



**HAL**  
open science

## Three essays on the equity of access to education in sub-saharan Africa

Esther Delesalle

► **To cite this version:**

Esther Delesalle. Three essays on the equity of access to education in sub-saharan Africa. Economics and Finance. Université de Cergy Pontoise, 2018. English. NNT : 2018CERG0964 . tel-02284329

**HAL Id: tel-02284329**

**<https://theses.hal.science/tel-02284329>**

Submitted on 11 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

## THREE ESSAYS ON EDUCATION IN RURAL TANZANIA

### Thèse de doctorat

délivrée par

l'Université de Cergy-Pontoise

Théorie Économique, Modélisation et Applications (THEMA), CNRS UMR 8184,

préparée à

l'Université de Cergy-Pontoise et à DIAL-IRD

présentée et soutenue publiquement le 16 février 2018 par

**Esther DELESALLE**

### Jury

Arnaud LEFRANC,	Professeur, à, Université de Cergy-Pontoise, IZA	Directeur
Flore GUBERT,	Directrice de recherche, à, IRD-DIAL, PSE	Co-directrice
Pierre ANDRE,	Maître de conférences, à, Université de Cergy-Pontoise	Co-encadrant
Philippe DE VREYER,	Professeur, à, Université Paris-Dauphine	Rapporteur
Paul W. GLEWWE,	Professeur, à, Université du Minnesota	Rapporteur
Marguerita COMOLA,	Professeur, à, Université Paris-Sud, PSE, IZA	Examinatrice



# Remerciements

J'ai longuement hésité avant d'entreprendre cette thèse mais aujourd'hui, je peux affirmer avec certitude que ces dernières années furent les plus enrichissantes de ma scolarité, et je tiens à saluer les personnes sans qui ce travail n'aurait pu voir le jour. Je témoigne en premier lieu ma gratitude à mes trois directeurs de thèse, Pierre André, Flore Gubert et Arnaud Lefranc. Malgré la configuration peu ordinaire, vous m'avez tous les trois accordé votre confiance.

Pierre, je ne pourrai jamais assez te remercier pour ton investissement. A travers notre papier, tu m'as guidée pas à pas dans les différentes étapes de la recherche. Merci d'avoir été si présent quand je sollicitais ton aide, d'avoir fait preuve de patience, mais aussi d'exigence. Ta rigueur m'a poussée à aller de l'avant et à me dépasser.

Flore, je te remercie pour ta bienveillance tout au long de la thèse. Tu m'as donné un cadre de travail optimal en m'ouvrant les portes de DIAL, et tu as aiguillé mon travail grâce à tes précieux conseils. J'ai toujours apprécié ton bon sens et ta capacité à prendre du recul pour mieux cerner les enjeux. Merci aussi d'avoir été à l'écoute lors de mes diverses interrogations.

Arnaud, je vous remercie pour votre encadrement et j'aimerais notamment revenir sur deux événements qui ont joué un rôle décisif. Le premier, sans doute vous le rappelez-vous... est lorsque vous m'avez empêchée de foncer droit dans le mur en identifiant un problème dans mon premier sujet de recherche. Suite à cela, j'ai abandonné un travail de six mois qui aurait pu se prolonger six mois de plus! Le deuxième événement que je tiens à rappeler ici est d'une toute autre nature. Au moment où j'étais sur le point d'abandonner mon second papier, vous m'avez donné un second souffle et m'avez encouragée à persévérer. Le papier est ce qu'il est maintenant, avec ses qualités et ses imperfections, mais il a le mérite d'exister.

J'ai énormément appris à vos côtés et je souhaite à tout doctorant de pouvoir vivre cette expérience enrichissante et privilégiée.

I would also like to thank Margherita Comola, Philippe de Vreyer, and Paul Glewwe who largely contributed to improve this work. Thank Margherita for accepting to be the director of the jury while the schedule was tight, and thank Philippe and Paul for your detailed remarks at the pre-defense. I should also express by gratitude to Paul who welcomed me at the University of Minnesota as a visiting student. Even if I was at first a little bit afraid of going to the Far North, this experience offered me great memories. I fully enjoyed this stay thanks to the stimulating environment at the university where I had the opportunity to follow classes, seminars and to interact with researchers. Beyond that, I discovered a wonderful researcher who turned out to be a very kind person. Thank you Paul for your advices, but also for your warm welcome (thank Mary too) and of course... for your « black-footed ferret » car.

J'aimerais ensuite remercier ma co-auteur Christelle Dumas pour les échanges productifs que l'on a pu avoir, notamment au cours des deux séjours à Fribourg. Merci d'avoir accepté cette co-écriture et d'avoir toujours fait preuve de persévérance, même lorsque l'on ne voyait pas le bout du papier.

J'aimerais manifester aussi ma sincère reconnaissance aux membres des laboratoires de recherche qui m'ont accompagnée ces dernières années. Au THEMA, je tiens avant tout à remercier Olivier Charlot qui, en tant que directeur de laboratoire, a réussi à dynamiser les échanges avec les doctorants. Merci d'avoir été à l'écoute et d'avoir encouragé les doctorants à interagir, au laboratoire et à des conférences. J'aimerais également saluer les chercheurs (Paul Maarek, Stefania Marcassa, Maelys de la Rupelle) et les doctorants (Amélie Adeline, Marie Baguet, Samia Badji, Jérémie Juste, Erwan Moussault, Jérôme Trinh) avec qui j'ai pu échanger. Merci Samia de m'avoir épaulée quand je commençais fraîchement ma thèse et merci Marie d'avoir partagé ton expérience américaine. Je tiens aussi à remercier Lisa Collin, Imen Defferrard et Yolande Gustave, pour assurer le bon fonctionnement du laboratoire.

My thanks also go to people that I met in Minneapolis: Marc Bellemare, Zach Erdman (for showing me Minneapolis, for your generosity and your open-mindedness) Ibrahim Keita (for all interesting research discussions from Mali to the US. You know how grateful I am for scanning, page by page, the Tanzanian school data!) Jason Kerwin, Aine Mc, Maxwell Mkondiwa, Lindsey Novak, Jeta Rudi, Liaila Tajibaeva (my office mate who becomes a precious friend. Thank you for the never-ending discussions!), Julie Buhl-Wiggers (for our breaks in and outside the university! It was really nice sharing this visiting experience with you), Benjamin Wuerflein (my dear housemate who was so considerate).

Je souhaite à présent remercier l'équipe DIAL de la rue d'Enghien qui a rendu cette expérience si chaleureuse. Quand je pense à DIAL, je pense à un lieu de travail mais aussi à un lieu de partage qui vous fait oublier que la thèse est un travail solitaire. Merci à Mohamed Ali Marouani, Oscar Barrera, Thomas Calvo, Raphael Cottin, Lisa Chauvet, Virginie Comblon, Anda David, Danielle Delmas, Axel Demenet, Siyavash Eslami Louti, Marin Ferry, Véronique Gille, Anne Hilger, Hoàng T. Quynh, Estelle Koussoube, Anne Legendre, Loïc Lepezennec, Marion Mercier, Sandrine Mesplé-Somps, Sarah Morcillo, Björn Nilsson, Christophe Nordman, Xavier Oudin, Laure Pasquier-Doumer, Mireille Razandrakoto, Anne-Sophie Robilliard, François Roubaud, Linda Rua, Camille Saint-Macary, Jean-Noël Senne, Marlon Seror, Marine de Talancé, Delia Visan, Jean-Michel Wachsberger et Claire Zanuso. Un merci tout particulier à Laure et Niri, mes collègues de bureau qui égalaient mon quotidien (c'est sûrement lié à la pilule bleue ou rose, ou au gingembre confit !) à Virginie pour ton oreille attentive et ta joie de vivre (et d'avoir sauvé la biblio), à Jean-Noël pour ta vivacité, ta bienveillance et tes conseils de dernières lignes droites, à Björn pour ta gentillesse inégalée, à Véronique pour ta curiosité, pour l'attention que tu portes au travail de chacun et pour avoir imposé la machine à café écolo, ce n'est pas rien – une petite révolution.

Voici maintenant le moment de remercier mes amis de longue date qui m'ont suivie et soutenue dans ce parcours. Il serait vain de croire que ces quelques lignes arriveront à exprimer toute l'affection que je vous porte, mais je ne peux pas écrire ces remerciements sans faire un saut du côté de la rue Charlemagne. Merci les « jerry-what-what » pour votre présence, d'être un bol d'air frais à chaque rencontre, d'avoir un humour... bien à vous ! Vous êtes irremplaçables. Ophélie, en écrivant ces lignes, j'ai le même pincement au cœur

que je te décrivais il y a maintenant 16 ans. Je suis heureuse de pouvoir, toujours et encore, m'arracher les cheveux lorsque l'on cherche à comprendre le pourquoi du comment. Je te remercie pour ton écoute, tes conseils et tes encouragements. Capucine, tu as beau ne pas être en économie, j'ai l'impression que tu as suivi ce parcours (et ses rebondissements) de près. D'abord à la BNF, puis aux pauses déjeuners dans le 10ème jusqu'au pays de ton enfance (le Béarn, tu sais les oies, ah non, les canards !), j'ai toujours apprécié nos discussions où le jugement n'avait pas sa place. Cela s'applique bien entendu à Barthélémy, Bérénice, Margot, et Maxence. Le temps passe mais je prends toujours plaisir à vous retrouver. Merci d'être fidèle à vous-même, de partager vos doutes et vos interrogations.

Frédérique... ces remerciements ne seront jamais assez « agréables » ni à la hauteur de ton amitié, j'espère que tu le sais. Avant de te remercier, j'aimerais juste te dire de ne rien changer. Garde ton intégrité et ton originalité, continue à t'offusquer quand quelqu'un te bouscule, et à rire comme tu le fais si bien. Merci d'être une amie si présente et compréhensive et de m'avoir soutenue à chaque tournant de cette thèse. Merci aussi pour toutes nos discussions (digressions) dont je ne me lasse pas.

Merci également à Sara aux quatre coins du monde qui reste toujours à l'écoute (reviens-moi vite !). Merci à Philippe de me communiquer sa passion pour l'économie et sa soif de culture. Et bien sûr, merci à Mélanie, rencontrée sur les bancs du primaire et retrouvée à Minneapolis.

Enfin, un grand merci à ma belle et grande famille, composée et recomposée. Maman, merci pour ton amour débordant, merci d'être là quand on a besoin de toi, même si cela implique de te mettre en quatre. Merci pour ton ouverture d'esprit et ta compréhension. Papa, malgré le scepticisme que j'ai cru entrevoir lorsque je me suis engagée dans cette voie... tu m'as fait confiance et tu as respecté mon choix. Tu me regardes, de près ou de loin, mais tu es là. Merci d'essayer de me sortir le nez des livres et de me pousser sans cesse à voir les choses différemment.

Merci à mes grands-parents pour leur tendresse dont ils m'ont entouré. Ma Nona, tu es un exemple de courage et d'acharnement et même avec cette thèse, je n'arriverai pas à ta cheville. Une pensée aussi à ma famille paternelle, aux oncles et tantes et aux 14 cousins cousines.

Armand, Fleur, Malvina, et bien sûr Stanislas, je me suis construite avec vous, et aujourd'hui, vous m'accompagnez chaque jour dans mes pensées. Je suis extrêmement chanceuse de vous avoir, et je vous remercie pour votre soutien inconditionnel.

Ma gratitude se dirige enfin vers Maxime, mon compagnon de tous les jours, avec qui je partage cette expérience depuis le début. Avant toute chose, je te remercie de toujours avoir respecté mes choix, même si cela signifiait être loin l'un de l'autre. Tu as su être patient, m'écouter, m'encourager lorsque la confiance me faisait défaut, et tu m'as aidée à lâcher prise lorsque j'en avais besoin. Merci de réussir à égayer le jour le plus maussade et de me rendre la vie si douce. J'espère que j'arriverai, aussi bien que tu l'as fait, à t'accompagner jusqu'au bout de ton travail qui te passionne tant.



