead to the development of parasitic and infectious diseases such as malaria and cholera. Exposure to infectious diseases, in turn, affects the body's ability to absorb nutrients, especially during early childhood when nutritional needs are high. Because these adverse effects of rainfall could offset the positive effects of improved harvest, the direction of the relationship between early-life rainfall, childhood health, and thus long-term outcomes for children is theoretically ambiguous. In this study, we expect the positive effect of rainfall on early-life health outweighs the negative effect for two reasons. First, rainfall's potential substitution effect is likely to be weak when more nutrition and less time input are required in early life. Second, compared to drought, flooding is not a recurring problem in the Kagera region because of the terrain's undulating nature and the focus on tree crop production.

5 Data

5.1 Kagera Health and Development Survey (KHDS)

The study uses long-running household panel data from the Kagera Health and Development Survey (KHDS)[3], conducted in the northwestern region of Tanzania. The survey was initially designed and implemented by the World Bank and the Muhimbili University College of Health Sciences. The KHDS has six waves of data collected be-

[3]The KHDS was originally adapted from the World Bank's Living Standards Measurement Study (LSMS) questionnaires

tween 1991 and 2010. The first four waves of the data were collected from 1991 to 1994 at six-month intervals. Overall, 915 households were interviewed during the baseline survey in 1991–94. The original households were selected from 51 clusters using two-stage stratified sampling. Excluding families for which all members were deceased, the survey re-interviewed 93 percent of the baseline households in 2004 and 92 percent in 2010. As the survey re-interviewed all baseline household members even if they moved out of their original households, the sample increased to 2700 households in 2004 (see Beegle et al. 2006, for details) and to 3300 households in 2010 (see De Weerdt et al. 2012, for more information). Overall, the re-contact rate is well above most known household panel surveys in developing countries summarized in Alderman et al. (2001).[4]

5.2 The Sample

The study sample comprised children of household heads aged ten years and below living with their parents during the baseline survey in 1991–94. That means the sample consists of cohorts born between 1981 and 1993. There were 1491 children aged ten years and below living with their parents in 512 households in the baseline survey. Out of these, 120 were deceased before 2010, and 243 were not traced during the survey's last round. We were left with 1128 children interviewed in 1991–94 and 2010. After further restricting the analysis sample to households with more than one child aged ten years and below from 1991–94, the final sample consists of 985 children from 309 households.

5.3 Description of Children in the Sample

Summary statistics for key variables are presented in Table 1. Boys make up 51 percent of the sample. The sample children's average age in 1994 was six years, with no significant difference between girls and boys. In 2010, the children's average age was 23 years, with 80 percent of the sample above 20 years old. In the same year, about 43

[4]The attrition rate reported in Alderman et al. (2001) ranges between 1.5 percent and 17.5 percent, with most of the surveys included in the study spanning short period.

percent of the children were married. Girls tend to marry earlier than boys do (53 vs 32 percent). Only 12 percent of the children were enrolled in school in 1994. The average years of schooling among those enrolled in school was 0.8 years, with no significant difference between girls and boys. In 2010, the average years of schooling increased to 7 years, with only 23 percent of the sample having above primary-level education. In the same year, about 62 percent of the children were involved in agriculture, and about 30 percent in paid employment (formal and informal). In comparison, about 8 percent were still enrolled in school. A larger proportion of boys (40 percent) were involved in paid employment outside agriculture as compared to girls (19 percent).

In all of the survey waves, anthropometric measurements were taken from all household members who were present at the time of the household interview. Trained anthropometrists from 1991–94 and enumerators in 2010 were responsible for taking the measurements. While heights of household members less than three years old were measured using a length board with a sliding foot piece, the heights of those older than three years were measured using a height board with a sliding headpiece. Infants below the age of two were weighted using hanging Salter scales. In contrast, standard scales in Wave 1 and Wave 2 and digital scales from Wave 3 were used for older cohorts. In 2010, anthropometric measurements were taken for 866 children out of the sample of 985 children. The most common reason why a measure was not taken is that children were away from home during the interview.

Children's nutritional status has been widely assessed by height-for-age (HAZ) and weight-for-age (WAZ) z scores. The HAZ and WAZ express height and weight as several standard deviations below or above the reference median value. For instance, a HAZ score of -1 shows that the child's height is one standard deviation below the median child of a healthy and well-nourished reference population for the same age and sex. The height and weight data were transformed into HAZ and WAZ scores using the 1990 British Growth Reference population data.

Overall, the sample children have poor nutritional status in childhood, measured by

Table 1: Summary Statistics of key variables

Variables	Mean	Std. Dev.	Min	Max	Obs
Childhood measures					
Height (centimeters)	106.46	22.76	47.3	159.1	982
Weight (kilograms)	18.41	7.64	3	45	982
Height-for-age z score	-1.17	1.4	-5.8	5.7	979
Weight-for-age z score	-1.16	1.2	-4.11	5.5	981
Adulthood measures					
Height (centimeters)	161.43	8.74	111.3	208.6	866
Years of schooling	7.18	2.49	0	15	862
Ln(expenditure per capita in household)	15.81	0.66	14.12	18.87	979
Self-reported rich economic status (indicator)	0.66				985
Asset index	-0.002	1.33	-1.98	2.79	985
Owns a house with good floor (indicator)	0.38				985
Owns radio (indicator)	0.70				985
Owns telephone (indicator)	0.60				985
Owns motor vehicle (indicator)	0.06				985
Rainfall measures					
Deviation of log rainfall from norm, year of birth	-0.027	0.149	-.421	0.362	985
Deviation of log rainfall from the norm, one year prior to the birth-year	-0.019	0.179	-2.153	0.405	984
Deviation of log rainfall from the norm, two years prior to the birth-year	-0.011	0.196	-2.153	0.963	983
Deviation of log rainfall from the norm, three years prior to the birth-year	0.015	0.189	-1.172	0.963	984
Deviation of log rainfall from the norm, one year after the birth-year	-0.047	0.125	-0.421	0.282	985
Deviation of log rainfall from the norm, two years after the birth-year	-0.041	0.120	-0.421	0.284	985
Deviation of log rainfall from the norm, three years after the birth-year	-0.037	0.113	-0.328	0.286	985

Height and weight in childhood were measured in 1991-94. The rest of the outcome variables were measured in 2010.

their HAZ score. In 1991-94, about 82 percent of the children had negative HAZ scores. The mean HAZ in the period was -1.3 for boys and -1.1 for girls. About 26 percent of the sample children were stunted[5] by WHO standards (i.e. they have a HAZ score of -2 or lower), compared to 42 percent reported in DHS for the whole of Tanzania National Bureau of Statistics 2011. Figure A1 in the Appendix depicts a distribution of the HAZ score by sex.

Table 2 compares the health and socioeconomic outcomes of children who were stunted during childhood and those who were not. In 1991-94, the mean HAZ score for stunted children was -2.7 compared to -0.6 for not stunted children. Children who were stunted during childhood were shorter, completed fewer years of schooling, scored less on the asset index, and had less per-capita household expenditure in 2010. In the same period, a smaller proportion of children who were stunted during childhood owned a house with a good floor, radio, telephone, and motor vehicle compared to those who were not stunted. In addition, a lower proportion of children who were stunted during childhood described their households as rich. The differences are statistically significant and suggest associations between childhood nutritional status and subsequent socioeconomic outcomes during adulthood.

5.4 Rainfall Data

We use historical rainfall data obtained from two sources. For 1981–2010, the data were obtained from NASA's Modern-Era Retrospective analysis for Research and Applications (MERRA). The data contains the daily total millimetres of rain. We also use additional monthly rainfall data obtained from the Tanzania Meteorological Agency for 1980-2004. The data was taken from 21 stations. The metrological data is linked to the 51 KHDS baseline villages (clusters) using GPS coordinates.

We use the information on the month of birth to identify the season during which a child was born. The month of birth is reported in the dataset for 971 observations.

[5] A child is considered as stunted if his or her HAZ score is below -2.

Table 2: Mean Outcomes of Children by Nutritional Status

Variables	Malnourished children (stunted)	Non-malnourished children (not stunted)
Childhood measures		
Height (centimeters)	101.86	108.16
Weight (kilograms)	16.67	19.05
Height-for-age z score	-2.78	-0.60
Weight-for-age z score	-2.33	-0.75
Adulthood measures		
Height (centimeters)	157.27	162.94
Years of schooling	6.67	7.36
Asset index	-0.36	0.13
Ln(expenditure per capita in household)	15.69	15.86
Self-reported rich economic status (indicator)	0.61	0.68
Owns a house with good floor (indicator)	0.29	0.41
Owns radio (indicator)	0.65	0.72
Owns telephone (indicator)	0.48	0.64
Owns motor vehicle (indicator)	0.02	0.08

Height and weight in childhood were measured in 1991-94. The rest of the outcome variables were measured in 2010.

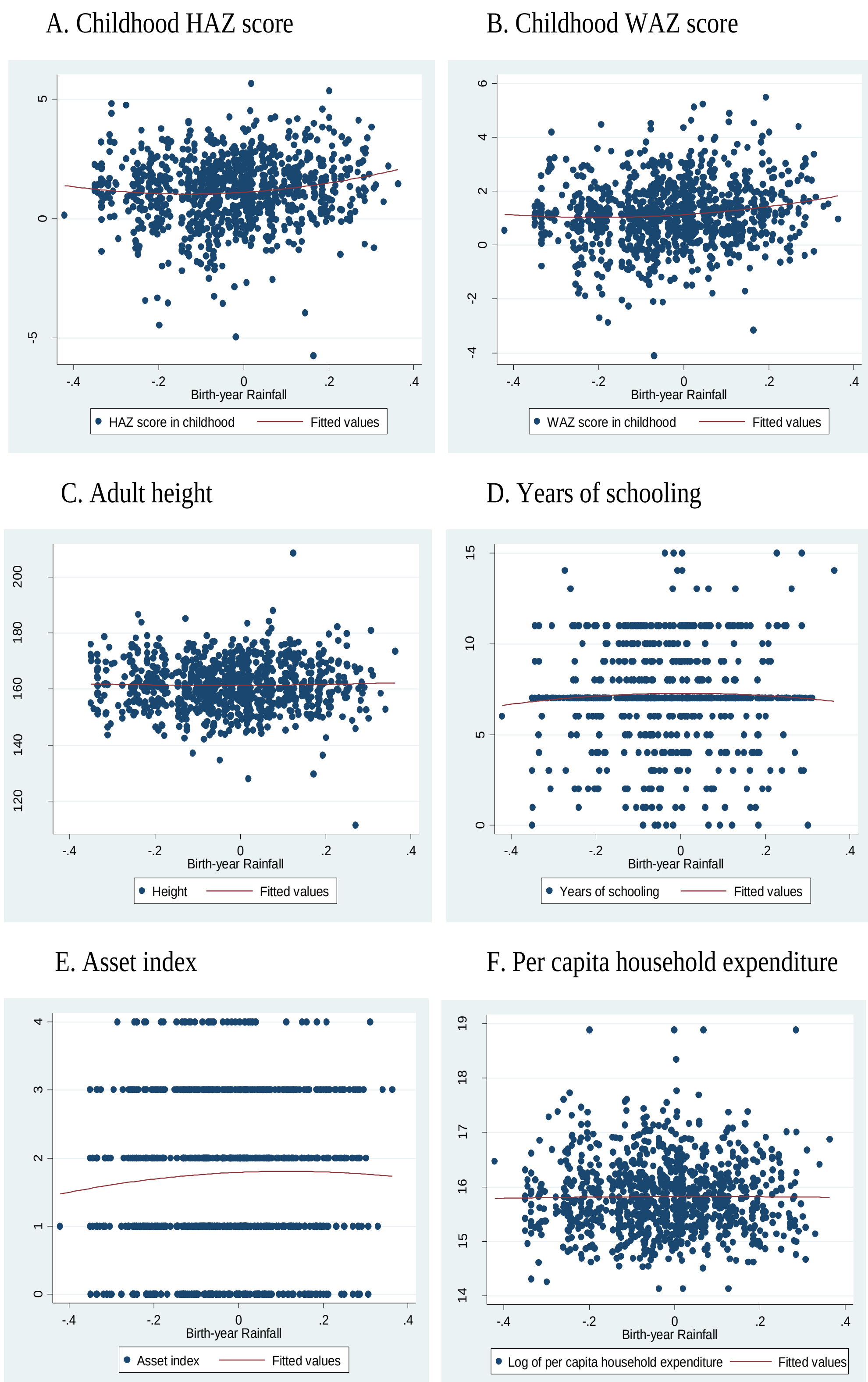
For the remaining 14 observations, we assume that children were born in the middle of two consecutive survey waves. Rainfall in a child's birth year is then calculated by focusing on rainfall in four successive seasons instead of on rainfall in the calendar year. For instance, if one was born in October or November, we calculate total rainfall in the following four seasons, starting from the short rainy season. For children born in the last month of a given season, birth year rainfall is calculated by starting from the season following the birth season. Our approach is similar to that of Maccini and Yang (2009) and closely related to agricultural activities' timing.