# Contents

Remerciements	i
General Introduction	1
Chapter 1: Returns to farm child labor in Tanzania	17
Chapter 2: The Impact of the Universal Primary Education Program on Labor Market Outcomes: Evidence from Tanzania	75
Chapter 3: Good or bad timing? The pro-cyclical and counter-cyclical effect of shocks on education investment and on schooling performance.	125



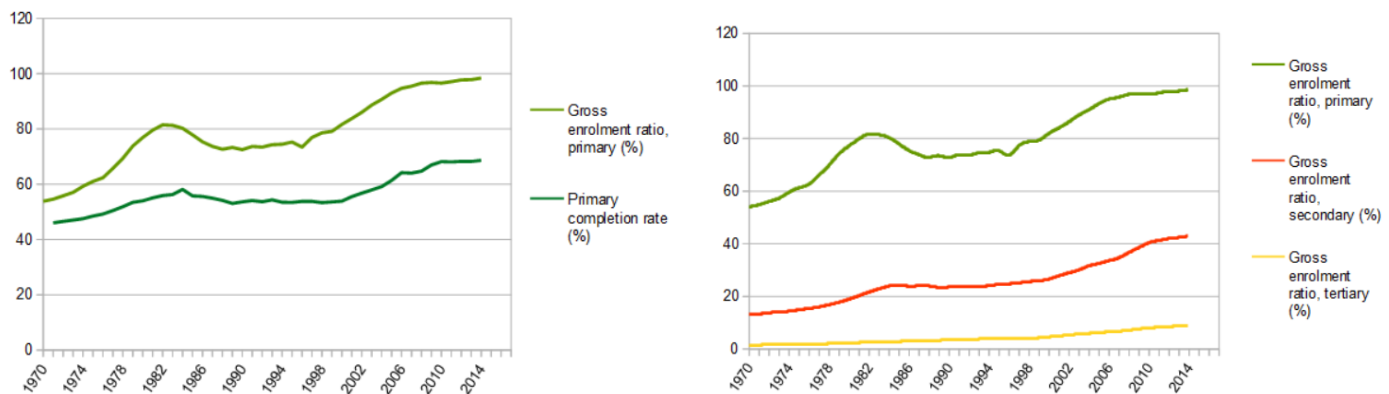


# General Introduction



The economic literature emphasizes the manifold benefits of education both from a microeconomic and from a macroeconomic perspective. Education is crucial to help eradicate poverty, counter the transmission of inequalities between generations and drive sustainable growth. This clear review of the situation convinced policymakers to make education a high priority and to achieve Universal Primary Education (UPE) by 2015.<sup>1</sup> With the joint efforts made by government and non-government organizations, sub-Saharan Africa has made great progress in providing access to school and in increasing the number of completed years of education. From 1999 to 2009, the primary Gross Enrolment Rate (GER) steadily increased by an average of 3.1 % per year<sup>2</sup> and almost reached 100% in 2014 (World Bank, 2015).

FIGURE 1: EVOLUTION OF ENROLLMENT RATES IN SUB-SAHARAN AFRICA



(A) PRIMARY ENROLLMENT RATE

(B) ENROLLMENT RATE BY SCHOOLING CYCLE

Source: Author's calculations based on World Bank data.

Figure 1 shows that in sub-Saharan Africa, the Gross Enrolment Rate in primary has doubled from 1970 to date. Despite this great progress, the dropout rate before completion of primary education is still high (primary education completion was less than 66 % in 2014), and relatively few children continue on to secondary and postsecondary education. In addition, the high enrollment rate is often offset by low attendance, which erodes the benefits of education. Attendance, especially in rural areas, is erratic. As a result of

<sup>1</sup>Universal Primary education is the second goal of the United Nations Millennium Development Goals.

<sup>2</sup>In comparison, the GER increased by 0.8% per year in the 1990s.

this low completion rate and poor attendance, 48 million children aged 15 to 24 in sub-Saharan Africa are still illiterate (UNESCO, 2013). This low educational achievement reflects diverse causes and poses new challenges that deserve to be further explored.

In this thesis, I analyze how to improve educational achievement in an agricultural environment. The analysis focuses on rural Tanzania. The thesis, composed of three chapters, uses a microeconomic approach and investigates access to education from different perspectives. The first chapter evaluates how child labor productivity in agriculture enters households' decisions, and how these decisions can hinder universal education. The second chapter utilizes an education policy implemented in Tanzania in the 1970's to estimate the benefits of education, where the benefits are measured in income and access to labor markets in a rural environment. Finally, the third chapter analyzes the impact of productivity shocks (climate variations and fluctuations in prices of cash-crop commodities) on child labor, education decisions, and children's cognitive skills.

To contextualize these research questions, this introduction chapter provides an overview of the specifics of the sub-Saharan African economy. In this context, I present the vulnerabilities of households and the strategies available to them to protect against shocks and insure their welfare. Then, I examine the consequences of such households' decisions on education outcomes. Finally, I present the outline of the thesis, the methodologies and the data used in this analysis.

## **0.1 A particular feature of sub-Saharan African countries: an economy highly dependent on agriculture**

The economy of sub-Saharan Africa (SSA) has to be assessed in the light of the territorial organization, characterized by the strong predominance of rural areas. As shown in figure 2, the percentage of the rural population has declined over time in almost every African country, but rural communities still constitute the largest share of the population. In 2015, the percentage of people living in rural areas was more than 60 % in sub-Saharan Africa

and was even higher in countries of the East African Community(79%).<sup>3</sup> Individuals living in rural areas can potentially work in various sectors, however, in practice, agriculture is extensively developed and remains the principal source of activities in SSA.<sup>4</sup> At the microeconomic level, this high dependency on agriculture implies that every change in agricultural productivity has an impact on the households' welfare, can keep people in poverty, and exacerbate their vulnerability (Baumeister and Kilian, 2014). Anderson and Brückner (2012) compute the shares of the real gross domestic product (GDP) by sector in SSA since 1950 and find that the agricultural sector accounts for more than a third of the GDP. More recently, this share is about 21 % and varies from 10 to 70 %, depending on the country (Sandri *et al.*, 2007; Devereux *et al.*, 2001). As a result, agricultural productivity shocks (weather shocks, fluctuations in commodity prices, etc.) constitute substantial factors of instability that can compromise food security, and have significant impact on economic growth (Addison *et al.*, 2016).

In this thesis, I focus on Tanzania, where the same stylized facts are observed. Agriculture in Tanzania accounts for 30 % of GDP and employs the majority of population. Most individuals involved in agriculture perform subsistence farming and cultivate small plot of land for a living.<sup>5</sup> Agricultural productivity is stalled by a limited access to technology and low development of irrigation systems, which further expose individuals to productivity shocks.

---

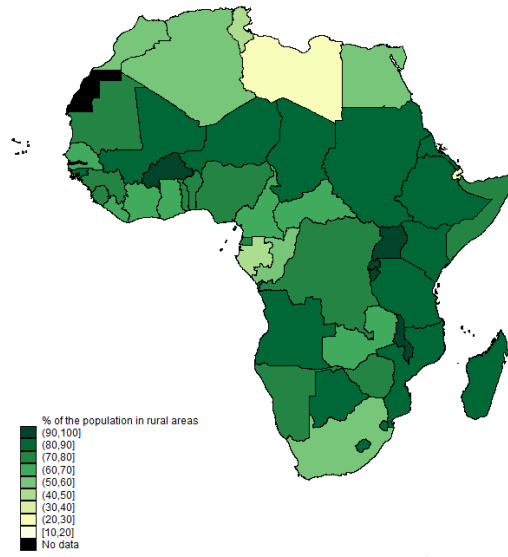
<sup>3</sup>The East African Community is located in the Great Lakes region and is comprised of Burundi, the Democratic Republic of the Congo, Kenya, Rwanda, Tanzania and Uganda.

<sup>4</sup>Sixty-two percent of the labor force performs agricultural activities (World Bank, 2015).

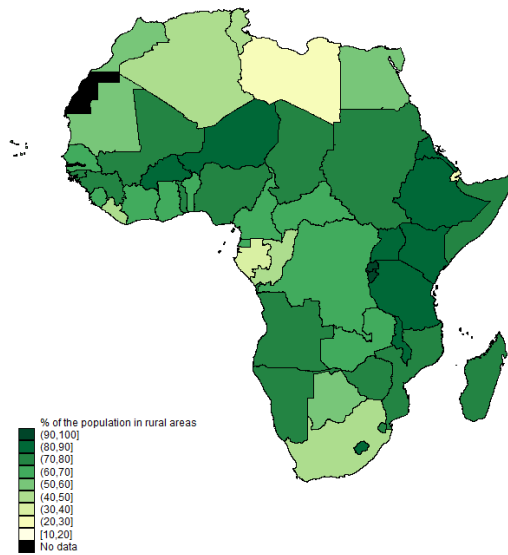
<sup>5</sup>The average plot size was about around 2.5 ha (World Bank report, 2011).

FIGURE 2: PERCENTAGE OF THE POPULATION LIVING IN RURAL AREAS

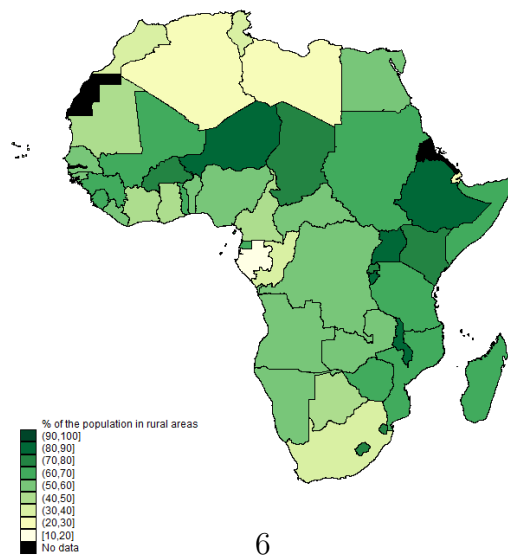
(A) IN 1980



(B) IN 1990



(C) IN 2015



## 0.2 Shocks in agriculture

Agriculture is a risky business where farmers face positive and negative income shocks that change their living standards. The literature commonly distinguishes two types of shocks, idiosyncratic and aggregate income shocks (Ferreira and Schady, 2009). Idiosyncratic shocks refer to individual variations of income, such as damage of crops, job loss, death or illness of household members, while aggregate shocks refer to covariate shocks that affect a broad community. The most widespread aggregate shocks are natural disasters (drought, floods, landslides, etc.), economic shocks (inflation, fluctuations of agricultural inputs and food prices), and civil conflicts.

Idiosyncratic and aggregate shocks differ from each other in their frequency and their magnitude. Idiosyncratic shocks are more common and explain between 75 to 96 % of total income variation (Morduch, 2005). However, as discussed in the next section, households can fall back on insurance systems to protect themselves against idiosyncratic shocks.

It is worth underlining that some shocks can change the opportunity cost of time. For instance, climate shocks, such as droughts, drastically reduce harvest and decrease labor productivity. Similarly, by changing the value at which households sell their agricultural outputs at the market, variations in commodity prices also change labor productivity. These shocks are called productivity shocks.

A cross-country comparison based on the LSMS-ISA data (Nikoloski *et al.*, 2018) describes the more widespread shocks that households face in five sub-Saharan African countries, including Tanzania.<sup>6</sup> The authors find that in Tanzania, most households are subject to frequent income shocks. About 60% of households have experienced sudden losses in income or assets, while the average number of reported shocks per household over the last five years reaches fourteen. The shocks that appear to be the most prevalent are weather shocks and food price variations.

---

<sup>6</sup>The five countries are Ethiopia, Niger, Malawi, Uganda, and Tanzania.



### 0.3 Which coping strategy to adopt?

In response to income shocks, households have access to several mechanisms. They can use insurance programs such as savings and credit. Although financial services recently expanded in sub-Saharan Africa, access to formal insurance systems remains limited and is not inclusive. In 2015, the World Bank reported that only 16 % of adults use formal savings and 6 % use formal borrowing systems in SSA countries, but these services are usually used by the richest households. This lack of access to protection systems contributes to inequalities and keeping people in poverty.

Alternatively, households may overcome these shocks by selling their assets and by smoothing their consumption (Nikoloski *et al.*, 2018). Although these decisions help households to cope with income shocks in the short-run, they may have negative long-run consequences. Indeed, selling off assets decreases households access to future resources and thus reduces their opportunities to get out of poverty. In comparison, reductions in consumption deprive individuals of nutrients they need which has adverse effects on children's health and cognitive development (Almond and Currie, 2011; Currie and Vogl, 2013).