We construct the measurement for rainfall shock in a birth year based on the deviation of birth year rainfall from the long-run average in one's birth village. We first calculate average annual rainfall for each village (cluster) using the period 1980–99. Each observation is then assigned the difference found in the logarithm between birth year rainfall and average annual rainfall. A similar approach is used to calculate the deviation of rainfall from the long-run average in the years before and after the birth year. The rainfall variable can be approximately interpreted as the percentage deviation from the average rainfall in a given village. For instance, a value of -0.1 means that rainfall is less than the village average by approximately 10 percent.

We estimate relationships between rainfall shock at various early-life stages and several outcomes in adulthood that we group into health, education, and socioeconomic outcomes. The health outcome is measured using height, an indicator of whether the person had been ill during the last four weeks prior to the survey in 2010, and the number of days ill. Educational attainment is measured by years of schooling. Our measure of years of schooling is assigned based on the number of years required to complete the highest grade attained. Socioeconomic status is measured by the log of per capita household expenditure, household asset index, and an indicator of whether a respondent described his household as rich. The household asset index is constructed using principal component analysis. Variables used to construct the index include indicators for ownership of a house with a good floor, radio, telephone, refrigerator, and motor vehicle.

A graphical representation of some of the relationships between each child's outcome and birth-year rainfall is displayed in scatter plots in Figure 1. Panel A shows the relationship between birth-year rainfall (on the horizontal axis) and childhood HAZ score (on the vertical axis). While the horizontal axis variable remains the same, the vertical axis in panels B, C, D, E, and F represents childhood WAZ score, adult height, years of schooling, household asset index, and per capita household expenditure, respectively. Panels A and B respectively show a positive relationship between birth-year rainfall and childhood HAZ score and between birth-year rainfall and childhood WAZ score. By contrast, the direction of relationships between birth-year rainfall and the respective adult outcomes are not apparent in the rest of the panels. In the following section, we will examine whether the relationships we observe from the graphs hold after controlling for various observable characteristics.

Figure 1: Birth Year Rainfall and Child Outcomes.



6 Results and Discussion

6.1 Early-Life Rainfall Shock and Adult Outcomes

We now turn to a discussion of the results. We examine the impact of rainfall shock around the time of birth on adult health, education, and socioeconomic outcomes observed in 2010. Table 3 presents results from sibling fixed-effect estimation of equation 6. Sibling fixed-effect model estimates the rainfall effect based on within-family differences in outcomes among siblings. Each row reports results from a separate regression of the outcome variable on rainfall around the time of birth. Each column presents a coefficient on rainfall variable in different years (the birth year and before and after the birth year). Year 0 is the birth year, year -1 is the year before the birth year, year +1 is year after the birth year, and so on. In addition to rainfall variables, all regressions include fixed effects for birth year and birth season, birth order, and a dummy variable indicating female, although coefficients for these additional controls are not shown in the table. As shown in the table, the effective sample size for dichotomous outcomes is significantly reduced, since only families in which outcome differences exist among siblings can be included in the estimation.

We begin with results for health outcomes. The coefficient on birth year rainfall (year 0) in height regression has a negative sign as opposed to expectation. Rainfall in the birth year has a negative relationship with an indicator for illness during the last four weeks and the number of days ill. However, the coefficients in regression for all the three health outcomes are not statistically significantly different from zero at conventional levels. The effect on an indicator for ill during the last four weeks and the number of days ill are not apparent, maybe because these variables measure short-duration illnesses and are thus generally weak discriminators for health.

Birth-year rainfall (year 0) has a statistically significant positive relationship with years of schooling attained. The coefficient on birth-year rainfall in regression for expenditure per capita in a household is positive, but it is not statistically significantly

Table 3: Early-life Rainfall Shock and Adult Outcomes

	Year-0	Year-1	Year-2	Year-3	Year+1	Year+2	Year+3
Adult Height (centimeters)	-0.682	1.781	3.618	-2.888	-2.855	-1.767	-4.569
(n=862)	(3.398)	(1.860)	(1.828)**	(1.908)	(3.801)	(4.037)	(4.030)
Ill during the last 4 weeks	-0.915	-0.460	1.528	-0.482	-1.690	-1.815	-2.451
(indicator) (n=604)	(1.214)	(0.701)	(0.822)*	(0.705)	(1.391)	(1.454)	(1.413)*
Days ill	-3.639	-0.597	2.384	0.791	-3.806	-8.061	-4.603
(n=953)	(2.559)	(1.510)	(1.454)	(1.498)	(2.950)	(3.161)**	(3.106)
Years of schooling	1.412	0.510	0.541	0.091	-1.681	1.017	1.076
(n=860)	(0.727)*	(1.064)	(0.660)	(0.622)	(1.203)	(1.330)	(1.269)
Ln (expenditures per capita	0.083	0.115	-0.047	0.279	-0.472	-0.216	-0.123
in household)	(0.244)	(0.151)	(0.139)	(0.143)*	(0.278)*	(0.297)	(0.288)
(n=975)							
Asset index	1.271	0.296	0.328	0.026	-0.928	0.269	0.121
(n=981)	(0.494)**	(0.294)	(0.283)	(0.290)	(0.563)*	(0.602)	(0.583)
Self-reported rich status	1.954	1.393	-0.112	-0.619	0.303	-1.217	-0.021
(indicator) (n=483)	(0.959)**	(1.431)	(0.654)	(0.782)	(1.513)	(1.601)	(1.725)

All outcome variables were measured in 2010. Standard errors are reported in parenthesis. $*p < 0.1; **p < 0.05; ***p < 0.01$. *The effective sample size for each outcome variable shown in parenthesis next to the variables.*

different from zero. Birth-year rainfall has a statistically significant positive effect on asset index and self-reported rich economic status.

We discuss the size of the estimated effects focusing on a 0.15 log point change in birth-year rainfall, which is the standard deviation of rainfall deviation from mean village rainfall in the year of birth. A 0.15 log point increase in rainfall in one's birth year and birth village leads to children having 0.21 more years of schooling and living in a household that scores 0.19 higher on an asset index. The same increase in birth-year rainfall is also associated with an increased likelihood of 30 percentage points to report rich economic status. This is quite a large effect compared to the base reporting propensity of 66 percent. Our results are very similar to Maccini and Yang (2009); they find that higher rainfall in the first year of life positively affects adult health, education, and socioeconomic status of women in Indonesia. However, unlike Maccini and Yang (2009), we find no significant relationship between birth-year rainfall and adulthood health outcomes.

In other years (before and after the birth year), rainfall has little or no relationship with adult outcomes. Most of the coefficients on rainfall in other years are not

Table 4: Effects of Birth-year Rainfall on Adult Outcomes

Adult Height (centimeters)	-1.128
	(3.190)
Ill during the last 4 weeks (indicator)	-0.487
	(1.079)
Days ill	-2.598
	(2.446)
Years of schooling	1.532
	(0.709)**
Ln (expenditures per capita in household)	0.126
	(0.232)
Asset index	1.156
	(0.470)**
Self-reported rich status (indicator)	1.824
	(0.907)**

All outcome variables were measured in 2010. Standard errors are reported in parenthesis. $*p < 0.1; **p < 0.05; ***p < 0.01$.

statistically significantly different from zero. We tested the joint significance of rainfall in the years before birth (year-1, year-2, year-3) together and in the years after birth (year+1, year+2, year+3) together for all the regressions in Table 3. The results reveal that, in most cases, we do not reject the hypothesis that the coefficients on rainfall in other years (before and after the birth year) are jointly insignificantly different from zero. This is evidence that only rainfall shock in the year of birth is vital in influencing adulthood outcomes.

Because coefficients on rainfall in other years (before and after birth year) are not statistically significantly different from zero, we also estimate a specification only with rainfall in the birth year. We include birth year and birth season, birth order, and a dummy variable indicating females as additional controls similar to above. The results are reported in Table 4. As shown in the table, the coefficients on birth-year rainfall do not change in magnitude or significance levels in most of the regressions. Similar to the result above, the coefficients on birth-year rainfall in regressions for health outcomes and expenditure per capita in households are not statistically significant. The coefficients in regressions for years of schooling, self-reported rich status, and asset index are positive and statistically significant. Overall, the result confirms that rainfall in the birth year matters the most.

6.2 Early-Life Rainfall Shock and Childhood Health Outcomes

The previous section established that rainfall in the birth year has a long-run effect on children. We find that rainfall shock in the birth year affects children's education and socioeconomic status in adulthood. For this evidence to be complete, we need to identify the initial effect of early-life rainfall shock. No studies in the area (except Alderman et al. 2006) do this because they lack the measurement of outcome variables in early-life. This section examines the relationship between early-life rainfall and childhood nutritional status in order to establish the initial effect. Our empirical specification in this section is very similar to that used in the previous section, except that our left-hand-side variable is now childhood nutritional status. As mentioned earlier, we measure children's nutritional status using height-for-age and weight-for-age z-scores.

Table 5 reports results from the sibling fixed-effect estimation. Column 1 shows the relationship between early-life rainfall and the HAZ score. The result indicates that rainfall in one's birth year and birth village is positively associated with the HAZ score. Column 2 presents the corresponding result for the WAZ score. The result demonstrates that rainfall in one's birth year and birth village has a statistically significant positive relationship with the WAZ score, echoing the HAZ regressions results.

We apply parameter estimates from sibling fixed-effect regression reported in Table 5 to identify the magnitude of the effect of change in birth-year rainfall. Once again, we focus our discussion on the effect of a 0.15 log point change in birth-year rainfall. A 0.15 log point incr

ease in rainfall in one's birth year and birth village increases the HAZ score by 0.20 (compared to the mean initial HAZ score of -1.17) and the WAZ score by 0.26 (compared to the mean initial WAZ score of -1.16). This means that higher birth-year rainfall leads to significant decreases in height and weight deficits in children relative

Table 5: Early-life Rainfall Shock and Childhood Health Outcomes

	HAZ	WAZ
Year 0	1.348	1.728
	(0.635)**	(0.553)***
year-1	-0.324	-0.138
	(0.345)	(0.302)
Year-2	0.282	0.371
	(0.339)	(0.296)
Year-3	-0.009	0.094
	(0.354)	(0.310)
year+1	0.763	0.572
	(0.704)	(0.617)
Year+2	0.291	0.466
	(0.754)	(0.659)
Year+3	0.659	0.752
	(0.748)	(0.654)

Notes: The sample includes children of the household head who, when measured in 1991–94, had height-for-age z and weight-for-age z scores between –6 and +6 and were re-interviewed in 2010. Height-for-age z scores less than -6 or greater than 6 indicate errors in measures of either height or age. Standard errors are reported in parenthesis. Asterisks indicate the level of significance: $*p < 0.1; **p < 0.05; ***p < 0.01$.

to children from a well-nourished population. Put differently, the result suggests that higher birth-year rainfall significantly improves childhood nutritional status.

In other years (before and after the birth year), rainfall has no relationship with childhood health outcomes. In both regressions, the coefficients on rainfall in other years are smaller in magnitude compared to the coefficients on birth-year rainfall. In addition, most of the coefficients on rainfall in other years have a negative sign, and none are statistically significantly different from zero. We test the joint significance of rainfall in the years before birth (year-1, year-2, year-3) together and in years after birth (year+1, year+2, year+3) together. The result shows we do not reject the hypothesis that the coefficients on rainfall in other years are jointly insignificantly different from zero. Our finding suggests that only rainfall shocks in the birth year are essential in influencing childhood nutritional status.

6.3 Discussion

This paper's key finding is that rainfall in a birth year is positively associated with childhood health status, adult education, and socioeconomic status. In interpreting this result, it is essential to understand that rainfall shock translates into income shock and nutritional deficiency with delay. The delay exists mainly because the fall in yields following drought occur during harvest season. In Kagera, harvest for major annual crops typically occurs between July and August, months after the long rainy season. Besides, rural households usually have stores of food enough to meet their basic nutritional needs for some period, which also delays the occurrence of drought-induced nutritional deprivation. Hence, nutritional deprivation resulting from rainfall shock in a birth year likely occurs during the second year of life. This implies that nutritional deprivation occurring during the second year of life is critical in influencing childhood health, adult education, and socioeconomic status.