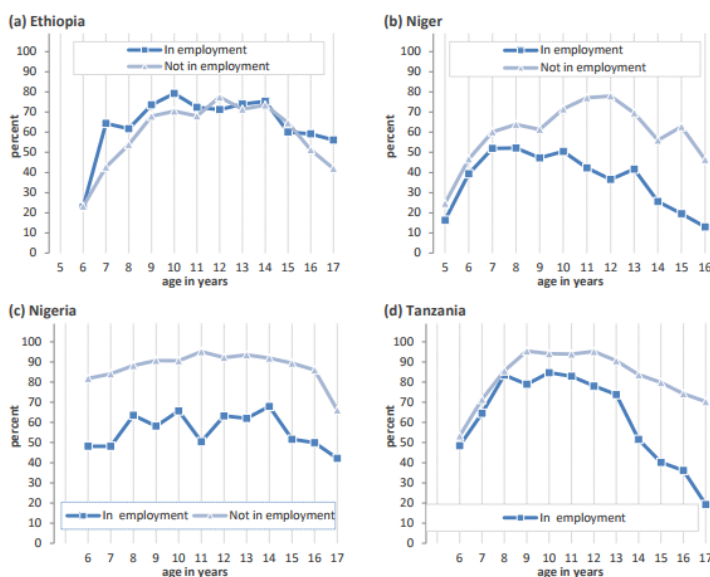
Since markets are imperfect and uncompetitive in developing countries (Jacoby, 1993; Skoufias, 1994; Chennareddy, 1967; Chavas *et al.*, 2005; Le, 2009), poor households rely on informal insurance from friends and family that tends to be inefficient when aggregated shocks hit a whole community at once. Alternatively, households can call on the available workforce and use marginal workers such as children to cope with shocks (Guarcello *et al.*, 2010; Dehejia and Gatti, 2005; Beegle *et al.*, 2006a).

The report of Nikoloski *et al.* (2018) underlines that, empirically, savings is the main coping strategy used by Tanzanian households. Formal insurance is limited to the wealthiest 60 percent of population, while poorer households engage in informal insurance systems, which are inefficient in case of aggregate shocks. Thus, adapting labor allocation remains a widely used strategy to cope with income shocks. This result is in line with the finding of Dumas (2015) showing that, in Tanzania, rainfall shocks affect child labor when credit and labor markets are imperfect.

## 0.4 Child labor, which consequences on education?

A number of international organizations such as Understanding Children’s Work (UCW)<sup>7</sup> have been dedicating their attention to child labor, which has severe consequences on development and children’s well-being. According to the ILO definition, child labor includes three categories of children, according to their age and the intensity of their activities.<sup>8</sup> This current definition of child labor is used to target and fight against all forms of child labor that may jeopardize education and children’s welfare. Indeed, an extensive literature has claimed that child labor may interact with education decisions. The recent cross-country study of nineteen developing countries (Guarcello *et al.*, 2016) highlights the negative correlation between work and school enrollment: working children have lower school enrollment than non-working children, and this is true for all countries.

FIGURE 3: ENROLLMENT IN SCHOOL, BY WORK STATUS AND AGE



Source: UCW calculations based on LSMS-ISA data in Ethiopia, Niger, Nigeria and Tanzania.

Based on the LSMS-ISA data, figure 3 compares the percentage of children enrolled in school by employment status and by age. In Tanzania, working children have lower school

<sup>7</sup>This is an inter-agency research institution created by ILO, UNICEF, and the World Bank.

<sup>8</sup>Children aged 5-11 years involved in any form of employment, children aged 12-14 years involved in any form of employment, excepting legal light work, children aged 15-17 years involved in any form of hazardous work.

enrollment than non-working children and the gap widens as children grow older. [Boozer and Suri \(2001\)](#) attempt to estimate the causal relationship of child labor on education and find evidence that child labor decreases contemporaneous enrollment in Ghana.

Although there is a negative correlation between child labor and education, the large majority of working children stay enrolled in school. Notwithstanding, working children who combine different activities are necessarily less involved in their education, which is probably harmful for their educational achievement. This big picture draws attention to the effect of child labor on education outputs such as children's performance at school. Empirically, there are still some debates on this relationship. Some studies have found no effect of child labor on schooling learning ([Dumas, 2012](#)), while others have found negative correlation between child labor and children's skills ([Akabayashi and Psacharopoulos, 1999](#)).

Despite these potential adverse effects on education, the frequency of child labor has not been reduced in sub-Saharan Africa. On the contrary, from 2012 to 2016, the percentage of children engaged in child labor has risen from 21,4 to 22,4% (ILO, 2016). These recent figures suggest that structural factors are not sufficient to curb child labor and that active public policies have to be implemented. However, banishing child labor can be inappropriate especially for the poorest households whom child labor can help escape poverty ([Basu and Van, 1998](#)). Thus, to be able to draw public policy recommendations that insure education progress in developing countries, it is necessary to understand the complexity of child labor and education decisions in rural sub-Saharan Africa.

## **0.5 Motivations and chapters' summaries**

In this thesis, I attempt to understand the demand for education in a rural risky environment. The objective of this analysis is to identify vulnerable children who are likely to drop out from school, and to draw public recommendations to protect and promote education. To do so, the three chapters of this thesis focus on factors that drive the demand for education, the costs and the benefits of education, and on factors that pull children out of school.

Direct and indirect costs of education are easily computed, but there remain

opportunity costs of education that are not straightforward to observe . These costs correspond to the additional income that children would have earned if they were not enrolled to school. In other words, this cost is the amount of extra income that parents have to give up if they send their children to school. Thus, these costs could explain why education investment is low even when there are no tuition fees. The first chapter of this thesis investigates children's productivity on farming to estimate a range of the values of one day of child labor and deduce the opportunity cost of children's time in agriculture.

Among the factors that explain education decisions, the returns to education also constitute a core determinant. To understand education investment in rural sub-Saharan Africa, it is necessary to assess the different benefits of education in a rural environment where the technology level is low and where the family farm is the dominant structure in agriculture. In the second chapter, I investigate the returns to education and the effects of education on access to the labor market in rural Tanzania.

To analyze factors that pull children out of school, the third chapter of this thesis explores whether short-run and recurrent positive and negative productivity shocks are detrimental to education attainment, and to educational achievement. To better assess vulnerable children, the effect of these shocks is disaggregated by children's age, from birth to secondary school age.

These research questions of public interest should help for formulating effective policies to protect children's education against shocks and to encourage parents to keep their children in school.

## **0.5.1 Chapter 1**

The purpose of the first chapter<sup>9</sup> is to study the opportunity costs of children's time that may significantly hinder universal education in developing countries. In this chapter, we are interested in the children's productivity in family farms, which represents the most widespread form of child labor in sub-Saharan Africa. To do so, we estimate the extent to which one additional day of child labor in the fields increases the households' production. We find that one day of child work raises the household production by US \$1.43-2.45, depending on the specification. Some heterogeneities can be observed:

---

<sup>9</sup>This first chapter has been jointly written with Pierre André (Cergy-Pontoise University) and Christelle Dumas (Fribourg University).

children's productivity increases with children's age, and girls appear far less productive than boys on the farm. This estimated range is also useful when calibrating the amount of Conditional Cash Transfers programs that aim to compensate the financial loss generated by schooling. Although children can combine education and work, we observe that on average children enrolled in school work 26 days less per year than children not enrolled in school. Thus, compensating children for the loss in income can be done by monthly payments between US\$3.1-5.3.

## **0.5.2 Chapter 2**

Although there is an extensive body of literature on the returns to education, few studies focus on sub-Saharan Africa. There are growing concerns about the quality of primary education in Africa, and most countries in the region have implemented policies to make primary education universal. The second chapter explores the impact of education on labor market participation and on households' consumption<sup>10</sup> in rural Tanzania. To address the endogeneity of education, I instrument the education of adults by exposure to the Universal Primary Education (UPE) program implemented in Tanzania in the 1970s. The exposure to this program is captured by variations across regions and over time. The results show a positive impact of education on households' consumption. This relationship is observed in every sector, but is more pronounced in agriculture. I find that education increases the probability of working in agriculture at the expense of non-agricultural self-employed activities. These results, at first glance surprising, illustrate the specifics of the schooling curriculum, which at the time of the program was composed of agricultural classes. Therefore, I find that returns to education are positive in agriculture, provided that skills taught at school are suitable for agriculture.

## **0.5.3 Chapter 3**

The third chapter investigates the effects of productivity shocks on education decisions and on human capital accumulation. This question lies at the crossroads between the literature on the fetal origin hypothesis and the literature that examines the effects of

---

<sup>10</sup>Consumption has the advantage of being better measured than income in developing countries, and is computable for every household.

contemporaneous shocks on education. The common thread of these two branches of literature is that, in developing countries, access to formal protection systems is imperfect and households have to develop informal mechanisms to cope with productivity shocks. In early life, productivity shocks change the available income, may affect the children's nutritional intakes, which could have long-lasting consequences on children's health and children's cognitive skills. When children are of school age and can work, productivity shocks still affect households' income, but also generate a substitution effect by changing the opportunity cost of children's time. Thus, the total effect of productivity shocks on education is ambiguous and depends on the relative sizes the income effect and the substitution effect. In this paper, I address this question and examine two particular aspects, the time at which these shocks occur and the length of these shocks. The first part of this chapter provides a very basic model to understand the main mechanisms involved in early life and in contemporaneous productivity shocks. Moving to the empirical part, I analyze the effect of two exogenous productivity shocks, climate and cash-crop price shocks. Results show that positive shocks in early childhood have persistent positive consequences on test scores. This can be explained by good nutrition in early age being favorable for the development of cognitive skills. In contrast, I find that contemporaneous positive shocks increase child labor and decrease education investment, meaning that the substitution effect is larger than the income effect. Finally, I show that cognitive skills result from a cumulative process and are especially sensitive to long-lasting shocks.

## 0.6 The methodology and the data

The three chapters which compose this thesis examine the determinants of education investments in order to understand how to protect education and avoid irregular attendance, which could compromise the benefits of education. Answering these questions requires solving identification issues that bias the results and lead to wrong interpretations.

Measurement of the relevant variables is the first challenge of this analysis. The literature recognizes the difficulty to correctly measure households' income<sup>11</sup> and

---

<sup>11</sup>Since income in agriculture is subject to significant variations, income is highly sensitive to prior events and may not be representative of the households' long-run wealth (Meyer and Sullivan, 2003). Another drawback is that income is not similarly measured between sectors of activity, which makes the comparison tenuous .

households' labor in agriculture <sup>12</sup> in developing countries. The second concern that puts into question the validity of the results is the difficulty of establishing a causal relationship between two events which are potentially endogenous. Indeed, simple correlation does not allow one to conclude a causal relationship and can simply be due to omitted variables. A well-known example is the omission of individuals' abilities when estimating the returns to education. Individuals with high income are likely to be more educated and to have higher abilities, while abilities are also likely to be positively correlated with education. Thus, if abilities are not observed, the positive relationship between income and education is not explained by education only. More generally, omitting variables that influence the decision process of a household prevents one from inferring causal relationships.

To address these endogeneity issues, I combine different methodologies and use diverse datasets which suit this analysis. In the first chapter, I estimate the productivity of one day of child labor performed in the fields by using the LSMS-ISA data. These data are particularly useful for this analysis and give a very accurate description of all agricultural inputs used by farmers. Since allocation of child labor is likely to be correlated with unobserved households' preferences, we take advantage of the panel dimension of the data and exploit variations over time. Instead of measuring the children's productivity by comparing production between households, we compare the households' production over time as child labor varies. If the unobserved households' characteristics, such as preference for labor, are time-invariant, this method would allow capturing the causal impact. Then, I control for all observable time-varying variables, such as rainfall shocks and pests, that could both influence child labor and households' production. Finally, I use an instrumental variable strategy to correct potential measurement error of child labor and to control for unobserved time-varying households determinants.

In the second chapter, I study the benefits of education in a rural environment by adopting a natural experiment approach. Generally, this empirical method exploits events such as natural shocks or changes in law or policy to capture exogenous variations of variables of interest. In this essay, I use variations in intensity of the Universal Primary

---

<sup>12</sup>As highlighted by Beegle et al., household labor in agriculture is largely misreported. Most farms are small-holder family farms that in general do not keep records of their inputs. Thus, households have to be reminded the quantity of labor they used during the recall period. On average, the hours of labor per person per plot is largely over-estimated, while the number of workers is under-estimated. Since the two biases are in the opposite directions, the labor aggregated at the household level does not suffer from a large amount of bias.

Education program to instrument education. Since this program was implemented during a limited period of time and targeted mainly regions with a low level of education, variations in intensity of this program are captured by a double difference, across locations and over time. To predict the treatment intensity of the program, I measure the education level before the program with two datasets. The first dataset is the 2002 Population and Housing Census in Tanzania, which gives a comprehensive picture of education. The second is administrative data that report the number of schools at the time of the program. These data are unique because they report the number of schools before the program was implemented and give an accurate measure of the school supply at this time.

Finally, in the third chapter I examine the effect of positive and negative productivity shocks on various sets of education outcomes to identify the factors that undermine educational achievement. To perform this analysis, I exploit variations across geographical areas and over time of two common aggregate shocks, climate shocks and fluctuation in commodity prices. The climate variable is constructed from standardized climate data that account for rainfall and temperature (Vicente-Serrano *et al.*, 2010), while the price standardized variable is constructed from the World Bank Commodities Price data. I also use the LSMS data that report detailed information of children's activities, and the Uwezo data that report test scores for children of school age. Contrary to most data on test scores, the Uwezo data are particularly interesting because they include test scores of children enrolled and not enrolled in school, which allow one to avoid sample selection bias.

\*

In this thesis, I adopt a set of approaches to understand why, despite the numerous investments that have been made in primary education in Sub-Saharan Africa, a significant share of children drop out of school prior to completing primary education. From an ethical point of view, this meaningful question suggests that children do not have equal opportunities (Dreze *et al.*, 1999) and do not necessarily acquire the basic reading and writing skills. To address this issue, I try to identify the factors that contribute to jeopardizing children's education. More specifically, I investigate whether the returns to education are positive in agriculture, whether children constitute a productive labor force on the farm, and whether productivity shocks are detrimental for education. This



subject is all the more relevant today when the number of productivity shocks, such as climate and price shocks, is growing. Although it is important to reduce the occurrence of shocks, it is also necessary to think about the ways of smoothing the effects of these shocks. This thesis provides some insight into the role of national social protection systems that could promote education.

# Chapter 1

## Returns to farm child labor in Tanzania



## 1.1 Introduction

In 2012, 144 millions children aged 5 to 14 were economically active. Child labour is primarily concentrated in agriculture (58.6%) and mostly performed within the household (68.4% of child labourers are unpaid family workers).<sup>1</sup> Unfortunately, we know very little of child labourers economic contribution to the household, which is a key determinant of household's time allocation decision, and as a consequence, a key aspect of any fight against child labor. Indeed, children who provide wage work are a minority of working children and are a very selected sample of this population. We should not infer anything from the wages they earn to the broader population. Our paper provides information on the productivity of children in household farming in rural Tanzania. For most children, this is a better estimate of the opportunity cost of children's time than the average child wage in the country.<sup>2</sup> In order to do so, we estimate several production functions to compute the productivity of children and exploit for identification the features of the LSMS-ISA panel dataset.

Our paper contributes to two different strands of literature. The first is the estimation of the opportunity cost of children, where the literature is scarce and mostly focuses on paid employment. *IPEC (2007)* studies different sectors which employ children and concludes to the diversity of situations in terms of children's return to labor: it ranges from children who are as productive as adults and are paid as such (in the Indian carpet manufacturing sector, for instance) to children who have a significantly lower productivity, or the same productivity but are paid much less (in the Indian garment industry, children are paid a sixth of their adult counterparts).<sup>3</sup> To our knowledge, this is the only study that evaluates the child productivity, and it does not cover on-farm labor supply. The large literature on market imperfections in developing countries incidentally addresses this

---

<sup>1</sup>Diallo *et al.* (2013).

<sup>2</sup>Indeed, in our sample, 80.65% of child workers are employed on the household farm.

<sup>3</sup>Children exhibit a productivity similar to the one of adults in carpet manufacturing in India but are paid less (Levison *et al.*, 1998; Anker *et al.*, 1998). In the fishing sector in Ghana and the construction sector in Uganda, adults and children receive a similar pay. However, in most other sectors considered in the study (chop bars in Ghana, pyrotechnics industry in the Philippines and vehicle repair in Delhi), children are paid less than adults and this reduction in pay is larger than the difference in productivity.

question (Jacoby, 1993; Skoufias, 1994; Lambert and Magnac, 1997) since they provide estimations of agricultural production functions. The opportunity cost of individuals who do not participate in the market is called the shadow wage and is equal to the marginal productivity of their on-farm labour. The papers by Jacoby (1993) (for Peru) and Skoufias (1994) (for India) do not compute the marginal productivity of children but the parameters associated to child labor in the Cobb-Douglas production function are fairly low compared to the ones for adult labour, indicating that children contribute only marginally.

The second strand of literature to which our article relates is the one on the effect of conditional cash transfers on children's time allocation. CCTs have been increasingly seen as an efficient tool for fighting current and future poverty. They have been implemented in almost all Latin American countries but are relatively rare in Africa. One notable exception is Tanzania. The evaluation of these programmes leads to the conclusion that child labor supply reacts very heterogeneously to cash transfers. For instance, Edmonds and Schady (2012) find that a transfer amounting to roughly 7% of the GNI per capita leads to a sharp decline in paid employment by 10 percentage points, and in unpaid economic activity by 19 percentage points, in Ecuador.<sup>4</sup> Large effects relative to the transfers (3% of the GNI per capita) are also found in the Food for Education programme in Bangladesh (Ravallion and Wodon, 2000). By comparison, Attanasio *et al.* (2010) find no effect of a CCT programme on child participation in economic activities in Colombia.<sup>5</sup> Other Latin American programmes have intermediate results, sometimes with much larger transfers.<sup>6</sup> From these results, it is difficult to extrapolate the amount of transfer that would lead to non negligible increases in schooling and reductions in child labor if one were to implement a conditional cash transfer in Africa. For instance, Kakwani *et al.* (2005) try to provide an ex-ante assessment of the implementation of a cash transfer programme conditional on school attendance in 15 Sub-Saharan African countries. However, the model fails to reproduce the conditionality for lack of data on the opportunity costs of

---

<sup>4</sup>This corresponds to a decline in paid employment by 41% and in unpaid employment by 34%.

<sup>5</sup>The transfer in Colombia is roughly 5% of the GNI per capita when a child is in secondary school.

<sup>6</sup>See for instance: Edmonds and Shrestha (2014); Bourguignon *et al.* (2003); Ravallion and Wodon (2000); De Leon and Parker (2000); Skoufias *et al.* (2001); Carpio *et al.* (2016); Galiani and McEwan (2013); Dammert (2009); Barrera-Osorio *et al.* (2011); Behrman *et al.* (2011); Schady and Araujo (2006).

children.<sup>7</sup> Since Bourguignon *et al.* (2003) have shown that ex-ante evaluations are useful tools for designing programmes, it is of interest to identify the key parameters of those models ahead of the implementation of the economic policy.

This is the contribution of this paper. Provided that most of working (African) children are unpaid family workers in the farm, we focus on them. This entails to identify their shadow wage. The estimation of production functions is delicate since inputs are chosen so as to maximize profit and might be plagued by endogeneity issues. In our case, we exploit the panel dimension of the data to control for unobserved and permanent determinants of household productivity and we instrument child labor by the number of children belonging to the household. With an imperfect labor market, characteristics of the household affect production choices such as labor demand (Singh *et al.*, 1986). Provided that we control for household fixed effects, the identification strategy relies on the aging of children between rounds of the panel. We also provide robustness checks to confirm that the endogeneity of other inputs does not plague our estimates. When allowing for different production functions, we find consistent results for child productivity. One day of work performed by a child between 10 to 15 years old leads to an increase in production value by US \$1.43-2.45, depending on specifications. Children enrolled in school work on average 26 days less in a year than non-enrolled children. Compensating them for the loss in income can be done by monthly payments between US\$ 3.1-5.3. We confirm that children 10 years old or younger are not productive and that children's productivity increases with age. We find that girls are far less productive than boys on the farm but this may be explained by the fact that they divide their day between different activities.

The remainder of the paper is organized as follow: section 3.3 describes the dataset, section 1.3 presents the specifications and the identification strategy, while section 2.4 provides the results. Section 1.5 computes the compensation that should be offered in a CCT, and section 1.6 assesses the productivity heterogeneity by child age and gender.

---

<sup>7</sup>The model therefore provides an ex-ante evaluation of an unconditional cash transfer.

## 1.2 Data

### 1.2.1 LSMS-ISA data

Our analysis is based on the Tanzanian LSMS-ISA (LSMS-Integrated Surveys on Agriculture) panel data. The panel is constituted of three waves: 2008-2009, 2010-2011 and 2012-2013.<sup>8</sup> The sample is representative at the national, the urban/rural and the major agro-ecological zones level. The original sample size was 3,265 households, spread over 409 enumeration areas across Tanzania and Zanzibar. Roughly two-thirds of these households are located in a rural area. The subsequent rounds revisit all households and includes potential split-offs. In the third round, all households previously interviewed are visited again. Given the large rate of split-off and the extremely low attrition rate in the panel (only 4.8% of households surveyed in 2008 are not observed in 2010 or in 2012), the third round of the panel interviews 5,015 households. The panel dimension of the dataset is crucial to our analysis and it is therefore extremely important to rely on a panel when attrition bias is unlikely to be an issue.

In addition, the data are particularly relevant for our analysis because they gather detailed information both on the production side and on the household side. All inputs and outputs are reported for each plot. Most importantly, the household members who have provided labor are recorded so that we are able to compute how many days of work have been provided by each household member (and a general category for non household members). The estimation of production functions is always delicate because it might be difficult to observe all inputs, and failure to observe all inputs increases the risk of biased estimates. In the case of LSMS-ISA, a large set of inputs is collected in addition to labour days. Organic and inorganic fertilizers, pesticides, and irrigation information are collected at the plot level. Land area is measured with a GPS, which is considered as much more accurate than estimations.<sup>9</sup> Self-reported land quality as well as erosion of the

---

<sup>8</sup>From October 2008 to 2009 for the first wave, from October 2010 to December 2011 for the second wave and from October 2012 to December 2013 for the last wave.

<sup>9</sup>In the first round, not all plots were measured with GPS. We use the GPS information when available

plot are also collected. In addition, the LSMS-ISA data are matched with information on rainfall, greenness and temperature, obtained from satellites measurements.<sup>10</sup> Appendix 1.7.1 describes the variables in detail.

## 1.2.2 Production

Even though inputs and outputs are collected at the plot level, the panel does not allow one to match plots from one date to the next and we aggregate the information of all the plots at the household level. As a consequence, our unit of analysis is a household observed in a given year. In order to obtain household farm output, we use the monetary value of each crop (in T-shillings), as declared by the household. We do not include farm outputs such as fruits growing and cattle herding in our analysis for the following reasons: 1) trees do not need much labor investment once they are planted, 2) in the data, it is not possible to compute the value of households' cattle, and 3) in both cases, labor time allocated to these activities are not collected. Neglecting cattle might be an issue when it comes to the analysis of child labor since this is one of the predominant activities of children. However, assuming that households behave efficiently, they should equalize the marginal productivity of child labor across the various possible farm activities, and thus focusing on activities in the field should not bias our estimates of child productivity.

Tanzania has two types of agriculture: in the North-North East, there is only one cropping season, that lasts roughly from November to May; the rest of the country has two cropping seasons, the short one taking place in October-February and the long one in February-July. In the LSMS-ISA data, households were retrospectively interviewed about working time, inputs and production during the long and short rainy seasons separately. They all answer for the same long-rainy season, but the month of interview (partly) determines which short season they provide information for. We follow the information provided by the LSMS by using rainfall data that correspond to the last long- and short- and use the respondents' estimation when not.

<sup>10</sup> [http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.FEWS/.Africa/.DAILY/.ARC2/.daily/.est\\_prcp/datafiles.html](http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.FEWS/.Africa/.DAILY/.ARC2/.daily/.est_prcp/datafiles.html)



rainy seasons for most of households. In practice, for the third round of the panel, collected between October 2012 and December 2013, we use the rainfall data between July 2012 and June 2013. This is the variable we call "Rainfall".

### 1.2.3 Child labor in the data

We have various sources of information in the data about child labor: whether the child has worked in the week before the survey (and how many hours), hours of domestic work in the day before the survey and, as previously described, days of agricultural work over the previous agricultural season. The data confirm the importance of the agricultural work for children living in rural areas of Tanzania. Among the 5 to 15 years old children who live in a rural area, 24 % have worked the last 7 days. Only 2% of children of this age range worked for a wage. The few who did so earned an average daily wage of 5,316TSh (equivalent to 2016 US\$5.1).<sup>11</sup> Two-thirds of children did not perform any domestic work during the day before the interview. Among those who provided domestic work, the median duration is only half an hour. As expected, girls are more likely to perform domestic chores (41% of them participate, compared to 25% for boys) but when they do so, they do not spend more hours than boys. Among children aged 5 to 15 and who belong to a land-endowed household, 22.4% have participated to the farming in the previous year. Here, there is no difference by gender: 22.7% of boys have participated, as have 22.1% of girls.