The literature gives us insight into why the second year of life can potentially be the most critical period. Children are protected from nutritional deficiency when they are intensively breastfed in the first year of life. Additionally, intensive breastfeeding buffers infants against pathogen exposure. In contrast, children just above breastfeeding age are vulnerable to malnutrition because the buffering role of breastfeeding attenuates, and children are not robust to shocks at that age. In Tanzania, breastfeeding is a widespread practice. Approximately 94 percent of infants are continually breastfed up to 12–15 months of age, but the proportion significantly decreases to 51 percent at 20–23 months of age (National Bureau of Statistics, 2005). The finding that nutritional deprivation during the second year of life influences childhood health, adult education, and socioeconomic status agrees with several bodies of literature (Glewwe and King, 2001; Hoddinott and Kinsey, 2001; Alderman et al., 2006; Maccini and Yang, 2009; Ampaabeng and Tan, 2013; Dercon and Porter, 2014)

Finally, it is instructive to compare the estimated impact of higher rainfall on adult outcomes with early-life nutrition intervention. In making the comparison, we focus on the effects on schooling, the most common outcome studied in the literature. Field et al. (2009) documented the intensive iodine supplementation program's impact on adult-child schooling in Tanzania. The authors found that children who benefited from iodine supplements in utero attain an average of 0.35 years of additional schooling. In our analysis, a 15 percent increase in rainfall in one's birth year and birth village leads to 0.21 more years of schooling, which is roughly two-thirds of the estimated impact in Field et al. (2009). It is worth noting that higher birth-year rainfall also affects outcomes other than schooling.

6.4 Robustness to Sample Selection

In this section, we explore whether or not selection into our sample might confound the results. Our sample consists of children of household heads aged ten years and below in 1991–94 who were re-interviewed during the last round of the survey in 2010. Individuals who were not traced in 2010 and, thus, whose long-term outcomes were not measured are not included in the data for analysis. This may present an obstacle to detecting the relationship between early-life rainfall and adult outcomes. In this study, attrition bias could result from selective mortality between birth and 2010 and attrition due to other reasons. We examine the likely impact of selective mortality and overall attrition in the data separately.

As mentioned earlier, 120 children in the relevant age cohort in 1991–94 died before 2010. Table A1 in the appendix shows the mean values of childhood outcomes by mortality status. As table A1 indicates, those who died before 2010 and those who survived through 2010 have many significant differences in characteristics. Children who died before 2010 were younger, had worse HAZ and WAZ scores, and lived in households with lower per capita consumption during childhood. Also, children who died before 2010 were more likely to experience stunting and be underweight during childhood. All differences in means are statistically significant at least at the 5 percent significance level. The clear implication of this pattern is that children who died before 2010 had worse childhood health. Suppose early-life rainfall affects the likelihood of survival through 2010. In that case, our estimates of the long-term effect of rainfall on adult outcomes based on surviving children will be biased downward.

As a test for mortality selection, we check whether early-life rainfall affects the likelihood of survival until 2010. Specifically, we regress an indicator for survival until 2010 on birth-year rainfall. The results are presented in Table A2 in the appendix. We find that the coefficient on birth-year rainfall is not statistically significantly different from zero. In column 2, we include rainfall in years before and after birth year and find that none of the rainfall variables bear a statistically significant association with the likelihood of survival until 2010. In column 3, the regression includes birth year, birth season and district fixed-effects, birth order, and a dummy variable indicating female comparable to the main analysis specification. Again, we find no evidence that early-life rainfall affects the likelihood of survival until 2010. Therefore, the results suggest that bias resulting from selective mortality is not a problem for our analysis.

Another source of attrition is the fact that some children were not traced in 2010.

Besides, the fact that information on adult height is missing for some children who were not present at the time of measurement further adds to attrition. In what follows, we examine whether overall attrition in the data systematically affects our result. We conduct three sets of tests of attrition, similar to Fitzgerald et al. (1998). First, we compare the means of childhood outcomes by attrition status. Specifically, we check whether childhood outcomes differ between those who lost to follow up and those who form the final sample. Children who lost to follow up include those who died, those who were not traced, and those who were not present at the time of anthropometric measurement. As Table A3 in the appendix shows, children who lost to follow up are younger and lived in a household with higher per capita consumption during childhood than those who form the final sample. A t-test shows that the differences in means are statistically significant. On the other hand, there are no apparent differences in childhood health outcomes. Although children lost to follow up have worse HAZ scores and are more likely to be stunted during childhood, the differences in means are small and not statistically significant.

We estimate probit equations for the probability of attrition in order to determine the presence of selection on observables. The dependent variable in the probits equals 1 if attrition occurred in 2010; 0 otherwise. We include childhood outcome variables (that is, in the language of Fitzgerald et al. (1998), lagged outcome variables) as regressors. We estimate expanding specifications of attrition probits, as shown in Table 6. In the first three columns, lagged outcome variables are included one at a time. The result shows that only per capita household consumption during childhood predicts attrition; that is, children who lived in households with higher per capita consumption during childhood are more likely to be lost to follow up. This pattern is also found from the comparison of means, as noted earlier. In column 4, all three lagged outcome variables are included at the same time. We find that none of the coefficients on lagged outcome variables are statistically significantly different from zero. In column 5, a broader set of variables (birth year, birth season, district fixed effects, birth order, and a dummy

Table 6: Probit of Attrition

	Single lagged outcome at a time			All three	Other controls
	1	2	3	lagged outcomes	included
HAZ score in childhood	-0.025			-0.034	-0.047
	(0.026)			(0.043)	(0.047)
WAZ score in childhood		-0.002		0.026	0.028
		(0.031)		(0.051)	(0.055)
Ln (per capita household consumption)			0.137	0.088	-0.010
			(0.068)**	(0.075)	(0.095)
N	1,278	1,275	1,359	1,270	1,263

Notes: Standard error are reported in parenthesis. Asterisks indicate the level of significance: $*p < 0.1; **p < 0.05; ***p < 0.01$

denoting female) are added. With the addition of these controls, none of the lagged outcome variables bear a significant association with attrition. The coefficient on per capita household consumption even takes the opposite sign. Therefore, it appears that there is no systematic attrition that would bias our results.

Lastly, we test whether coefficient estimates for childhood outcome regressions using only non-attriting sample differ from those using the total sample. We estimate the relationship between early-life rainfall and childhood health outcomes (measured by height-for-age and weight-for-age z scores) separately for the whole and the non-attriting sample. Our empirical specification here is the same as that used in section 7.2. Results from sibling fixed-effect estimations are presented in Table 7. It can be seen that the results for the non-attriting sample are virtually identical to those found in section 7.2. We find that none of the coefficients on rainfall variables in either HAZ or WAZ regressions are significantly different between the total and non-attriting sample. We test the joint significance of the differences in rainfall coefficients and fail to reject the null hypothesis that coefficients on rainfall variables do not differ across the two subsamples for either HAZ or WAZ regressions. This further strengthens our claim that attrition bias does not affect our results. Therefore, we conclude that bias resulting from selective mortality and overall attrition in the data is not a problem for our analysis.

Table 7: Early-life Rainfall Shock and Childhood Health Outcomes

	Panel A: HAZ		Panel B: WAZ	
	Whole sample	Non-attriting sample	Whole sample	Non-attriting sample
Year 0	0.913	1.348	1.014	1.728
	(0.503)*	(0.635)**	(0.423)**	(0.553)***
year-1	-0.378	-0.324	-0.114	-0.138
	(0.313)	(0.345)	(0.264)	(0.302)
Year-2	0.307	0.282	0.269	0.371
	(0.298)	(0.339)	(0.251)	(0.296)
Year-3	0.113	-0.009	-0.004	0.094
	(0.300)	(0.354)	(0.253)	(0.310)
year+1	0.275	0.763	0.357	0.572
	(0.558)	(0.704)	(0.470)	(0.617)
Year+2	0.025	0.291	0.287	0.466
	(0.609)	(0.754)	(0.513)	(0.659)
Year+3	-0.002	0.659	0.182	0.752
	(0.585)	(0.748)	(0.493)	(0.654)
N	1,274	897	1,271	899

Notes: Standard errors are reported in parenthesis. Asterisks indicate the level of significance: $*p < 0.1; **p < 0.05; ***p < 0.01$

6.5 Additional Robustness Checks

In the previous section, we established that our main results are robust to the sample selection. In this section, we present results from additional robustness checks. First, we estimate the OLS and cluster fixed-effect regression of Equation 6. The results are presented in Table A4 in the appendix. In general, coefficient estimates in OLS and cluster fixed-effect regressions are similar to those obtained from sibling fixed-effect regressions. The main difference is that coefficients on birth-year rainfall are statistically insignificant in years of schooling and asset-index regressions for OLS and cluster fixed-effect regressions.

Second, we carried out a Hausman test to compare the sibling fixed-effect model with a random-effect model. The null hypothesis of a Hausman test is that unobserved family-level factors are uncorrelated with the observed covariates. Rejection of the null hypothesis suggests that only the sibling fixed-effect estimator is consistent. In contrast, failure to reject suggests that both sibling fixed- and random-effect estimators are consistent and that the random effect is efficient. As shown in Table A5 in the appendix, the test statistics reject the null hypothesis for two of the seven regressions estimated. Thus, it is appropriate that we used the fixed-effect model for all regressions, based on consistency.

Third, we tested how sensitive our results are to excluding children from urban areas from our sample. About 11 percent of children in the sample are from the Bukoba urban district. We included these children in the analysis sample because most households in this district practised garden production in 1991-94. As Bengtsson (2010) finds that families residing in the Bukoba urban district are less dependent on the weather, we estimated the original model (sibling fixed-effect regression) after excluding these children from the sample as a check on our results. The results are reported in Table A6 in the appendix. Most of the main results survive.

Fourth, we show that a younger child effect does not drive our results. We included birth order in the main regression to control any resource competition effect due to birth order. In addition to the order, spacing between the births of children potentially affects children's later socioeconomic outcomes due to a younger child effect. To rule out this concern, we show a distribution of the distance between the birth years of siblings in the sample(see Figure A2). The average spacing between siblings included in the regression is 2.5 years, with 80 percent of the siblings having greater than two years. This reinforces that the results are not likely driven by competition for resources due to spacing between siblings.

Finally, we explore the robustness of our results in terms of alternative rainfall specification. In the primary analysis, we focused on the deviation of birth-year rainfall from the long-run village average, since we aim to identify the effect of more "typical" variation in rainfall on adult outcomes. Alternatively, it may also be relevant to focus on the effect of the more extreme rainfall variability. Accordingly, we estimate a specification in which rainfall shock is measured using an indicator for drought as a robustness check. Following Shah and Steinberg (2017), we define drought as 1 if rainfall in a year of birth is less than 75 percent of the long-run village average rainfall. A similar approach is used to define drought in the years before and after birth year. Because only a few (26) sibling pairs have differences in drought exposure, there may not be enough variation for sibling fixed-effect estimation. We instead report random effect results in Table A7 in the appendix. Coefficients on drought exposure in year of birth have the expected (negative) signs and are statistically significant in HAZ, height, years of schooling, and asset index regressions. The result lends support to our main finding.

7 Conclusion

Most previous research on developing countries studies the impact of more extreme types of early-life shocks such as civil war, famine, and pandemics compared to less severe types of shocks. Nevertheless, less extreme shocks such as rainfall shocks that rural households in developing countries commonly face have high policy relevance to the rural population, whose livelihoods heavily depend on rain-fed agriculture. This paper combines historical rainfall data with unique panel data from Kagera Health and Development Survey (KHDS) to examine the long-term effect of early-life rainfall shock on adult health, education, and socioeconomic outcomes of individuals in Tanzania. We control for unobserved family-level and district-level heterogeneity by estimating differences among siblings. Our analysis indicates that rainfall in birth year affects the education and socioeconomic status of children in adulthood. We estimate that a 15 percent increase in rainfall in one's birth year and birth village (relative to average village rainfall) leads children to have 0.21 more years of schooling and live in a household in 2010 that score 0.19 higher on an asset index. Such an increase in birth-year rainfall is also associated with being 30 percentage points more likely to report a rich economic status.

We then explore the relationship between early-life rainfall and childhood nutritional status and find that higher birth-year rainfall leads to significant decreases in height and weight deficits in children. Our estimates show that a 15 percent increase in rainfall in one's birth year and birth village (relative to average village rainfall) improves height-for-age z score by 0.20 and weight-for-age z score by 0.26. Robustness checks show that our findings are robust to sample selection.