The ILO definition of child labor includes all economically active children aged from 5 to 15. However, children from 5 to 9 years old work only marginally: Table 1.1 shows that the share of children working significantly rises with age and that older children also devote more time to agricultural work.<sup>1213</sup> Estimating the labor productivity of young children

---

<sup>11</sup>1,000TSh in 2008 = US\$ 0.95 in 2016, authors' computation based on World Bank series of inflation rates in consumer price index and official average exchange rates. Sources: [http://donnees.banquemondiale.org/indicateur/FP.CPI.TOTL.ZG?locations=TZ&name\\_desc=true](http://donnees.banquemondiale.org/indicateur/FP.CPI.TOTL.ZG?locations=TZ&name_desc=true).

<sup>12</sup>Very few children are economically active in other sectors than the agriculture: among the 5 to 9 years old, 0.10% earn a wage and 0.06% work as self-employed.

<sup>13</sup>This is also consistent with the fact that 99% of 5 to 9 years old children are enrolled in school, while 87% of 10-15 years old are enrolled.

is difficult because few of them do work. We will therefore focus on the productivity of children older than 10 and younger than 15. However, we will provide estimates of the productivity differential by child age.

TABLE 1.1 : AGRICULTURAL CHILD LABOR BY AGE

Age	Share of working children	Number of days (of working children)	Obs.
5	1.09 %	22.04	483
6	2.59%	31.39	522
7	4.12%	24.92	462
8	5.36%	38.61	546
9	13.89%	41.88	500
10	20.79%	50.99	566
11	22.98%	59.38	478
12	32.59%	46.35	552
13	36.51%	50.29	587
14	46.15%	51.69	658
15	48.82%	61.17	610

In addition, we will restrict our analysis to farm labour because it encompasses the main activity of children and because this is the only activity for which the output is measurable. Our definition of child labor is therefore the number of farming days in the last year performed by children aged 10 to 15 years old. When aggregating at the household level, we observe on average 16 days of child labor, while there are 185 days of adult labor. Child labor therefore constitutes 8% of the total farm labor, but they contribute as much labor (in days) than non-household workers.<sup>14</sup> Roughly a third of child farm labor is devoted to preparing and planting; a third to weeding and a third to harvesting. Adults from the surveyed households have a fairly similar work allocation than them; if anything, adults spend slightly more time on the preparation of the field and less on harvesting (Figure A3.6 in the Appendix).<sup>15</sup> Last, children mostly allocate their farm work to the same crops as their parents (see Figure A1.6 in the Appendix) except for cassava: 34% of their working days are allocated to cassava, against only 20%

<sup>14</sup>Table A1.10 , in the Appendix, shows that these number of days are quite stable across years.

<sup>15</sup>Non-household members spend significantly more time on weeding (almost 40%) and significantly less on harvesting (22%).

for the adults. <sup>16</sup> Cassava is known to be an easy crop to cultivate and might be an “easy” task to give to a child.

### 1.2.4 Sample

Our sample consists of households who farm land and where at least one adult member participates to the farming. Because we implement a household fixed-effect strategy, we only keep households surveyed at least twice. Households who are re-surveyed but who have moved in a new location are discarded since, essentially, our fixed-effect strategy is also implemented to control for unobserved soil characteristics. Our total sample is constituted of 4996 household-year observations (1703 surveyed in 2008, 1711 in 2010 and 1582 in 2012). Special attention must be paid to households who split during the panel (20% of households at each wave). We treat differently the household who is considered as the original household and the split-off. In practice, the “original” household after the split is mostly constituted of members from the original household (85%) while the split-off has as 35% of original members, i.e. on average less than 2 members from the original household. Very often the split-off household declares himself as living in a new location, while the original household is in the same location. As a consequence, we treat the split-off as a totally new household entering the panel.

## 1.3 Production functions and identification

Estimating the productivity of children on the farm entails to estimate a production function, which leads to a series of choices. First, different functional forms are available, with different properties, and sometimes different ways of estimating these functional forms. Second, one has to choose the set of inputs; third, one has to deal with null values of inputs and output; fourth, one should choose what type of productivity is reported. Last but not least, inputs on the farm are the result of a choice made by households and

---

<sup>16</sup>This is not driven by a gender allocation of tasks since women allocate as many of their days (23%) to cassava as do men.

could therefore be endogenous. In this section, we describe our set of choices and explain how we deal with the question of identification.

Before turning to this set of choices, let us clarify also why we have made the choice to estimate production functions rather than profit functions. Indeed, profit functions can be expressed as functions of (some) input prices rather than input levels. This is convenient because prices are more naturally assumed exogenous to the household, while input levels are the result of precisely the profit maximization and should therefore be considered as endogenous. However, estimating profit functions runs into two additional issues. First, collected prices must reflect relevant prices for the households, which is not necessarily the case if markets are imperfect and in particular for some inputs that the household can acquire without the market (seeds, for instance). Second, some households end up with negative profits. While this may be due to shocks to production, it could also be due to a wrong estimation of the production costs, and it is therefore difficult to properly address this situation.<sup>17</sup> Dealing with production functions will allow us to avoid these difficulties.

### 1.3.1 Production function

There is a large choice set when it comes to specifying a production function. Insofar as we are interested in estimating child labor productivity, in particular by comparison with adult labor productivity, we have allowed for specifications that varied the degree of substitutability between both types of labor. Much less emphasis has been put on the analysis of the link between labor as a whole and the other inputs. As a consequence, we posit a Cobb-Douglas relationship between (total) labor and the other inputs. The elasticities of substitution between labor and other inputs are therefore assumed constant equal to one.

We focus our attention on the possible substitution between child and adult labor. A

---

<sup>17</sup>Bos and Koetter (2011) study three imperfect solutions to deal with negative profits: (1) censoring observations with negative profits, (2) using  $\log(\pi - \min(\pi, -1))$  instead of  $\log \pi$  and (3) replacing  $\log \pi$  by  $\log(\max(1, \pi))$  on the left hand side and controlling for  $|\min(\pi, -1)|$  on the right hand side.

Cobb-Douglas specification for different types of labor does not seem attractive to us since it entails an infinite marginal productivity of child labor in zero, which is not supported by the data: a large share of households do not use child labor. A quite flexible functional form would be a CES function:

$$Y = A[L_a^{-\rho} + \gamma_c L_c^{-\rho}]^{-\alpha/\rho} X^\beta \quad (1.1)$$

where  $L_a$  and  $L_c$  are respectively adult and child labor,  $X$  are the other inputs (land, fertilizers...) and  $A$  is a productivity factor. Turning to the parameters,  $\rho \in [-1, +\infty[$  is the substitution parameter between both types of labor,  $\gamma_c \in [0, 1]$  is the relative productivity of child labor compared to adult labor and  $\alpha, \beta < 1$ . Eq. (1.1) can be log-linearized:

$$\log Y = \log A - \frac{\alpha}{\rho} \log(L_a^{-\rho} + \gamma_c L_c^{-\rho}) + (\log X)\beta \quad (1.2)$$

but cannot be linearly estimated and often leads to very unstable results (Henningsen and Henningsen, 2012).

Based on this general specification, we offer three different specifications and linearizations.

**1.3.1.0.1 Perfect substitutes** The first assumes that child and adult labor are perfect substitutes ( $\rho = -1$ ). In that case, the total amount of efficient labor is  $L_a + \gamma_c L_c$ . The corresponding production function is:

$$\log Y = \log A + \alpha \log(L_a + \gamma_c L_c) + (\log X)\beta \quad (1.3)$$

When  $\gamma_c L_c \ll L_a$ , this can be linearly approximated by:

$$\log Y \approx \log A + \alpha \log L_a + \alpha \gamma_c \frac{L_c}{L_a} + (\log X)\beta \quad (1.4)$$

**1.3.1.0.2 Imperfect substitutes** The second venue consists in departing from the perfect substitution approach ( $\rho > -1$ ). Assuming that  $\gamma_c \left(\frac{L_c}{L_a}\right)^{-\rho} \ll 1$  and that  $\rho$  remains close to -1, equation (1.2) can be approximated by a Taylor development as:

$$\log Y \approx \log A + \alpha \log L_a + \alpha \gamma_c \frac{L_c}{L_a} + \alpha(1 + \rho) \gamma_c \frac{L_c}{L_a} \left(1 - \log \frac{L_c}{L_a}\right) + (\log X)\beta \quad (1.5)$$

If the fourth term is equal to 0 (but  $\alpha \gamma_c \neq 0$ ), then both types of labor are perfect substitutes.

**1.3.1.0.3 Translog** Finally, we can simply use a translog function.<sup>18</sup> This is useful because the translog function is very flexible and allow us to avoid any assumption on the parameters. The translog function has also been demonstrated to approximate the CES function when  $\rho \approx 0$  (Kmenta, 1967).

$$\begin{aligned} \log Y \approx \log A + \alpha_a \log L_a + \alpha_c \log L_c + \alpha_{aa}(\log L_a)^2 + \alpha_{cc}(\log L_c)^2 \\ + \alpha_{ac} \log L_a \log L_c + (\log X)\beta \end{aligned} \quad (1.6)$$

Obviously, the translog specification is the most flexible one among the three offered specifications, but the coefficients cannot be interpreted as structural parameters. This will determine how we compare the results of the three specifications.

## 1.3.2 Child labor productivity

Our ultimate goal is to compute child productivity per day of farm labor. There are different ways to do this. First, in the specification where child and adult labor are perfect substitutes (eq. 1.4),  $\gamma_c$  is a measure of the productivity rate of children compared to adults. If we know the adult wage rate and assuming that the labor market equates the marginal productivity to the wage, then we can compute the child marginal productivity

---

<sup>18</sup>In this specification, we adjust the translog function to impose our hypothesis that the substitution between labor and other inputs is constant equal to 1. We provide robustness tests for that assumption.

by:  $w_c = \gamma_c w_a$  where  $w_a$  is the adult wage. However, this series of assumptions (perfect substitutes and perfect labor market) is doubtful.

In addition, we may not be interested in the marginal productivity but rather on the average productivity. Indeed, if marginal returns are decreasing, marginal productivity informs us on the production obtained with the last day of work. A cash transfer would aim to reduce substantially, and maybe even suppress, child labor. In order to achieve this, families would have to be compensated for a larger number of days and average productivity on those days is the relevant concept.

As a consequence, we will provide for each specification the *average semi-elasticity of output with respect to days of child labor*. More precisely, our estimates of equations (1.4), (1.5) and (1.6) allow us to compute the expected production in absence of child labor, using the actual number of days of child labor for each farm, based on the households' use of the other inputs. We then compute for each household:

$$E_{Y_{cit}} = \frac{\log \widehat{Y}_{it}(L_c = L_{cit}^o) - \log \widehat{Y}_{it}(L_c = 0)}{L_{cit}^o} \quad (1.7)$$

where  $L_{cit}^o$  is the number of days of child labor observed in household  $i$  at date  $t$ . The numerator is therefore the predicted difference in the production (expressed in logs) between the situation where the child does not work and the situation where he works the actual number of days.  $E_{Y_{cit}}$  is the average labor productivity of children in household  $i$  at date  $t$ , provided that all other inputs remain the same. Obviously, this can only be estimated for households that use child labor, but this is our sample of interest (in spite of the fact that we use all farming households for the estimation).

We then weight households by amount of child labor and average the individual semi-elasticities to obtain an aggregate measure of child labor productivity:

$$\bar{E}_{Y_c} = \frac{1}{\sum_{i,t} L_{cit}^o} \sum_{i,t} L_{cit}^o \cdot E_{Y_{cit}} \quad (1.8)$$

Because adult productivity is a benchmark against which child productivity should be

evaluated, we compute that as well. Adult on-farm labor is rarely null, so to compute the average productivity on a meaningful margin, we take as a base the 10th percentile of adult labor that is observed in our sample.<sup>19</sup> The adult semi-elasticity  $\bar{E}_{Y_a}$  is computed as:

$$E_{Y_{ait}} = \frac{\log \widehat{Y}_{it}(L_a = L_{ait}^o) - \log \widehat{Y}_{it}(L_a = \underline{L}_a)}{L_{ait}^o - \underline{L}_a} \quad (1.9)$$

$$\bar{E}_{Y_a} = \frac{1}{\sum_{i,t} L_{ait}^o} \sum_{i,t} L_{ait}^o \cdot E_{Y_{ait}} \quad (1.10)$$

where  $\underline{L}_a$  is the the 10th percentile of adult labor (38 labor days per year for a farm).<sup>20</sup>

### 1.3.3 Inputs

The inputs that are included in the estimation are the following ones: adult labor, child labor, cultivated land size, use of organic and inorganic fertilizer, spending on pesticides, erosion of the plot, irrigation of the plot and productive assets. Non-household labor days are aggregated with household adult labor days (no non-household child labor is used on the farms). Several inputs such as child labor, fertilizers and pesticides have frequently null values, which raises a problem for our specifications in logs. In order not to restrict the sample to the households who have positive values of all inputs, which would lead to selection bias, we follow MacKinnon and Magee (1990); Burbidge *et al.* (1988); Pence (2006) by using a modified function of the logarithm that is defined in 0:

$$\log^M(x) = \log \frac{1}{2} (x + \sqrt{1 + x^2}) \quad (1.11)$$

This function behaves similarly to the log function when  $x$  is large. As a consequence, for all inputs that have large values, the estimated coefficient reflects the increase in the production (expressed as a percentage) associated to an increase by 1% of the input. Given that child labor is often equal to 0 in our data, we cannot use this approximation for the

---

<sup>19</sup>The estimated production function fits the data only for the range of adult labor that is observed. We do not want to extrapolate outside of this range.

<sup>20</sup>More precisely, in this computation, we discard the observations with less than 38 adult labor days.



interpretation of the coefficient. The same holds as well for the other inputs that tend to be close to zero. The semi-elasticities of production with respect to labor are computed taking into account that the  $\log^M$  function (instead of  $\log$ ) is used for the estimation. The details for these computations are provided in Appendix 1.7.3.

In the data, we also have a non negligible number of households who declare a null production, despite non-zero inputs. This is due to disasters such as droughts and pests. We choose to keep these observations with null production in order to avoid a selection bias. We therefore use the same modified function. Given that the expected value of output is always large, we consider our function to be well approximated by the logarithm function and interpret it accordingly (the effect of one additional unit of input is expressed as a percentage increase in the production).

### 1.3.4 Identification and specifications

Ordinary Least Squares (OLS) estimates of equations (1.4), (1.5) and (1.6) may suffer from endogeneity bias for several reasons. First, unobserved household permanent characteristics (wealth, abilities, education, network affiliations, etc) may influence both household's production and household's labor allocation decisions. In addition, households observe the circumstances of the production much better than econometricians and therefore may adapt their farm allocation decisions to determinants of production that are not observed (sunlight, rainfall, temperature, pests, etc).

Our solution to this endogeneity issue combines three techniques. First, we exploit the panel dimension of our data and control for household fixed-effects. Second, we control for observed household time-varying characteristics such as rainfall shocks, as well as idiosyncratic and covariates shocks that the household has declared. We also controlled for temperature and greenness (measured by satellites), but given that their coefficients were not significantly different from zero, and that our estimates were unchanged, these variables were removed from the list of covariates. We also control for village- year fixed effects, which should capture common shocks. Third, because unobserved and individual

time-varying determinants may still be correlated with labor decisions and productions, we instrument child labor.

According to the standard agricultural household model (Singh *et al.*, 1986; De Janvry *et al.*, 1991), when markets are complete and competitive households' production decision are separable from their consumption decisions, and labor demand only depends on inputs and output prices. The literature that has tested whether markets are perfectly competitive in developing countries almost unanimously reject the hypothesis (Jacoby, 1993; Skoufias, 1994; Chennareddy, 1967; Chavas *et al.*, 2005; Le, 2009). In addition, they use household composition as a determinant of farm labour supply. For instance, Benjamin (1992) tests whether household labor demand is independent from family composition in rural Java and rejects the separability assumption. We exploit the labor market imperfections in rural Tanzania (?)<sup>21</sup> and use the number of children aged 10 to 15 years old as an instrument for child labor. Given that we control for household fixed effects, this approach amounts to predicting variations in farm labor based on variations in the household of the age of household members.

More precisely, for each household, we build a pool of children which will be aged 10 to 15 years old at some point during the course of the panel. This pool of children consists only of children who are offspring of at least one member of the household. Fostered children are excluded for endogeneity reasons: it could be the case that they are fostered in the household precisely because the household has a large land endowment and needs manpower.<sup>22</sup> Ideally, we would like to use all offspring of adult household members. However, the dataset lacks information on those children: we do not know if adults have children living elsewhere. The pool of children is therefore constituted of all offspring of household members, who were recorded as belonging to the household for at least one round of the panel.<sup>23</sup> The underlying assumption is that this pool of children

---

<sup>21</sup>? shows that only 5 % of households hire external workforce. She also shows that positive rainfall shocks increase child labor suggesting that the separability assumption does not hold.

<sup>22</sup>Indeed, Safir (2009) shows that the household composition (in Senegal) reacts to shocks.

<sup>23</sup>In the case of households who split during the course of the panel, we adjust the pool of children accordingly. More precisely, we build the pool of children based on the years for which the household is supposed to remain the same, as we did for the definition of the fixed effect. For instance, a household observed in 2008 and that split by 2010 gives birth to two households in 2010 and 2012: the original

constitutes all children who may belong to the household in case of shocks (offspring who never belonged to the household during the six years of the panel are assumed not to be available for help). This pool of children  $\mathcal{N}_{ci}$  for household  $i$  is then fixed for the entirety of the panel and  $N_{cit}$  is the number of children  $k$  who belong to this pool and are aged 10 to 15 in year  $t$ :

$$N_{cit} = \sum_{k \in \mathcal{N}_{ci}} 1(10 \leq \text{age in year } t \text{ of child } k \leq 15). \quad (1.12)$$

This procedure is of interest to us because only the aging of children provides variation in  $N_{cit}$ , not the actual presence of children in the household, which could be correlated to the needs of the household. More precisely, for all the specifications where we do not allow for heterogeneity in child productivity by age, only the entry/exit of children into/out of the 10-15 age bracket provides the exogenous variation that is used in the estimation. Children who provide time variation in a given household are children who are included in the pool and who cross the age limit between two rounds. Since rounds are spaced by two years, they are aged 10 or 11 when they enter the age bracket and they are aged 14 or 15 before exiting the age bracket. As a consequence, the productivity estimation relies on a mixed composition of children aged 10, 11, 14 and 15 years old. Figure A1.1, in the Appendix, provides the distribution of children belonging to the household by age, across panel years. We see that no age pattern emerges and that the children are balanced with respect to age. When no heterogeneity in productivity is allowed, we estimate the average productivity on children of the previously mentioned ages, which should be close to the average productivity of children between 10 and 15 years old. As a robustness check, we also instrument child labor by the number of children by age, comprised between 10 and 15 year old  $N_{10it}, \dots, N_{15it}$ .<sup>24</sup>

However, in spite of the fact that our instrument is not manipulated by the household, 

---

household and the split-off. The split-off is only observed in 2010 and 2012. The original is observed from 2008 to 2012. The pool of children for the split-off is based on offspring observed at least once in 2010 and 2012, while the pool of children for the original household is based on offspring observed at least once in 2008, 2010 or 2012. Households not observed more than once are not used for the estimation.

<sup>24</sup>By doing so, the productivity estimation is computed from children aged 10 to 15.

we still could face violations of the exclusion restriction if these entry/exit of children into/out of the age range had other consequences in terms of production.<sup>25</sup> Let us start with entry into the 10-15 age range. One additional child in the 10-15 age range is associated to one fewer child in the 5-9 age range. If children of this age range actually work and are productive, then our production function is misspecified (because labor performed by young children is not included) and this has consequences for the identification. The increase in production associated with the entry is in fact only the difference in productivity between young children and older children, instead of being the whole productivity of older children. This is unlikely to be a serious problem since young children provide little labor. Conversely, the exit out of the age range is in fact associated with an increase in adult labor time. Here, the problem is slightly different since we do control for adult labor. However, this control is imperfect if actually the adult productivity is heterogenous by age. Allowing for heterogenous productivity by child age will help us check whether the exclusion restriction is violated but we postpone this analysis to section 1.6.

To summarize, our specifications will be the following, with  $i$  indexing households in village  $v$ , and  $t$  indexing dates:

$$\log Y_{it} = \alpha \log L_{ait} + \alpha \gamma_c \frac{L_{cit}}{L_{ait}} + (\log X_{it})\beta + \mu_i + \zeta_{vt} + \varepsilon_{it} \quad (\text{Perfect})$$

$$\begin{aligned} \log Y_{it} = & \alpha \log L_{ait} + \alpha \gamma_c \frac{L_{cit}}{L_{ait}} + \alpha(1 + \rho)\gamma_c \frac{L_{cit}}{L_{ait}} \left(1 - \log \frac{L_{cit}}{L_{ait}}\right) \\ & + (\log X_{it})\beta + \mu_i + \zeta_{vt} + \varepsilon_{it} \quad (\text{Imperfect}) \end{aligned}$$

$$\begin{aligned} \log Y_{it} = & \alpha_a \log L_{ait} + \alpha_c \log L_{cit} + \alpha_{aa}(\log L_{ait})^2 + \alpha_{cc}(\log L_{cit})^2 \\ & + \alpha_{ac} \log L_{ait} \log L_{cit} + (\log X_{it})\beta + \mu_i + \zeta_{vt} + \varepsilon_{it} \quad (\text{Translog}) \end{aligned}$$

with  $\mu_i$  standing for household fixed-effects and  $\zeta_{vt}$  for village-by-year fixed effects (not

---

<sup>25</sup>To test whether the arbitrary thresholds are not specific, we also provide estimates for children aged 11 to 16.

systematically included). The instrumentation equations are the following in each case:

$$\frac{L_{cit}}{L_{ait}} = \delta_1 \frac{N_{cit}}{N_{ait}} + \delta_2 \log L_{ait} + (\log X_{it})\delta_3 + \nu_i + \xi_{vt} + \eta_{it} \quad (\text{Perfect and Imperfect})$$

$$\log L_{cit} = \delta_1 N_{cit} + \delta_2 \log L_{ait} + \delta_2 2 \log L_{ait}^2 + (\log X_{it})\delta_3 + \nu_i + \xi_{vt} + \eta_{it} \quad (\text{Translog})$$

with  $\nu_i$  household fixed-effects and  $\xi_{vt}$  village-by-year fixed effects.

$N_{cit}$  is called the number of children from 10 to 15 but is actually the variable defined in equation (1.12).  $N_{ait}$  is built similarly as  $N_{cit}$ : it is number of adults from 16 to 65 at date  $t$  among the adults observed at least once in household  $i$  during the panel.

The demographic structure is a valid instrument if:

$$E\left(\frac{N_c}{N_a} \cdot \varepsilon \mid \log(X), \log(L_a), \log(L_a), \mu, \zeta\right) = 0 \quad (\text{Perfect and Imperfect})$$

$$E(N_c \cdot \varepsilon \mid \log(X), \log(L_a), \mu, \zeta) = 0 \quad (\text{Translog})$$

A second issue to deal with in **Imperfect** and **Translog** specifications is that the term to be instrumented appears more than once and with different interactions or functional forms. If we instrument each of these terms, it would lead to extremely low partial R-squared for each of the instrumented term. Instead, we follow [Wooldridge \(2015\)](#) and implement a control function approach. This amounts to predicting  $\hat{\eta}_{it}$  in the first stage equation and including it as a control in the main equation. The underlying idea is that  $\hat{\eta}_{it}$  captures the endogeneity of the household behavior and then is controlled for, in the same spirit as the inverse-Mills ratio in a selection equation. [Wooldridge \(2015\)](#) shows that this parsimonious control function approach leads to more efficient estimates than the IV estimates when the LHS variable is not linear in the endogenous variable.

So far, we have discussed only how to deal with the endogeneity of child labor. However, the households are also expected to choose the other inputs. We should be cautious in the interpretation the estimates of those other inputs. However, part of the identification strategy for the child labor productivity already deals with similar endogeneity concerns for other inputs. In particular, household and village-by-year fixed

effects likely deal with a substantial amount of joint determination issues. In addition,  $\hat{\eta}$  should also pick up unobserved and idiosyncratic shocks occurring to households, that are also relevant for the determination of other inputs use, in particular adult labor.

In spite of this, it is worth discussing the consequences of possible violations of exogeneity for the other inputs. Notably, it is important to recognize that adult productivity is in itself a result of interest, to which child labor productivity will be compared. If endogeneity issues remain for adult labor, then we should be cautious in the comparison. However, we can provide additional tests to assess the extent of this remaining endogeneity. First, we will test for child labor exogeneity, conditional on household and village by year fixed effects. If exogeneity is not rejected for child labor, then it is less likely that adult labor is endogenous, provided that we also condition on the same fixed effects. Second, neglecting the endogeneity of other inputs could lead to a bias in the child labor estimates under certain circumstances. Since we instrument child labor by  $Z$  ( $Z = N_c$  or  $\frac{N_c}{N_a}$ ) and using the Zellner property, the estimates on child labor can be biased only if  $Z$  is correlated with the other inputs, conditional on household and village-by-year fixed effects. This could be invalidated, for instance, if households anticipate that their children get older and adjust for other inputs based on the increased productivity of children. We return to this question in the next section, providing additional evidence supporting the validity of our approach.