Because rainfall shock translates into income shock and food shortage with delay, nutritional deprivation resulting from rainfall shock in a birth year likely occurs during the second year of life. Therefore, we conclude that nutritional deprivation occurring during the second year of life is important in influencing childhood health, adult education, and socioeconomic status. This suggests children are most vulnerable to shocks in the period after weaning from breast milk. This is in contrast to significant importance placed on the time from early gestation to the first six months of life. The absence of a rainfall effect during gestation is likely because pregnant mothers serve as an effective buffer for the developing fetus against shocks. Children are also protected from nutritional deficiency when they are intensively breastfed in the first year of life. In contrast, children just above breastfeeding age are vulnerable to malnutrition because the buffering role of breastfeeding attenuates, and children are not robust to shocks at this age. Overall, our findings suggest that anti-poverty interventions that promote providing nutritional supplements during the postweaning period could have significant long-run payoffs. The findings also underscore the importance of policies that help rural households smooth consumptions.

The literature on the long-term effects of exposure to early-life shocks in developing countries covers a wide range of shocks related to early-life nutrition (see Currie and Vogl 2013 for a recent summary of this literature). The main results in the existing studies fall into three broad categories. The first finds that exposure to shock during gestation has a long-term effect (Thai and Falaris, 2014; Shah and Steinberg, 2017; Chi et al., 2018; Yamashita and Trinh, 2021; Chang et al., 2022). The second finds exposure to shock during early childhood has a long-term effect (Glewwe et al., 2001; Hoddinott and Kinsey, 2001; Alderman et al., 2006; Yamauchi, 2008; Chen and Zhou, 2007; Godoy et al., 2008; Hoddinott et al., 2008; Maccini and Yang, 2009; Alderman et al., 2009; Maluccio et al., 2009; Ampaabeng and Tan, 2013; Nübler et al., 2021). The third finds exposure to shocks in utero and during early childhood has a long-term effect Meng and Qian (2009); Dercon and Porter (2014); Umana-Aponte et al. (2011); Cornwell and Inder (2015); Fitz and League (2020). Our paper adds to the literature on the second category.

One potential concern with our results is that parental response to rainfall shock could constitute part of the estimated reduced-form effect. Parents might adjust intra-household resource allocation in response to early-life shock on their children. We cannot account for this effect in our regression because we do not have adequate measures of parental human-capital investment for each child in our data. Parents could adopt a compensatory strategy or a reinforcing approach by investing relatively more or fewer resources, respectively, in a child who has suffered from early-life shock. Failing to account for parental investment response to rainfall shock in early childhood would underestimate or overestimate the true effect in adulthood depending on how parents compensate or reinforce the initial effect.

Some studies suggest that parents simultaneously make a compensatory investment in health and reinforcing investment in education in response to early-life shocks (Ayalew, 2005; Yi et al., 2015). Suppose parents in our sample follow the same strategy. In that case, shock-affected children may catch up with their unaffected siblings in health but lag in education. This may provide a potential explanation for why we find that rainfall shock affects childhood health but not adulthood health. It may also explain the result that early-life rainfall shock affects adulthood education without affecting adulthood health. Therefore, the reduced-form effect we identified may underestimate the impact of rainfall shock on adulthood health and overestimate the impact on adulthood education.

From a policy perspective, understanding how parents respond to early-life rainfall shock is essential. If parental response to shock is ignored or poorly understood, the effectiveness of early childhood nutritional intervention programs is likely to be undermined. This is because parents can aggravate or reduce the effect of early-life rainfall shock by adjusting the allocation of resources within the family. While there is considerable advancement regarding the impact of early-life shock, there is limited literature that accounts for parental responses to shocks. More research on the impact of early-life shock that accounts for parental responses is needed to develop effective policy responses. We hope future research will shed more light on parental responses to shock in Tanzania, so our results can be considered in its full context.

References

Harold Alderman, Jere R Behrman, Hans-Peter Kohler, John A Maluccio, and Susan Cotts Watkins. Attrition in longitudinal household survey data: some tests for three developing-country samples. *Demographic research*, 5:79–124, 2001.

Harold Alderman, John Hoddinott, and Bill Kinsey. Long term consequences of early childhood malnutrition. *Oxford economic papers*, 58(3):450–474, 2006.

Harold Alderman, Hans Hoogeveen, and Mariacristina Rossi. Preschool nutrition and subsequent schooling attainment: longitudinal evidence from tanzania. *Economic Development and Cultural Change*, 57(2):239–260, 2009.

Douglas Almond. Is the 1918 influenza pandemic over? long-term effects of in utero influenza exposure in the post-1940 us population. *Journal of political Economy*, 114 (4):672–712, 2006.

Samuel K Ampaabeng and Chih Ming Tan. The long-term cognitive consequences of early childhood malnutrition: the case of famine in ghana. *Journal of health economics*, 32(6):1013–1027, 2013.

Elsa V Artadi. Going into labor: Earnings vs. infant survival in rural africa. *Mimeograph, Bocconi University*, 2005.

Tekabe Ayalew. Parental preference, heterogeneity, and human capital inequality. *Economic Development and Cultural Change*, 53(2):381–407, 2005.

David James Purslove Barker. *Mothers, babies, and health in later life*. Elsevier Health Sciences, 1998.

Kathleen Beegle, Joachim De Weerdt, and Stefan Dercon. Kagera health and development survey 2004 basic information document. *The World Bank. www. worldbank. com/lsms/country/kagera2/docs/KHDS2004% 20BID% 20feb06. pdf[accessed March 13, 2007]*, 2006.

Niklas Bengtsson. How responsive is body weight to transitory income changes? evidence from rural tanzania. *Journal of Development Economics*, 92(1):53–61, 2010.

Samuel Bowles. Towards an educational production function. In *Education, income, and human capital*, pages 11–70. NBER, 1970.

Anne Case, Angela Fertig, and Christina Paxson. The lasting impact of childhood health and circumstance. *Journal of health economics*, 24(2):365–389, 2005.

G Chamberlain. Analysis of covariance with qualitative data, the review of economic studies, vo1. 47, no. 1ij f. *Econometrics Issue*, 1, 1980.

Grace Chang, Marta Favara, and Rafael Novella. The origins of cognitive skills and non-cognitive skills: The long-term effect of in-utero rainfall shocks in india. *Economics & Human Biology*, 44:101089, 2022.

Yuyu Chen and Li-An Zhou. The long-term health and economic consequences of the 1959–1961 famine in china. *Journal of health economics*, 26(4):659–681, 2007.

Y Chi, Eduardo Fe, et al. Exposure and contemporaneousness: What can we learn about the effect of drought on children's cognitive development? *Eduardo, Exposure and Contemporaneousness: What Can We Learn about the Effect of Drought on Children's Cognitive Development*, 2018.

Katy Cornwell and Brett Inder. Child health and rainfall in early life. *The Journal of Development Studies*, 51(7):865–880, 2015.

Janet Currie and Tom Vogl. Early-life health and adult circumstance in developing countries. *Annu. Rev. Econ.*, 5(1):1–36, 2013.

J. De Weerdt, K. Beegle, HB. Lilleør, S. Dercon, K. Hirvonen, M. Kirchberger, and S. Krutikova. Kagera health and development survey 2010: Basic information document. rockwool foundation working paper series 46, 2012.

Stefan Dercon and Catherine Porter. Live aid revisited: long-term impacts of the 1984 ethiopian famine on children. *Journal of the European Economic Association*, 12(4): 927–948, 2014.

Erica Field, Omar Robles, and Maximo Torero. Iodine deficiency and schooling attainment in tanzania. *American Economic Journal: Applied Economics*, 1(4):140–69, 2009.

Dylan Fitz and Riley League. The impact of early-life shocks on adult welfare in brazil: Questions of measurement and timing. *Economics & Human Biology*, 37:100843, 2020.

John Fitzgerald, Peter Gottschalk, and Robert A Moffitt. An analysis of sample attrition in panel data: The michigan panel study of income dynamics, 1998.

Ewout Frankema, Kostadis Papaioannou, et al. Rainfall patterns and human settlement in tropical africa and asia compared. did african farmers face greater insecurity? Technical report, CEPR Discussion Papers, 2017.

Paul Glewwe and Elizabeth M King. The impact of early childhood nutritional status on cognitive development: Does the timing of malnutrition matter? *The world bank economic review*, 15(1):81–113, 2001.

Paul Glewwe, Hanan G Jacoby, and Elizabeth M King. Early childhood nutrition and academic achievement: a longitudinal analysis. *Journal of public economics*, 81(3): 345–368, 2001.

Ricardo Godoy, Susan Tanner, Victoria Reyes-García, William R Leonard, Thomas W McDade, Melanie Vento, James Broesch, Ian C Fitzpatrick, Peter Giovannini, Tomás Huanca, et al. The effect of rainfall during gestation and early childhood on adult height in a foraging and horticultural society of the bolivian amazon. *American Journal of Human Biology*, 20(1):23–34, 2008.

Michael Grossman. On the concept of health capital and the demand for health, 80 j. *Pol. Econ*, 223(10.2307):1830580223, 1972.

John Hoddinott and Bill Kinsey. Child growth in the time of drought. *Oxford Bulletin of Economics and statistics*, 63(4):409–436, 2001.

John Hoddinott, John A Maluccio, Jere R Behrman, Rafael Flores, and Reynaldo Martorell. Effect of a nutrition intervention during early childhood on economic productivity in guatemalan adults. *The lancet*, 371(9610):411–416, 2008.

Mary Abihud Lema and Amos E Majule. Impacts of climate change, variability and adaptation strategies on agriculture in semi arid areas of tanzania: The case of manyoni district in singida region, tanzania. *African Journal of Environmental Science and Technology*, 3(8):206–218, 2009.

Youhong Lin, Feng Liu, and Peng Xu. Effects of drought on infant mortality in china. *Health Economics*, 30(2):248–269, 2021.

Sharon Maccini and Dean Yang. Under the weather: Health, schooling, and economic consequences of early-life rainfall. *American Economic Review*, 99(3):1006–26, 2009.

John A Maluccio, John Hoddinott, Jere R Behrman, Reynaldo Martorell, Agnes R Quisumbing, and Aryeh D Stein. The impact of improving nutrition during early childhood on education among guatemalan adults. *The Economic Journal*, 119(537): 734–763, 2009.

Xin Meng and Nancy Qian. The long term consequences of famine on survivors: evidence from a unique natural experiment using china's great famine. Technical report, National Bureau of Economic Research, 2009.

Tanzania National Bureau of Statistics, Dar-es-Salaam. Tanzania demographic and health survey 2004-2005, 2005.

Tanzania National Bureau of Statistics, Dar-es-Salaam. Tanzania demographic and health survey 2010, 2011.

Laura Nübler, Karen Austrian, John A Maluccio, and Jessie Pinchoff. Rainfall shocks, cognitive development and educational attainment among adolescents in a drought-prone region in kenya. *Environment and Development Economics*, 26(5-6):466–487, 2021.

Pedram Rowhani, David B Lobell, Marc Linderman, and Navin Ramankutty. Climate variability and crop production in tanzania. *Agricultural and forest meteorology*, 151 (4):449–460, 2011.

Manisha Shah and Bryce Millett Steinberg. Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital. *Journal of Political Economy*, 125(2):527–561, 2017.

Thuan Q Thai and Evangelos M Falaris. Child schooling, child health, and rainfall shocks: Evidence from rural vietnam. *Journal of Development Studies*, 50(7):1025–1037, 2014.

Marcela Umana-Aponte et al. *Long-term effects of a nutritional shock: the 1980 famine of Karamoja, Uganda.* Centre for Market and Public Organisation, University of Bristol, 2011.

URT. National sample census of agriculture 2007/2008: Regional report – kagera region, 2012a.

URT. Population and housing census: Population distribution by administrative units; key findings, 2012b.

Suhas Pralhad Wani, TK Sreedevi, Johan Rockström, YS Ramakrishna, et al. Rainfed agriculture–past trends and future prospects. *Rainfed agriculture: Unlocking the potential*, 7:1–33, 2009.

Nobuaki Yamashita and Trong-Anh Trinh. Effects of prenatal exposure to abnormal rainfall on cognitive development in vietnam. *Population and Environment*, pages 1–21, 2021.

Futoshi Yamauchi. Early childhood nutrition, schooling, and sibling inequality in a dynamic context: evidence from south africa. *Economic Development and Cultural Change*, 56(3):657–682, 2008.

Junjian Yi, James J Heckman, Junsen Zhang, and Gabriella Conti. Early health shocks, intra-household resource allocation and child outcomes. *The Economic Journal*, 125 (588):F347–F371, 2015.

Mina Zamand and Asma Hyder. Impact of climatic shocks on child human capital: evidence from young lives data. *Environmental hazards*, 15(3):246–268, 2016.

Appendices

Appendix A:

Table A1: Childhood Characteristics by Mortality Status

Variable	Not dead before 2010	Dead before 2010
Age in years	5.9	4.7
Percent of female	49.4	44.1
HAZ score in childhood	-1.1	-1.6
WAZ score in childhood	-1.1	-1.4
Percent with stunting in childhood	25.2	42.2
Percent with wasting in childhood	20.8	30.4
Ln(Per capita household consumption)	12.6	12.4

Table A2: Impact of Rainfall Shock on Probability of Survival until 2010

	only birth year rainfall	only rainfall variables	other controls included
Year 0	0.183	0.075	-0.324
	(0.340)	(0.379)	(0.750)
year-1		0.395	0.283
		(0.306)	(0.427)
Year-2		-0.191	-0.524
		(0.318)	(0.593)
Year-3		0.386	0.074
		(0.310)	(0.472)
year+1		-0.717	-1.923
		(0.477)	(0.874)**
Year+2		0.113	-0.537
		(0.500)	(0.903)
Year+3		-0.530	-1.264
		(0.486)	(0.853)
N	1,293	1,289	1,139

Notes: Standard error are reported in parenthesis. Asterisks indicate the level of significance: $*p < 0.1; **p < 0.05; ***p < 0.01$

Figure A1: Distribution of HAZ Score in Childhood by Sex.

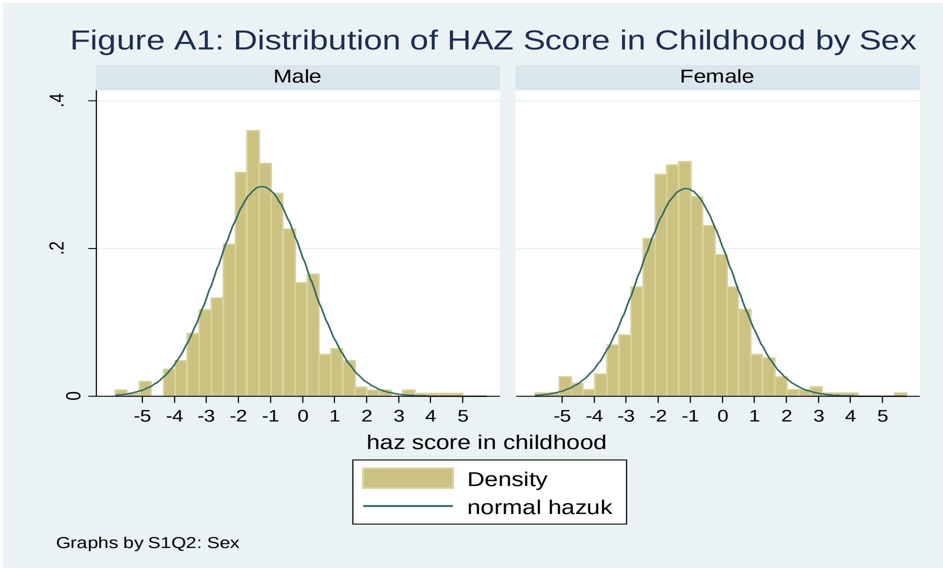


Table A3: Childhood Characteristics by Attrition Status

Variable	Not attriting sample	Attriting sample
Age in years	6.1	5.2
Percent of female	51.5	43.9
HAZ score in childhood	-1.1	-1.2
WAZ score in childhood	-1.1	-1.1
Percent with stunting in childhood	25.6	29.1
Percent with wasting in childhood	21.7	21.5
Ln(Per capita household consumption)	12.5	12.6

Table A4: Effects of Birth Year Rainfall on Childhood and Adult Outcomes

	OLS	Cluster fixed effect
HAZ	1.628	1.495
	(0.495)***	(0.582)**
WAZ	1.753	2.104
	(0.416)***	(0.595)***
Adult Height (centimeters)	-3.873	-4.163
	(3.094)	(4.559)
Ill during the last four weeks	0.881	0.134
	(0.908)	(1.214)
Days ill	1.376	1.081
	(2.325)	(2.318)
Years of schooling	0.098	0.288
	(0.964)	(1.365)
Ln (expenditures per capita in a household)	0.156	-0.096
	(0.257)	(0.352)
Asset index	0.451	0.298
	(0.516)	(0.551)
Self-reported rich status	1.066	2.495
	(0.498)**	(0.853)***

Notes: Standard error are reported in parenthesis. Asterisks indicate the level of significance: $*p < 0.1; **p < 0.05; ***p < 0.01$

Table A5: Hausman test statistics

	Chi square	Prob Chi square
Adult Height (centimeters)	32.16	0.1231
Ill during the last four weeks	28.62	0.2346
Days ill	39.73	0.0228
Years of schooling	33.97	0.0852
Ln (expenditures per capita in a household)	15.60	0.9020
Asset index	24.81	0.4163
Self-reported rich status	11.94	0.9806

Table A6: Effects of Birth Year Rainfall on Childhood and Adult Outcomes –Excluding Urban Sample

HAZ	1.377
	(0.685)**
WAZ	1.654
	(0.593)***
Adult Height (centimeters)	-0.887
	(3.666)
Ill during the last four weeks	-1.288
	(1.277)
Days ill	-4.310
	(2.666)
Years of schooling	1.153
	(0.755)
Ln (expenditures per capita in a household)	0.277
	(0.250)
Asset index	1.653
	(0.515)***
Self-reported rich status	2.007
	(1.032)*

Notes: Standard error are reported in parenthesis. Asterisks indicate the level of significance: $*p < 0.1; **p < 0.05; ***p < 0.01$

Figure A2: Distribution of spacing between the births of siblings in the sample

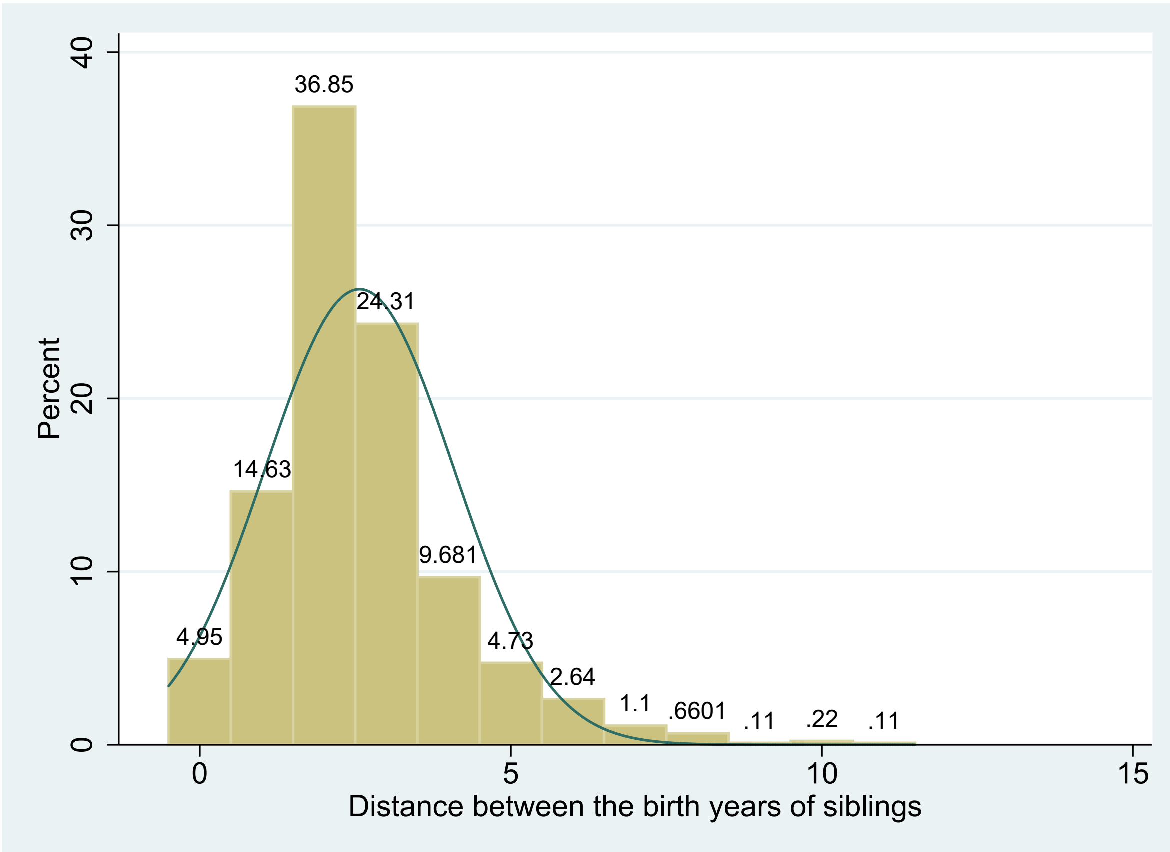


Table A7: Effects of Birth Year Drought on Childhood and Adult Outcomes

HAZ	-0.535
	(0.270)**
WAZ	-0.368
	(0.235)
Adult Height (centimeters)	-3.214
	(1.667)*
Ill during the last four weeks	0.254
	(0.487)
Days ill	0.096
	(1.158)
Years of schooling	-1.573
	(0.541)***
Ln (expenditures per capita in a household)	-0.067
	(0.120)
Asset index	-0.490
	(0.245)**
Self-reported rich status	-0.937
	(0.683)

Notes: Standard error are reported in parenthesis. Asterisks indicate the level of significance: $*p < 0.1; **p < 0.05; ***p < 0.01$

Chapter 3

Emigration and education: the schooling of the left behind in Nigeria

Journal of Migration and Development, 2020

Emigration and education: the schooling of the left behind in Nigeria

Biniam Bedasso *[1], Ermias Gebru Weldesenbet †[2], and Nonso Obikili ‡[3]

[1]Center for Global Development
[2]University of Copenhagen
[3]Economic Research Southern Africa and Stellenbosch University

February 23, 2022

Abstract

The potential effects of migration on the welfare of the left behind consist in an important part of the debate around migration. In this paper, we use household survey data from Nigeria to examine the impact of family migration on educational attainment. Because the migration status of households is endogenous, we use the proportion of migrants in a local district and historical exposure to foreigners as proxied by distance to foreign missionary station in 1921 as instruments for migration of household members. We find that being in a migrant household increases the probability of completing secondary school and attending some postsecondary education. We also find that belonging to a migrant household increases the probability of own future migration. We further explore channels through which migration of family member affects education. We provide tentative evidence suggesting that anticipation of own future migration may be behind increased educational attainment.

Keywords: Children left-behind; Human Capital; Migration; Nigeria.
JEL Code: D12, I21, I25, O15, O55

*bbedasso@cgdev.org
†ermias.weldesenbet@econ.ku.dk
‡me@nonsoobikili.com

1 Introduction

Despite recent political backlash in Europe and the United States, international migration remains a formidable force with wide-ranging consequences in destination and origin countries. The presence of a migrant family member living in a developed country could have multifaceted implications for the welfare of the left behind. This paper attempts to examine the impacts of international migration on the education of family members in Nigeria. Specifically, we study the net effects of the presence of a migrant family member living in a foreign country on the educational attainment of family members back at home at both secondary and postsecondary levels. By conducting such analysis for Nigeria, the most populous country on the African continent, we hope to shed light on the human capital implications of migration in one of the poorest regions in the world. Moreover, the year in which the data were collected, i.e. 2009, marks the culmination of a decade that saw the growth in the stock Nigerian migrants around the world nearly double compared to the decade before.

The link between emigration and education is often viewed from the point of the view of the emigrant without much consideration for spillover effects on those left behind. That is why the issue of migration from developing countries tends to raise the specter of brain drain. This is a legitimate concern given the selectivity involved in attracting skilled migrants from developing countries that do not share borders with developed countries. However, this concern has been countered with the argument that the prospect of migration may lead to a net increase in the stock of human capital of the origin country by encouraging more individuals than who will eventually emigrate to invest in education (Mountford, 1997; Vidal, 1998; Beine et al., 2008). In this regard, the potentially positive impact of migration on the education of the left behind might not be limited to household wellbeing. It could also extend to improving aggregate human capital despite initial loss through brain drain.