## 1.4 Results

### 1.4.1 OLS estimations

**1.4.1.0.1 Perfect substitutes.** We start with OLS estimations of the three production functions. Table A1.4 provides estimates if child and adult labor are assumed to be perfect substitutes. The  $\alpha$  parameter estimate is 0.656 (coefficient of  $\log L_a$ ) when not controlling for household fixed effects and 0.795 with household fixed effects. This is consistent with decreasing marginal returns to labor (as the coefficients are below 1). The

TABLE 1.2 : AVERAGE SEMI-ELASTICITIES OF LABOR: PERFECT SUBSTITUTE SPECIFICATION (SIMULATIONS FROM TABLE A1.4 ).

	(1)	(2)	(3)	(4)
Adult	-0.000226 (0.000944)	0.00269** (0.00114)	0.00283** (0.00113)	0.00285** (0.00132)
Child	0.00446*** (0.000686)	0.00308*** (0.000865)	0.00287*** (0.000853)	0.00256** (0.000997)
Households F.E		×	×	×
Climatic factors			×	
Village*year F.E				×
Simulation sample	1,482	1,482	1,460	1,482

Note: The estimation sample includes all households with adult labor and cultivated lands while the simulation sample is constrained to households with adult, child labor and cultivated lands. Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means, respectively, that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Climatic factors are the precipitation at date  $t$ , the number of idiosyncratic and covariate shocks.

relative productivity of children  $\gamma_c$  compared to adults is estimated to  $1.24 = 0.815/0.656$  when we do not control for household fixed effects (column 1). However, once controlling for household fixed effects, the relative productivity of children decreases:  $\gamma_c$  is estimated at  $0.71 = 0.564/0.795$ . This suggests that more productive households are the ones who tend to make their children work. This is consistent with previous evidence that households with more work opportunities are the ones employing children (Bhalotra and Heady, 2003; Dumas, 2007). From a methodological point of view, it shows that using panel data to estimate child productivity is crucial. When controlling for observable shocks to production (column 3), the estimate remains the same, while it decreases somewhat when allowing for unobservable shocks at the village level. The other inputs that display significant positive effects in the specification with household fixed effects are: land area, inorganic fertilizer, and pesticides. Covariates shocks have a negative impact on household production. In contrast, rainfall shocks do not, maybe because declared shocks are more accurate. The estimated semi-elasticities, computed with formulas 1.8 and 1.10 are provided in Table 1.2 . They are precisely estimated but we find in this specification that the average productivity of children is similar to the one of adults. On average, one more day of child work is associated to an increase by 0.3% of the production.

**1.4.1.0.2 Imperfect substitutes.** Table A1.5 provides estimates of production function for which adult and child labor are imperfect substitutes. The adult productivity is very close to the one estimated earlier and again, controlling for households fixed effects lowers the estimate of child productivity. Looking at the upper panel of Table 1.3, we find that, on average, one day of adult work is associated with a 0.6% increase in production, while one day of child work is associated with a 0.4% increase in production. The last column, however, gives lower semi-elasticities of both adult and child work. The estimates for the control variables are similar to the ones obtained in the previous sub-section.

Given that the Imperfect model is a more general model of the Perfect model, we can test whether the Perfect model is rejected. The bottom part of Table A1.5 shows that in all specifications but one we can reject the Perfect model (hypothesis that  $\alpha(1+\rho)\gamma_c = 0$ ). Table A1.9, in the Appendix, collects the estimates for  $\rho$  under different specifications but these estimates vary widely and the confidence intervals are large. In addition, allowing for the imperfect term does not improve the explanatory power of the estimation since the R-squared coefficients are the same.

TABLE 1.3 : AVERAGE SEMI-ELASTICITIES OF LABOR: IMPERFECT SUBSTITUTES SPECIFICATION (SIMULATIONS FROM TABLE A1.5 ).

	(1)	(2)	(3)	(4)
Adult	0.00338** (0.00157)	0.00628*** (0.00163)	0.00654*** (0.00160)	0.00343 (0.00232)
Child	0.00575*** (0.000960)	0.00464*** (0.00125)	0.00439*** (0.00124)	0.00281* (0.00150)
Test $\alpha(1+\rho)\gamma_c = 0$	6.130	5.888	5.487	0.0695
Prob>F	0.0138	0.0158	0.0198	0.792
Households F.E		×	×	×
Climatic factors			×	
Village*year F.E				×
Simulation sample	1,482	1,482	1,460	1,482

Note: The estimation sample includes all households with adult labor and cultivated lands while the simulation sample is constrained to households with adult, child labor and cultivated lands. Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means, respectively, that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Climatic factors are the precipitation at date  $t$ , the number of idiosyncratic and covariate shocks.



**1.4.1.0.3 Translog.** Lastly, Table A1.7 provides the estimates for the translog specification and shows that the effects of covariates are the same in this new specification. Table 1.4 shows that the child semi-elasticity obtained from the Translog is strikingly similar to the one obtained from the Imperfect specification.

Based on this specification, we can test whether the Cobb-Douglas specification is rejected. This amounts to testing the joint significance of the interacted terms between  $\ln L_a$  and  $\ln L_c$  ( $\alpha_{aa}$ ,  $\alpha_{cc}$  and  $\alpha_{ac}$  in equation (1.6)), and the test is provided in the bottom panel of Table 1.4 . We clearly reject the Cobb-Douglas specification, as expected.

TABLE 1.4 : AVERAGE SEMI-ELASTICITIES OF LABOR: TRANSLOG SPECIFICATION (SIMULATIONS FROM TABLE A1.7 ).

	(1)	(2)	(3)	(4)
Adult	0.00248*** (0.000820)	0.00463*** (0.000983)	0.00476*** (0.000988)	0.00364*** (0.00116)
Child	0.00575*** (0.00102)	0.00463*** (0.00146)	0.00439*** (0.00146)	0.00268 (0.00171)
$\alpha_{aa} = \alpha_{cc} = \alpha_{ac} = 0$	9.830	5.748	4.913	5.012
Prob>F	3.18e-06	0.000766	0.00237	0.00207
Households F.E		×	×	×
Climatic factors			×	
Village*year F.E				×
Simulation sample	1,482	1,482	1,460	1,482

Note: The estimation sample includes all households with adult labor and cultivated lands while the simulation sample is constrained to households with adult, child labor and cultivated lands. Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means, respectively, that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Climatic factors are the precipitation at date t, the number of idiosyncratic and covariate shocks.

## 1.4.2 First-stage of IV specifications

TABLE 1.5 : EFFECT OF THE NUMBER OF CHILDREN  $N_c$  ON CHILD LABOR (FIRST STAGE)

IV variables	perfect and imperfect		translog	
	(1)	(2)	(3)	(4)
$\frac{N_c}{N_a}$	0.202*** (0.0293)	0.186*** (0.0334)		
$N_c$			0.582*** (0.0606)	0.558*** (0.0619)
R-squared	0.075	0.266	0.081	0.296
F-test	47.47	30.99	92.05	81.25
HH F.E	×	×	×	×
village-year F.E		×		×
Observations	4,922	4,992	4,924	4,994

Note: The estimation sample includes all households with adult labor while the simulation sample is constrained to households with adult and child labor. Standard errors are clustered and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are pesticide and inorganic fertilizer spending, the use of organic fertilizer, the precipitation at date t, the number of idiosyncratic and covariate shocks, the erosion and the irrigation status, the productive assets, the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the gender of the household's head, the access to the plot GPS information, survey month dummies and year dummies.

We now turn to the identification of the child labor productivity based on the exogenous variation provided by children's aging. The results are presented in Table 1.5. We first explore whether the aging of children belonging to the household provides sufficient explanatory power, in spite of the fact that we take into account all children related to the household, and not just present children. There are only two different specifications since the Perfect and Imperfect instrumentation equations are identical. From now on, all estimates control for household fixed effects. In the even-numbered columns of Table 1.5, we control for village-by-year fixed effects, but not in the odd-numbered columns. We find that household composition predicts well the supply of labor, even conditional on household and village-by-year fixed effects. In particular, if adults

have one more child of the age 10-15, it increases child labor days by 0.6. The F-statistics are very high, ranging from 31 to 92 depending on the specification, which indicates no problem of “weak” instrument.<sup>26</sup>

### 1.4.3 Validity tests for IV specifications

Before turning to the estimates of child labor productivity based on this strategy, we assess the validity of our approach. One key aspect is whether the likely endogeneity of the other inputs will bias the estimated child productivity. In an IV estimation, it would be the case if an endogenous input is correlated with the instrumented child labor and if it is correlated with the instrument. One of these likely endogenous inputs is adult labor days, of course, which should be correlated with child labor days. It is therefore important to evaluate whether our instruments are correlated with adult labor days. In Table 1.6, we estimate the effect of labor on other inputs. For a given cell, the reported coefficient is the estimate of the effect of the variable stipulated on the left on the input variable stipulated at the top. Each of these correlations are conditional on household fixed effects and village-by-year fixed effects (and 5 different equations are estimated in each column). For instance, the upper cell provides the parameter  $\theta$  from the following regression:

$$\log L_a = \theta \log L_c + \nu_i + \xi_{vt} + \eta_{it} \quad (1.13)$$

The first two lines show that most inputs “react” to adult and child labor days, which is consistent with households choosing simultaneously the level of all the inputs. However, the correlations with  $N_c$  are mostly no significantly different from 0. The correlation between  $N_c$  and productive assets is significant at the 5% level but as long as productive assets do not have a significant effect on farm production (this is what OLS estimations show, see Table A1.4 for instance), the correlation of our instruments with productive assets does not threaten our identification. The correlation with  $\log L_a$  is significant at the

---

<sup>26</sup>As a robustness check, we instrument child labor by the number of children by age, and we see that child labor is mostly predicted by the number of children aged 13 to 15 (see Table A1.12). The F-statistics are large for the perfect and imperfect specifications, but remain small for the translog function.

5% level, but the relationship between  $L_c$  and  $\log L_a$  seems stronger than the relationship between  $N_c$  and  $\log L_a$ . One cannot directly compare the size of the coefficient with the one in the first line since the two variables ( $L_c$  and  $N_c$ ) are scaled differently. In order to make such a comparison, we express these “effects” in standard deviation of the RHS variable ( $L_c$  and  $N_c$ ). An increase by one standard deviation in  $L_c$  is associated to an increase in  $L_a$  by 8.3%<sup>27</sup> while an increase by one standard deviation in  $N_c$  is associated to a decrease in  $L_a$  by 3.5%. In our case, we do take into account  $L_a$ , but our estimation requires the coefficient on  $L_a$  to be unbiased. However, failure to control appropriately for  $L_a$  leads to a contamination of the endogeneity bias on the child labor parameter in the OLS estimates. In our case, such a contamination via  $L_a$  could exist but should be smaller when instrumenting than when not.<sup>28</sup>

---

<sup>27</sup>The standard deviation in  $L_c$  is 50.89 and the standard deviation in  $N_c$  is 1.058.

<sup>28</sup>Overall, the picture is the same when using  $\frac{N_c}{N_a}$  except for the very precisely estimated correlation between  $\frac{N_c}{N_a}$  and  $\log L_a$ . We are unable to explain this significant correlation, especially since  $N_a$  is not correlated with  $\log L_a$ , conditional on fixed effects.