The presence of a migrant family member in a foreign, and usually, more developed country could influence the education of the left behind through a number of potential channels. The first potential channel is remittances. Remittances may help to relax the credit constraint that is often behind underinvestment in education in many developing countries. As a second potential channel, the absence of family members from home may affect the education of the left behind negatively by depriving them of proper guidance and role models or burdening them with extra household responsibilities. The third potential channel is the improved probability of future migration of the left behind. The argument for this channel rests on two assumptions. First, the presence of family members and other social networks in destination countries plays an important role in encouraging new emigration to those destinations (Massey and España, 1987; Palloni et al., 2001). Second, on average, there is a higher return to education in destination countries than in origin countries.

In this paper, the primary objective is identifying the net effect of the emigration of a family member on education that may have been transmitted through a variety of channels. But, as a secondary objective, we also examine the role of the prospect of future own migration, inspired by the presence of a family member abroad, as a potential channel mediating the effect of emigration on family education. We focus on the probability of future emigration as a potential channel for two reasons. First, the prospect of future emigration is deemed to affect the expected returns to education more directly than competing channels such as remittances, which exert influence through the budget constraint. Second, as far as our geographical focus is concerned, more migrants from African countries move to the US and the European Union through the help of family reunification programs than any other means, lending credence to the importance of family network for migration (Lucas, 2015).

Although the key propositions we test are rooted in human capital theory, the scope of this study is mainly empirical. We use the completion of secondary school and the attendance of postsecondary education as alternative outcome variables. Completion of secondary schooling has become an important indicator of social progress in developing countries since the expansion of primary schooling in most of these places in recent years has led to an increase in the number of young people eligible for secondary education. Moreover, secondary education is arguably the minimum requirement for most migrants to be able to cope up with life and work in foreign countries[1]. In this regard, we expect that the migration of a family member has a positive impact on the completion of secondary education by the left behind. Postsecondary education, on the other hand, may involve more strategic considerations than secondary schooling. For instance, people might defer investing in postsecondary education in their home countries if they expect to migrate to another country in the near future. But, it could also be the case that postsecondary education increases both the chance of migration and the return to education in prospective destination countries. Thus, the effect of migration on postsecondary education depends on a number of factors, including timing, the perceived probability of migration, and the comparative quality and cost of education in origin and destination countries.

We employ household and individual-level data from the migration and remittances survey conducted by the World Bank in 2009 for the empirical analysis. Making causal inference regarding the link between migration and education requires identifying an exogenous source of variation for migration. We instrument having a migrant household member in a foreign country with the proportion of international migrants from the same town as the respondent and distance to a foreign mission/church in 1921. Accordingly, we use bivariate probit estimation to measure the net effect of migration on the education of the left behind. Once we have estimated the main coefficients of interest, we then test the relevance of the prospect of future migration of the left behind as a channel. This is done by estimating the predicted probability of migration as a function of having a migrant household member in a foreign country. We then estimate the correlation between the predicted probability of future migration and educational attainment.

Firstly, we find a positive and significant impact of having a migrant member of the

[1] For example, the United States Diversity Visa program that grants 50,000 permanent residence visas annually via lottery requires winners to have attained a minimum of 12 years for formal schooling

household on the probability of completing secondary school and attending postsecondary school. This result is robust to various specifications and estimations techniques and is validated by using various exogenous measures of migrant networks as instruments. Secondly, we find that being in a migrant household increases the probability of own future migration. Finally, we find that the probability of own future migration is positively correlated with the probability of completing secondary school or attending postsecondary school. Our results help understand the dynamics and channels through which migration influences human capital development of those left behind.

The existing evidence on the impact of family member's migration on education of children left behind in home country is generally inconclusive (see Antman 2013 and Démurger 2015 for recent reviews). Some studies provide evidence of overall positive effect of migration on education (Hanson and Woodruff, 2003; Edwards and Ureta, 2003; Yang, 2008; Antman, 2012). These studies stress that remittances play a major role in determining children's educational outcomes. Other studies find overall negative effect of migration on education (Giannelli and Mangiavacchi, 2010; McKenzie and Rapoport, 2011; Antman, 2011; Hu, 2013). These studies emphasize that the disruptive effect of parental absence dominates the positive effect of migration through remittance.

Two main reasons might explain the mixed findings in the literature. First, only some of the studies employ careful identification strategies that take in to account the endogeneity of migration. Second, the effects of each channel through which migration affects the education of children left behind may differ depending on each country's specific circumstances. For instance, studies find both negative and positive effects of the prospect of migration channel depending on the source and destination countries analyzed. McKenzie and Rapoport (2011) report that individuals with higher probabilities of migration to the United States invest less in education in Mexico because the return to Mexican education in the United States is lower than in Mexico. In contrast, Batista et al. (2012) find that individual's migration prospects increase the probability of completing intermediate secondary school in the case of Cape Verde. Most of the existing studies in the literature cover Latin American countries, especially Mexico and some Asian countries. There is limited work on the impact of migration on the education of children left behind in sub-Saharan Africa, although it is a major migrant-sending region in the world. Our paper contributes to understanding the human capital implication of migration in the region.

The remainder of this paper is organized as follows. Section 2 presents the conceptual framework of the paper. Section 3 discusses the empirical strategy. Section 4 describes the data used for the analysis, and section 5 discusses the results. Finally, section 6 concludes.

2 Conceptual framework

The link between emigration of family members and education of the left behind is anchored in two separate theoretical frameworks. One is human capital theory of education while the other is the social capital theory of migration. The human capital theory of education characterizes schooling decisions as one of investment, which is a function of future returns, opportunity cost and direct costs. Generally, schooling decision can be specified as follows:

$$S_i = s(R_i, C_i, Z_i) \tag{1}$$

where S is years of schooling, R is expected returns, C is expected cost and Z is a vector of other factors that influence schooling decision.Individuals face heterogeneous returns to education depending on, among other things, their level of access to high-return foreign labor markets. This assertion is based on the assumption that global labor markets are segmented according to wealth into high and low return markets. In the presence of credit market imperfections, individuals also face heterogeneous direct costs, which may depend on their endowments and other flows such as remittances. Finally, nonpecuniary factors including environmental conditions such as role models influence education. In this regard, the physical absence of family members due to migration can reduce their use as role models to individuals who attend school in the country of origin. Thus, we can claim that expected own migration or current migration of other family members can potentially affect schooling decision through one or more of the constructs identified in equation 1.

Among the two dimensions of migration identified above as consequential to educational outcomes, expected own migration can be predicted using the presence of migrant family members in a foreign country. This is where the social capital theory of migration comes into play. According to social capital theory, migrant networks exist due to interpersonal ties between potential migrants and previous migrants occurring in the form of kinship, friendship, and shared community origin. Migrant networks "increase the likelihood of international movement because they lower the costs and risks of movement and increase the expected net returns to migration" (Massey 1988, pp. 42-3). In this light, the presence of a migrant family member in a foreign country can be expected to have a significant impact on the probability of own future migration.

Assuming that migrants can acquire schooling in foreign countries, the level of schooling individuals with a non-zero probability of migration attain in the country of origin depends on the following considerations. On the one hand, education acquired in country of origin may increase the probability of migration itself. On the other hand, education obtained in destination country may be more relevant and worthy of higher return than education obtained in country of origin. If the latter consideration outweighs the former, we can expect the presence of a migrant family member in a foreign country to have a negative effect on the education of the left behind by encouraging them to defer investing in education until they emigrate. On the contrary, if the gain in the probability of migration from domestic education outweighs the return differential due to foreign education, the presence of a migrant family member in a foreign country encourages immediate investment in education in the home country.

Considering all direct and indirect channels, the net effect of the presence of a migrant family member in a foreign country on the education of the left behind can be either positive or negative depending on the magnitude of competing effects. In the case of the current study, we expect the emigration of a family member to have a positive effect on the completion of secondary school. The potentially negative impact of emigration that could occur via parental absence is minimized because temporary labor migration, which is often the reason for parental absence, is not common in Nigeria. Moreover, deferment of education until prospective emigration is unlikely to happen for basic education at a young age. In the case of postsecondary education as well, we expect the positive effects of remittances and expected own migration to outweigh the negative impacts of deferment and parental absence.

3 Empirical strategy

We begin formulating our empirical strategy by specifying a reduced form of the human capital investment function laid out in equation 1,

$$S_i = \alpha_0 + \alpha_1 M_i + \alpha_2' X_i + \varepsilon_i \tag{2}$$

where S is either completion of secondary schooling or attainment of postsecondary education, M_i is an indicator variable for having a migrant in the household, X_i is a vector of control variables that often determine the demand for education. ε is a random disturbance term. α_i captures the net effect of having a migrant in the household on education of the left behind. Primarily, we estimate equation 2 to reveal this effect. Secondarily, we attempt to explicate the impact of the probability of future own emigration as a channel. Accordingly, expected own emigration is written as the following latent variable,

$$Q_i^* = \beta_0 + \beta_1 M_i + \beta_2 S_i^* + \zeta_i \tag{3}$$

where ζ is a random disturbance term. Expected own emigration is a function of having a migrant in the household, M, and expected level of schooling, S^*. Finally, equation 2 can be written as a function of expected own migration as follows,

$$S_i = \gamma_0 + \gamma_1 Q_i^* + \gamma_2' X_i + \varsigma_i \tag{4}$$

where ς is a random disturbance term. The above simultaneous equation system suggests that the portion of expected own migration that is explained by the presence of a migrant household member in a foreign country can be used to predict level of schooling. Accordingly, in order to determine the effect of expected own emigration as a channel, we predict the latent probability of migration using equation 3. We then use those results in equation 4 to estimate the impact of family emigration as channeled by the prospect of own migration.

In this paper, we argue that expected own migration, indirectly captured by the presence of migrant family members in a foreign country, influences the level of schooling. It is however likely that the presence of migrant family members in a foreign country is endogenous, i.e. both migration and education in a given family could be influenced by a common unobserved factor. For instance, family members such as parents and elder siblings may decide to migrate because they wish to improve the educational opportunities of younger members of the family through remittances and future migration to developed countries with better schools and universities. In this case, both migration and education of the left behind are determined by family-level preference for better human capital.

This problem of endogeneity poses a challenge for our estimation strategy. To deal with this problem, we argue that migrant networks could act as a source for exogenous variation in the probability of migration. The migrant networks would however not directly influence the probability of any given person's level of schooling.

The idea that migrant networks influence an individual's probability of migration has been established in the literature. Theoretically, the presence of migrant networks reduces the costs of migration. Having friends or family members who have successfully migrated implies that new migrants have contacts who will guide and help them settle more easily. The reduced cost of migration, all things being equal, should increase the probability of migration. This idea of the impact of migrant networks, and its usefulness as a source of exogenous variation, is not unique to this paper and has been used in other studies examining various impacts of migration. Batista et al. (2012) use the length of migration of a family member, arguing that longer history of migration implies deeper access to migrant networks and lower migration costs. Woodruff and Zenteno (2007) and McKenzie and Rapoport (2011) on the other hand use the historic level of migration in a given geographical area as proxy for migrant networks.

In this paper, we use two separate proxies to capture the strength of migrant network to which any given individual is exposed to. First, we use the proportion of international migrants in the town in which the individual is located. This instrument is similar to that used by Woodruff and Zenteno (2007). Towns with a larger proportion of migrant family members are assumed to have deeper migrant networks. The variation in the proportion of migrant family members is however due to historic factors not influenced by an individual's level of schooling. A possible concern with this instrument is that having own household in the computation of town level proportion makes the instrument endogenous. In order to address this concern, we exclude own household from the computation of proportion of migrants in town.

Our second instrument uses one of these historic factors as a proxy for deeper migrant networks. Specifically, we use the distance from the town in which the migrant is located to the nearest missionary station in 1921. Missionary stations were typically associated with Europeans who presumably still maintained networks to their home countries. This maintenance of links to home countries is highlighted by one of the key factors behind the decisions on where to locate missions, the ability to import supplies from Europe (Johnson, 1967). According to Johnson (1967), mission station locations were influenced by previous mission stations that provided information on surrounding areas. The result was a network of stations that together formed a series of transshipment points from the coast to the interior (Nunn 2010, pg. 148). Missions with these networks to their home countries implied that people living closer to these missions were historically much more likely to tap into these networks. This would imply lower costs of migration and therefore a higher probability of historical migration which in turn predicts larger present migration.

Although one could argue that the level of schooling is intergenerational, and therefore historic factors that influence migration would also influence the level of schooling. For instance, the work of missionaries in Nigeria was directly linked to their distribution of European education (Horton, 1971). To deal with this threat, we control for the level of education of the father and mother of the individual. Controlling for the level of education of the parents can potentially cancel out the possible influence of historic factors directly influencing the level of schooling of the individual. We also control for the relative level of development in the area, proxied by the average intensity of night light in the local government area where the individual is located, and the number of schools relative to the population of the state. This should capture all the town-level effect that the distance to the historic mission may have on the individuals' level of schooling.

4 Data

The study uses data from the Migration and Remittance Household Survey (MRHS) conducted by the World Bank in 2009. Although the MRHS was designed to shed light on various aspects of migration and remittances, it also collected information on a variety of demographic, social, and economic characteristics of all household members, including emigrants. The MRHS interviewed a nationally representative sample of 2251 households. The households were selected using multistage stratified random sampling from 17 states and 93 towns. We restrict the sample for the study to children of household head aged 16 to 30 years who were living in the household in 2009. The sample consists of 2628 children living in 1156 households.

Migration and education are the key variables of interest in our study. Regarding migration, the MRHS asked all households if they currently have a member living outside of Nigeria. Generally, surveys on migration tend to ask whether migration of household members occurred within the last five years prior to the survey. As opposed to other similar surveys, the MRHS captures information on migrants even if they left home a long time back. This adds to the suitability of our data for examining the impact of migration on education considering that the impact of migration on education persists over a long period of time. We define a child as living in a migrant household if the household has at least one international migrant family member who had departed before the child left behind completed schooling. Summary statistics for key variables are presented in Table 1. About 30 percent of children in our sample were living in a migrant household, ensuring enough variation in the explanatory variable.

The education system in Nigeria involves 6 years of primary education, 3 years of junior secondary education and 3 years of senior secondary education. Majority of Nigerians begin primary education at the age of 4. Primary education (grades 1 to 6) is free and compulsory since 1976. Junior secondary education (grades 7 to 9) also became compulsory with the introduction of the Universal Basic Education (UBE) program in 1999. We choose 16 as the starting age for our sample because most students in Nigeria normally complete secondary school from age 16 onward. Around 38 percent of 16-year-old children in our sample completed secondary school.

Education is measured using the number of years required to complete the highest grade attained. We use data on years of schooling to define two binary indicators showing whether the child completed secondary school and whether the child attained postsecondary school. Table 1 shows that children living in a migrant household are more likely to complete secondary school and attain postsecondary school. Eighty percent of children living in migrant households completed secondary school compared to 60 percent in non-migrant households. Thirty-two percent of children living in migrant households attained postsecondary school compared to 14 percent in non-migrant households. Likewise, children living in migrant households have higher mean schooling than children in non-migrant households (12.5 vs. 10). Given that the differences are statistically significant, it suggests that there is a positive association between having a migrant in the household and education of children. This is also corroborated by the fact that the coefficient of correlation between being in a migrant household and years of schooling of children is 0.3 (p-value=0.000). The empirical analysis sheds light on whether this relationship holds after we take into account the endogenous nature of migration behavior and control for various observable characteristics which also affect schooling.

Table 1 also presents descriptive statistics for an array of control variables. We use an indicator for female, father's years of schooling, mother's years of schooling, number of children in the household, number of schools per 1000 population, and average night light intensity as controls. While most of the control variables are taken from MRHS, number of schools, night light intensity and distance to missionary station data come from secondary data sources. Women make up 38 percent of the sample. Average age of the sample is 21.5 years. Average father's and mother's years of schooling are 8.5 and 7.7 respectively. Children living in migrant households have more educated parents than children in non-migrant households. This could be seen as an indication of family preference for education being a potential source of endogeneity. Looking at the instruments, households with migrant are located in towns with higher proportion of migrants and close to missionary stations relative to households without migrant. This suggests that both proportion of migrants in the town and distance to missionary station are likely to perform well as instruments for migration.

Our main sample discussed above is restricted to only household members who still lived at home in 2009. The analysis of channel of migration, however, requires both current household members and migrant children in order to estimate the effect of being in a migrant household on the probability of own migration. Accordingly, we add migrant children of head who were aged 16 to 30 years prior to their migration

Table 1: Summary statistics of key variables

	Number of observations	Non-migrant households	Migrant households	All households
Household variables				
Proportion of households with a migrant	2628	0	1	0.307
Number of children aged between 8 and 30 years	2628	5.018	4.844	4.965
		(2.773)	(2.904)	(2.814)
Individual variables				
Proportion of female	2625	0.372	0.404	0.382
Age in years	2628	21.27	22.23	21.57
		(4.205)	(4.160)	(4.213)
Years of schooling	2628	10.09	12.54	10.84
		(4.494)	(3.122)	(4.273)
Proportion of children who completed secondary school	2628	0.605	0.808	0.667
Proportion of children who attained post-secondary school	2585	0.140	0.324	0.197
Father´s years of schooling	2293	7.897	10.25	8.587
		(6.134)	(6.121)	(6.222)
Mother´s years of schooling	2537	6.931	9.542	7.731
		(6.021)	(5.939)	(6.115)
District variables				
Proportion of migrant in the district	2575	0.0612	0.104	0.0746
Distance to church	2628	0.326	0.151	0.272
		(0.420)	(0.187)	(0.373)
Average night light intensity for the period 2011-2012	2628	23.05	34.06	26.43
		(31.48)	(34.69)	(32.89)
State variable				
Number of schools per 1000 population	2628	0.491	0.451	0.479
		(0.157)	(0.145)	(0.154)

Notes. Sample is children of head aged between 16 and 30 years. Standard deviations are in parenthesis.

into our main sample. After dropping observations with missing data, our sample for the estimation of channel of migration consists of 2796 children from 1030 households. The summary statistics show that around 6 percent of children of head aged 16 to 30 years in the sample are international migrants. Higher proportion of migrant children completed secondary school and attained some postsecondary education prior to their migration relative to non-migrant children.

5 Empirical results

This section reports the empirical estimates on the relationships hypothesized in the conceptual framework. While we present estimates of different models for the purpose of comparison, we mainly interpret results of bivariate probit model, which is better suited for our purpose. We begin with documenting the baseline association between having a migrant in the household, and the probability of completing secondary school as in equation 3.1. Accordingly, Table 2 reports OLS results using different specifications and subsamples. Columns (1) shows that there is a positive and statistically significant relationship between being in a migrant household and completing secondary school. In column (2), we control for sex, age, father's education, mother's education, and number of children in the household. These controls are typically regarded as being important for education outcomes. We also control for the number of schools per 1000 people in the state where the household is located and the average night lights intensity of the local government area where the household is located. Both variables help capture some of the aggregate differences in level of development and schooling resources across areas which should also influence educational outcomes. Finally, we control for age group fixed effects to account for some of the differences across cohorts. The results show that the positive relationship holds and remains significant despite these controls.

In columns (3) and (4), we restrict the sample to male and female individuals respectively. We do this to ensure that the relationship we find is not driven by gender-specific factors that we have not captured. In both cases, the relationship between being in a

Table 2: Impact of being in a migrant household on probability of completing secondary school. OLS estimates.

	(1)	(2)	(3)	(4)	(5)
Child is in a migrant household	0.203	0.116	0.141	0.074	0.117
	(0.045)***	(0.040)***	(0.043)***	(0.043)*	(0.040)***
Controls included	No	Yes	Yes	Yes	Yes
R^2	0.04	0.25	0.29	0.19	0.25
Observation*s*	2,628	2,216	1,389	827	2,214

Notes. Sample in columns (1) and (2) is children of head aged between 16 and 30 years. Sample in column (3) is restricted to male children. Sample in column (4) is restricted to female children. Sample in column (5) is restricted to children whose parents were both not migrants. Controls include sex, age, age group fixed effect, father's education, mother's education, number of children, number of schools per 1000 population and local average night light intensity. Standard errors clustered at district level. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

migrant household and the probability of completing secondary school remains positive and significant. However, the relationship is more robust for male children than for females. Finally, in column (5), we restrict the sample to only individuals neither of their parents are migrants. This is done to rule out the effect of parental absence on educational outcomes. The result again shows a strong positive effect of being in a migrant household on the probability of completing secondary school.

The results in Table 3 suggest that there is a positive correlation between being in a migrant household and the probability of finishing secondary school. However, as discussed earlier, it is also possible that there are unobserved variables which influence both the probability of being in a migrant household and the probability of finishing secondary school. In Table 3, we use two indicators of exposure to migrant networks as instruments. Specifically, column (1) uses the proportion of migrants in the district as an instrument, and column (2) uses the distance to a foreign mission in 1921. The 2SLS estimates confirm that being in a migrant household does increase the probability of completing secondary school. Using distance to foreign missionary station in 1921 as an instrument produces higher point estimates than using proportion of migrants in the district. Comparison of OLS and 2SLS results shows the impact of being in migrant household on probability of completing secondary school becomes stronger once we instrument for migration. The F-statistics indicate that the excluded instruments explain

Table 3: Impact of being in a migrant household on probability of completing secondary school. 2SLS and bivariate probit estimates.

	2SLS	2SLS	Bivariate probit	Bivariate probit
Child is in a migrant household	0.747 (0.216)***	1.496 (0.522)***	1.347 (0.195)***	1.421 (0.224)***
Marginal effect			0.167 (0.033)***	0.168 (0.030)***
Instrument	A	B	A	B
P-value of endogeneity test	0.01	0.00		
F-statistic on excluded instrument	26.8	13.8		
Observations	2,165	2,216	2,165	2,216

Notes. Instrumental variable A: proportion of migrant in the district. Instrumental variable B: distance to foreign missionary station in 1921. All regressions include sex, age, age group fixed effect, father's education, mother's education, number of children, number of schools per 1000 population and local average night light intensity as controls. Sample is children of head aged between 16 and 30 years. Standard errors clustered at district level. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

variation in migration of family members at conventional levels. Columns (3) and (4) in Table 3 mirror columns (1) and (2) but make use of a bivariate probit estimator due to the binary nature of both the dependent and endogenous variables. We compute marginal effect of change in household migration status to identify the magnitude of the effect. Depending on the instrument used, living in a migrant household increases the probability of completing secondary school by 14-17

The results above show that the probability of completing secondary school is higher in migrant households. However, the impact of family migration on education may extend to postsecondary schooling. In Table (4), we show that the positive relationship also holds for the probability of attending some postsecondary education. Table 4 mimics table 3 in that we use exposure to migrant networks to test the effect of being in a migrant household on the probability of attending postsecondary school. Generally, there is a positive and significant relationship between family migration and the probability of attending some postsecondary schooling. However, the results are not as robust as the ones for secondary schooling. This suggests that the impact of family migration on higher levels of education might not be as straight forward as the impact on basic education. The magnitude of the marginal effect is also much smaller in the

Table 4: Impact of being in a migrant household on probability of attending post-secondary school.

	OLS	2SLS	2SLS	Bivariate probit	Bivariate probit
Child is in a migrant household	0.145	0.474	0.250	1.672	1.133
	(0.032)***	(0.127)***	(0.157)	(0.173)***	(0.539)**
Marginal effect				0.009	0.028
				(0.006)	(0.009)***
Instruments		A	B	A	B
Observations	2,180	2,129	2,180	2,129	2,180

Notes. Instrumental variable A: proportion of migrant in the district. Instrumental variable B: distance to foreign missionary station in 1921. All regressions include sex, age, age group fixed effect, father's education, mother's education, number of children, number of schools per 1000 population and local average night light intensity as controls. Sample is children of head aged between 16 and 30 years. Standard errors clustered at district level. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

case of postsecondary schooling. Being in a migrant household increases the probability of attending postsecondary school by less than 3 percent.

The above results show the overall impact of family migration on education outcomes. We now move on to exploring the role of expected own migration as an intermediate factor channeling the impact of family migration. The working hypothesis is that expected own migration could serve as the motivation for the increase in educational attainment. We test this hypothesis using a two-step approach. First, we pool the samples of migrants and non-migrants to estimate the impact of being in a migrant household on the probability of own migration as in equation 3.2. Second, we use the predicted probabilities of own migration from the previous estimation to estimate the effect of expected own migration on educational outcomes as in equation 3.3.

Table 5 presents the results of the first-stage estimation linking family migration to own migration. As expected, there is a positive relationship between the presence of a migrant family member in a foreign country and the probability of own migration. Since the same concern of endogeneity as in earlier estimations pervades the relationship between family migration and own migration, we apply the two instruments we used before jointly in columns (3) and (4). The impact of family migration on own migration is more robust when instruments are used.

Table 5: Impact of being in a migrant household on probability of own future migration

	OLS	Probit	2SLS	Bivariate probit
Child is in a migrant household	0.031	0.252	0.215	1.114
	(0.018)*	(0.130)*	(0.085)**	(0.364)***
Instruments			C	C
Observations	2,796	2,796	2,745	2,745

Notes.Instrumental variable set C: proportion of migrant in the district + distance to foreign missionary station in 1921. All regressions include sex, age, age group fixed effect, years of schooling, father's education, mother's education, number of children, number of schools per 1000-population and local average night light intensity as controls. Sample is children of head (both household members and migrants) aged between 16 and 30 years. Standard errors clustered at district level. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 6: Impact of own future migration on probability of completing secondary school

	OLS	Probit	OLS	Probit
Probability of own future migration	0.398	1.604	0.303	1.416
	(0.135)***	(0.522)***	(0.031)***	(0.245)***
R^2	0.25		0.43	
Observations	2,216	2,216	2,141	2,141

Notes.Columns (1) and (2) use predicted probability of own future migration from 2SLS model as the main regressor. Columns (3) and (4) use predicted probability of own future migration from bivariate probit model as the main regressor. All regressions include sex, age, age group fixed effect, father's education, mother's education, number of children, number of schools per 1000 population and local average night light intensity as controls. Sample is children of head (both household members and migrants) aged between 16 and 30 years. Standard errors clustered at district level. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

We calculate the predicted probabilities of own migration for the sample of migrant and non-migrant individuals using coefficients presented in columns (3) and (4) in Table 5. Table 6 estimates the effect of expected own migration, which already takes account of family migration through the aforementioned estimation, on the probability of completing secondary school. Table 7 presents the corresponding results in the case of postsecondary schooling. All results show that expected own migration exerts positive and significant effect on educational outcomes.

Table 7: Impact of own future migration on probability of attending postsecondary school.

	OLS	Probit	OLS	Probit
Probability of own future migration	0.496	2.193	0.193	1.107
	(0.111)***	(0.414)***	(0.020)***	(0.100)***
R^2	0.30		0.38	
Observations	2,180	2,180	2,105	2,105

Notes.Columns (1) and (2) use predicted probability of own future migration from 2SLS model as the main regressor. Columns (3) and (4) use predicted probability of own future migration from bivariate probit model as the main regressor. All regressions include sex, age, age group fixed effect, father's education, mother's education, number of children, number of schools per 1000 population and local average night light intensity as controls. Sample is children of head (both household members and migrants) aged between 16 and 30 years. Standard errors clustered at district level. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

6 Robustness tests

The above analysis has found that being in a migrant household increases the probability of completing secondary school and attending some postsecondary education. In this section, we explore whether our results are robust to alternative specifications. First, we test how sensitive our results are to including household wealth in our regression. Since household resources net of subsistence needs (available for education of children) vary with household wealth, it is important to control for household wealth in our regression. Unfortunately, household wealth reported in our data is likely affected by remittances received from migrant family members. When the reported wealth is influenced by remittances, controlling for household wealth partly takes away some of the effects due to the presence of migrant family member in a foreign country. So, we opted to exclude household wealth from our regression in the main analysis in order to avoid underestimating the effect of having migrant family member. In this section, we extend the analysis by including household wealth in our regression.

We constructed household wealth index using principal component analysis. Variables used to construct the wealth index include whether a household owns a dwelling unit, type of dwelling, major construction material of exterior walls, existence of separate room for cooking and household members per room. We included these variables

Table 8: Impact of being in a migrant household on probability of completing secondary school using household wealth index as additional control.

	2SLS	2SLS	Bivariate probit	Bivariate probit
Child is in a migrant household	0.701	2.088	1.093	0.978
	(0.312)**	(1.078)*	(0.308)***	(0.344)***
Instrument	A	B	A	B
Observations	2,094	2,141	2,094	2,141

Notes.Instrumental variable A: proportion of migrant in the district. Instrumental variable B: distance to foreign missionary station in 1921. All regressions include sex, age, age group fixed effect, father's education, mother's education, number of children, number of schools per 1000 population, local average night light intensity and wealth index as controls. Sample is children of head aged between 16 and 30 years. Standard errors clustered at district level. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 9: Impact of being in a migrant household on probability of attending postsecondary school using household wealth index as additional control.

	2SLS	2SLS	Bivariate probit	Bivariate probit
Child is in a migrant household	0.536	0.057	1.586	-0.089
	(0.187)***	(0.308)	(0.177)***	(1.564)
Instrument	A	B	A	B
Observations	2,058	2,105	2,058	2,105

HNotes. Instrumental variable A: proportion of migrant in the district. Instrumental variable B: distance to foreign missionary station in 1921. All regressions include sex, age, age group fixed effect, father's education, mother's education, number of children, number of schools per 1000 population, local average night light intensity and wealth index as controls. Sample is children of head aged between 16 and 30 years. Standard errors clustered at district level. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

in the principal component analysis because they are less likely to be affected by remittances. Only 5 percent of the households reported that they spent remittances received from family members on building new house or rebuilding house.

Table 8 reports coefficient estimates for probability of completing secondary school when we include household wealth as an additional control. Column (1) and (2) report two-stage least square results, while column (3) and (4) present results from bivariate probit regressions. Column (1) and (3) use the proportion of migrants in the district as an instrument, and column (2) and (4) use the distance to foreign missionary station in 1921. All specifications show a strong positive effect of being in a migrant household on the probability of completing secondary school, echoing the result in the main analysis. The only difference compared to the baseline regression presented in Table 3 is that coefficient estimates are now reduced in size in three of the four cases. The corresponding result for the probability of attending some postsecondary education is presented in Table 9. The results reveal that being in a migrant household has a statistically significant positive effect on probability of attending some postsecondary education only in regressions using proportion of migrants in the district as instrument. Therefore, it appears that most of our results survive even after controlling for household wealth.

Second, we explore robustness of our result to using different age groups. Our sample in the main analysis is children of household head aged 16 to 30 years in 2009. As mentioned earlier, we chose this age group because children in Nigeria complete secondary school from age 16 onward, and lately enrolled children could still be attending some postsecondary education until the age of 30. It could be argued that the age range of our sample may influence the estimated coefficients. In order to address this concern, we restrict our sample to a narrower age cohort. For probability of completing secondary school, we limit our sample to children of head aged 16 to 20 years. Table 10 reports results of estimating the impact of being in a migrant household on probability of completing secondary school. It can be seen that results in Table 10 remain the same as in the baseline specifications (Table 3). Similarly, we estimate the impact of being in migrant household on probability of attending some postsecondary education using children aged 20 to 25 years as shown in Table 11. We find that the coefficients on migration are statistically significant in three of the four specifications. Overall, our result on the relationship between being in a migrant household and probability of completing secondary school is significantly robust, whereas the result for probability of attending some postsecondary education is only partially robust. Therefore, we conclude that the robustness checks further strengthen our previous findings.

Table 10: Impact of being in a migrant household on probability of completing secondary school of children aged 16 to 20.

	2SLS	2SLS	Bivariate probit	Bivariate probit
Child is in a migrant household	1.119	1.200	1.356	1.059
	(0.315)***	(0.465)***	(0.212)***	(0.405)***
Instrument	A	B	A	B
Observations	1,098	1,122	1,098	1,122

Notes. Instrumental variable A: proportion of migrant in the district. Instrumental variable B: distance to foreign missionary station in 1921. All regressions include sex, age, age group fixed effect, father's education, mother's education, number of children, number of schools per 1000 population, and local average night light intensity as controls. Sample is children of head aged between 16 and 20 years. Standard errors clustered at district level. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

Table 11: Impact of being in a migrant household on probability of attending postsecondary school of children aged 20 to 25.

	2SLS	2SLS	Bivariate probit	Bivariate probit
Child is in a migrant household	0.842	0.501	1.887	1.136
	(0.249)***	(0.207)**	(0.147)***	(0.818)
Instrument	A	B	A	B
Observations	818	839	818	839

Notes.Instrumental variable A: proportion of migrant in the district. Instrumental variable B: distance to foreign missionary station in 1921. All regressions include sex, age, age group fixed effect, father's education, mother's education, number of children, number of schools per 1000 population, and local average night light intensity as controls. Sample is children of head aged between 20 and 25 years. Standard errors clustered at district level. * Significant at 10%; ** Significant at 5%; *** Significant at 1%.

7 Conclusion

The welfare implications of migration are expected to be significantly different when the impact on family members in origin countries is accounted for than when only the migrants themselves are considered. This paper has attempted to assess the impact of family migration on the education of the left behind at secondary and postsecondary levels in Nigeria. We find that being in a migrant household increases the probability of completing secondary school and attending some postsecondary education. We employed instrumental variables to show that the estimated relationship is not spurious. In terms of the channels through which the migration of family members increases educational attainment, we find suggestive evidence that the probability of own future migration could be playing a positive role.

This paper has furnished an important piece of evidence illuminating the relationship between migration and household welfare in a major African country. However, so many unanswered questions could be better investigated with more data at the household and local levels. For instance, the tentative conclusion regarding the positive impact of family migration on postsecondary education could be enunciated more clearly if the data allowed disentangling the theoretically identified competing effects of expected own migration. In this regard, there is a lot more room for future research to elucidate the many dimensions of the relationship between emigration and the education of the left behind.

References

Francisca M Antman. The intergenerational effects of paternal migration on schooling and work: What can we learn from children's time allocations? *Journal of Development Economics*, 96(2):200–208, 2011.

Francisca M Antman. Gender, educational attainment, and the impact of parental migration on children left behind. *Journal of Population Economics*, 25(4):1187–1214, 2012.

Francisca M Antman. The impact of migration on family left behind. In *International handbook on the economics of migration*. Edward Elgar Publishing, 2013.

Catia Batista, Aitor Lacuesta, and Pedro C Vicente. Testing the 'brain gain'hypothesis: Micro evidence from cape verde. *Journal of Development Economics*, 97(1):32–45, 2012.

Michel Beine, Frederic Docquier, and Hillel Rapoport. Brain drain and human capital formation in developing countries: winners and losers. *The Economic Journal*, 118 (528):631–652, 2008.

Sylvie Démurger. Migration and families left behind. *IZA World of Labor*, 2015.

Alejandra Cox Edwards and Manuelita Ureta. International migration, remittances, and schooling: evidence from el salvador. *Journal of development economics*, 72(2): 429–461, 2003.

Gianna Claudia Giannelli and Lucia Mangiavacchi. Children's schooling and parental migration: Empirical evidence on the 'left-behind'generation in albania. *Labour*, 24: 76–92, 2010.

Gordon H Hanson and Christopher Woodruff. Emigration and educational attainment in mexico. Technical report, Mimeo., University of California at San Diego, 2003.

Robin Horton. French summary of 'african conversion'. *Africa*, 41(3):244–245, 1971.

Feng Hu. Does migration benefit the schooling of children left behind? evidence from rural northwest china. *Demographic Research*, 29:33–70, 2013.

Hildegard Binder Johnson. The location of christian missions in africa. *Geographical review*, pages 168–202, 1967.

R. E. B. Lucas. "african migration". In *Barry R. Chiswick, Paul W. Miller (Editors), Handbook of the Economics of International Migration*, volume 1. North-Holland, 2015.

D. Massey. Economic development and international migration in comparative perspective. *Population and Development Review*, 14(3):383–413, 1988.

Douglas S Massey and Felipe García España. The social process of international migration. *Science*, 237(4816):733–738, 1987.

David McKenzie and Hillel Rapoport. Can migration reduce educational attainment? evidence from mexico. *Journal of Population Economics*, 24(4):1331–1358, 2011.

Andrew Mountford. Can a brain drain be good for growth in the source economy? *Journal of development economics*, 53(2):287–303, 1997.

N. Nunn. Economic development and international migration in comparative perspective. *American Economic Review Papers and Proceedings*, 100(2):147–152, 2010.

Alberto Palloni, Douglas S Massey, Miguel Ceballos, Kristin Espinosa, and Michael Spittel. Social capital and international migration: A test using information on family networks. *American journal of sociology*, 106(5):1262–1298, 2001.

Jean-Pierre Vidal. The effect of emigration on human capital formation. *Journal of population economics*, 11(4):589–600, 1998.

Christopher Woodruff and Rene Zenteno. Migration networks and microenterprises in mexico. *Journal of development economics*, 82(2):509–528, 2007.

Dean Yang. International migration, remittances and household investment: Evidence from philippine migrants' exchange rate shocks. *The Economic Journal*, 118(528): 591–630, 2008.