TABLE 1.6 : EFFECT OF HOUSEHOLD LABOR AND OF HOUSEHOLD DEMOGRAPHIC CHARACTERISTICS ON THE USE OF INPUTS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	log La	log(cult. land)	pesticide	organic fertilizer	inorganic fertilizer	nbr. idio prod. shocks	nbr. cov. prod. shocks	Eroded land	Irrigated land	Productive assets
Household labor										
$\log L_c$	0.00164*** (0.000255)	0.000838*** (0.000226)	0.00185 (0.00122)	-1.49e-05 (0.000148)	0.00101 (0.00117)	-1.50e-05 (9.30e-05)	0.000347 (0.000222)	8.31e-05 (0.000129)	-2.77e-05 (3.13e-05)	0.000475** (0.000209)
$\log L_a$	-	0.251*** (0.0171)	0.480*** (0.0916)	0.0253*** (0.00917)	0.294*** (0.0738)	0.00229 (0.00737)	0.0182 (0.0142)	0.0113 (0.00945)	0.00559* (0.00311)	0.0198 (0.0514)
$\frac{L_c}{L_a}$	-0.326*** (0.0605)	0.0509 (0.0432)	0.345* (0.201)	-0.00487 (0.0227)	-0.0515 (0.160)	0.0101 (0.0233)	-0.0282 (0.0389)	0.0225 (0.0300)	-0.0141 (0.00932)	0.0977** (0.0417)
Household demographic characteristics										
$N_c$	-0.0338* (0.0201)	-0.00938 (0.0151)	-0.0970 (0.103)	-0.00780 (0.00948)	0.117 (0.0773)	0.00314 (0.00764)	-0.00418 (0.0170)	-0.00939 (0.00990)	-0.00540 (0.00389)	0.0424** (0.0202)
$N_a$	0.0217 (0.0141)	-0.00364 (0.0113)	0.0231 (0.0840)	-0.000103 (0.00718)	0.142* (0.0770)	-0.00126 (0.00533)	0.000326 (0.0113)	-0.0122* (0.00655)	0.00261 (0.00196)	0.00782 (0.0164)
$\frac{N_c}{N_a}$	-0.135*** (0.0502)	0.00638 (0.0338)	-0.149 (0.215)	-0.0273 (0.0228)	0.219 (0.167)	-0.00254 (0.0198)	-0.0720* (0.0411)	-0.00991 (0.0239)	-0.00978 (0.00780)	0.0471 (0.0388)
Households F.E	×	×	×	×	×	×	×	×	×	×
village*year F.E	×	×	×	×	×	×	×	×	×	×
Obs.	4,991	4,991	4,991	4,991	4,991	4,991	4,991	4,991	4,991	4,991
Number of households	1,838	1,838	1,838	1,838	1,838	1,838	1,838	1,838	1,838	1,838

Note: Sample: households with adult labor and with cultivated lands. Standard errors are clustered and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Survey month dummies and year dummies are added.

#### 1.4.4 IV estimations

We now turn to the IV estimations, that are performed by the inclusion of the first-stage residual. The full specifications are reported in the Appendix (Table A1.8 ). Table 1.7 reports tests on the specifications and the estimated labor semi-elasticities with IV (but also reports OLS results for the ease of comparison). The test of exogeneity is simply given by the significance of the coefficient of  $\hat{\eta}$ . Quite surprisingly, we never reject the exogeneity hypothesis of child labor. However, we lose substantial amount of precision of the estimates and all estimates become insignificant, including in the Translog specification where the F-statistic was particularly large and which was the least questionable specification according to the previous section. In the specification without village-by-year fixed effects, we find that the semi-elasticity of production with respect to child labor is roughly divided by two when controlling for first-stage residuals. The estimates are consistent across the various specifications: on average, one day of child labor increases production by 0.2%. In the second part of the panel, the precision is again an issue, but the estimates tend to move upwards. One interesting aspect is that the IV estimates without controls for village-by-year fixed effects are strikingly close to the OLS estimates with controls for village-by-year fixed effects. Controlling for village-by-year effects absorbs a large share of unobserved shocks on households, that drive factors allocation, and it might well be that this is sufficient to absorb unobserved heterogeneity. These OLS semi-elasticities of production to child labor range from 0.26% to 0.28%, which is a very small range.<sup>29</sup> By comparison, the estimated adult labor semi-elasticities range from 0.29% to 0.36%. We need to be more cautious about this last set of estimates, for which we have not properly addressed the question of endogeneity. However, the ratio between the two, ranging from 0.712 (Translog) to 0.82 (Perfect), is plausible and points towards a high efficiency of child labor. This might be driven by the fact that children

---

<sup>29</sup>IV estimates are left unchanged when we instrument by the number of children of each age category (see Table 1.7 ). Coefficients are imprecisely measured but remain close to the OLS estimates, especially for the perfect and imperfect specifications. Since the reported residuals are not significant, we do not reject the exogeneity of child labor.

TABLE 1.7 : AVERAGE SEMI-ELASTICITIES OF LABOR, SIMULATION FROM OLS AND IV ESTIMATIONS.

Variables	Perfect		Imperfect		Translog	
	OLS	IV	OLS	IV	OLS	IV
	Without village-year F.E					
Adult	0.00283** (0.00113)	0.00393 (0.00592)	0.00654*** (0.00160)	0.00814 (0.00666)	0.00476*** (0.000988)	0.00472*** (0.000962)
Child	0.00287*** (0.000853)	0.00198 (0.00486)	0.00439*** (0.00124)	0.00315 (0.00477)	0.00439*** (0.00146)	0.00746 (0.00734)
$\widehat{\eta}$		0.173 (0.958)		0.246 (0.977)		-0.0583 (0.136)
$\alpha_{aa} = \alpha_{cc} = \alpha_{ac} = 0$					4.913	4.94
$\alpha(1 + \rho)\gamma_c = 0$			5.49	21.79		
Prob>F			0.0198	0.0000	0.00237	0.0024
Climatic factors	×	×	×	×	×	×
Households F.E	×	×	×	×	×	×
Simulation sample	1460	1460	1460	1460	1460	1460
	With village-year F.E					
Adult	0.00285** (0.00132)	0.00141 (0.00769)	0.00343 (0.00232)	0.00203 (0.00910)	0.00364*** (0.00116)	0.00364*** (0.00109)
Child	0.00256*** (0.000997)	0.00377 (0.00647)	0.00281* (0.00150)	0.00393 (0.00635)	0.00268 (0.00171)	0.00818 (0.00789)
$\widehat{\eta}$		-0.231 (1.274)		-0.218 (1.319)		-0.104 (0.152)
$\alpha_{aa} = \alpha_{cc} = \alpha_{ac} = 0$					5.012	5.76
$\alpha(1 + \rho)\gamma_c = 0$			0.0695	0.0794		
Prob>F			0.792	0.294	0.0024	0.0008
Households F.E	×	×	×	×	×	×
Simulation sample	1482	1482	1482	1482	1482	1482

Note: The estimation sample includes all households with adult labor and cultivated lands while the simulation sample is constrained to households with adult, child labor and cultivated lands. Standard errors reported in parentheses are clustered at the ward level and are bootstrapped for IV estimates. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. For the perfect and the imperfect function, the IV is  $\frac{N_c}{N_a}$ . For the translog function, the IV is  $N_c$ .

working on-farm are only household children while adult work encompasses a more diverse category: household males, households females and non-household members. This last category of workers may have a lower productivity due to lack of incentives. In addition, adult labor includes all labor provided by individuals aged between 16 and 65. It could well be that older individuals are much less productive than younger ones. Last, the tests for comparing the different functional forms give the same results as previously: we reject the Perfect substitutes production function in favor of the Imperfect one, and we cannot reduce the Translog to a Cobb-Douglas. However, given the extreme proximity in the estimates, we do not view this result as extremely important.

## 1.5 How much should children be compensated?

We are now equipped with estimates of different specifications. We can convert them in Tanzanian Shillings (the unit in which the production is measured). Table 1.8 displays the elasticities of production with respect to one day of labor, separately for children and adults. These elasticities are expressed in 2008 TSh and we convert them into 2016 US\$ for ease of comparison.<sup>30</sup> Significant estimates of child productivity lie between 951TSh and 1645TSh (US\$0.91-US\$1.6). As Gertler and Glewwe (1992), we compute the opportunity cost of education for children by comparing the number of days of labor between enrolled and non-enrolled children. On average, non-enrolled children work 26 days more. Therefore, the opportunity costs for which children should be compensated lie between 24,731TSh (US\$23.6) and 42,770TSh (US\$40.8). This is equivalent to monthly payments between 2,061TSh (US \$2) and 3,564TSh (US \$3.4).

As previously discussed, using the child wage on the market to assess the child time opportunity costs is difficult to defend. However, it is interesting to compare the adult agricultural wage with our estimates of adult productivity. In our data, wages can be reported per day or per month, up to the choice of the respondent.<sup>31</sup> The average per

---

<sup>30</sup>See footnote 11 for the computation.

<sup>31</sup>Actually, the period for wage declaration can be reported per hour, day, week, fortnight, month, quarter, half year or year, but most of people either report wage per day or per month, and agricultural



TABLE 1.8 : AVERAGE VALUE OF SEMI-ELASTICITIES IN TANZANIAN SHILLINGS

Variables	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	Perfect IV	Imperfect OLS	Imperfect IV	Translog OLS	Translog IV
Without village-year F.E						
Adult	2,360*** (179.8)	2,151 (6.563e+11)	-2.697e+07 (1.831e+29)	-1.459e+11 (6.364e+18)	1,777*** (278.6)	2,062*** (186.6)
Child	1,020*** (309.8)	732.5 (1,964)	1,645*** (456.5)	1,278 (6,718)	1,113 (1,098)	-8,473 (12,345)
Simulation Sample	1,460	1,460	1,460	1,460	1,460	1,460
With village-year F.E						
Adult	2,297*** (205.1)	2,212 (1.045e+25)	2,259 (1.606e+25)	2,035 (5.598e+33)	2,011*** (222.2)	1,837*** (265.9)
Child	951.2** (462.2)	627.7 (3,141)	1,178 (851.8)	828.2 (3,688)	290.1 (1,737)	-14,999 (41,766)
Households FE	×	×	×	×	×	×
Simulation Sample	1,482	1,482	1,482	1,482	1,482	1,482

Note: The estimation sample includes all households with adult labor while the simulation sample is constrained to households with adult and child labor. Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are pesticide and inorganic fertilizer spending, the use of organic fertilizer, the precipitation at date t, the number of idiosyncratic and covariate shocks, the erosion and the irrigation status, the productive assets, the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the gender of the household's head, the access to the plot GPS information, survey month dummies and year dummies.

day wage among adults who are paid to the day is 4,818TSh (2016 US\$4.6).<sup>32</sup> This is slightly higher than most of the adult elasticities obtained in Table 1.8. This is expected since the existence of labor market imperfections would lead to a discrepancy between observed wages and shadow wages. In particular, household workers would presumably have difficulties being hired full-time on the market and taking into account the likelihood of remaining unemployed a given day reduces the opportunity cost of time, and increases on-farm employment. This discrepancy confirms that we cannot completely rely on child wages to calculate the opportunity costs of child time, but the difference between adult wages and adults estimated productivity suggests that our estimates are meaningful.

wages are most of the time reported per day.

<sup>32</sup>When doing a similar computation for adults paid to the month, we obtain a monthly average agricultural wage of 96,334TSh, which corresponds to 20 days of the day wage. Therefore, there does not seem to be a striking difference between pay per day and pay per month since 20 days of work in a month seems sensible.

## 1.6 Heterogeneity in productivity

We now explore the heterogeneity in productivity by child characteristics.

### 1.6.1 Specification

We do so based on the Perfect specification. While we have shown that the data reject the hypothesis that both types of labor are perfect substitutes, we have also seen that the estimates are strikingly close from one functional form to the other. Testing for heterogeneity by children's characteristics is straightforward with the Perfect functional form. We explore the productivity differences of children by gender and by age.

Starting with equation 1.3, but allowing for a different  $\gamma_c$  by child's gender, and then linearly approximating, we obtain:

$$\log Y = \log A + \alpha \log(L_a + \gamma_b L_b + \gamma_g L_g) + (\log X)\beta \quad (1.14)$$

$$\log Y \approx \log A + \alpha \log L_a + \alpha \gamma_b \frac{L_b}{L_a} + \alpha \gamma_g \frac{L_g}{L_a} + (\log X)\beta \quad (1.15)$$

$$\approx \log A + \alpha \log L_a + \alpha \gamma_b \frac{L_c}{L_a} + \alpha(\gamma_g - \gamma_b) \sum_k 1(k = girl) \frac{L_k}{L_a} + (\log X)\beta \quad (1.16)$$

where  $L_b$  stands for labor of boys (10-15 y.o.),  $L_g$  stands for labor of girls and  $L_k$  is the amount of work done by child  $k$ . For identification, we have two options: first we can instrument each type of child work separately, as suggested by equation 1.15. In this case  $N_b/N_a$  and  $N_g/N_a$ , respectively the ratio of number of boys to adult members and the ratio of number of girls to adult members, are good candidates. Alternatively, we may simply instrument for  $L_c$  and assume that the allocation of work to children of different genders is exogenous (equation 1.16).<sup>33</sup>

For the heterogeneity by age, we allow for a parsimonious specification where productivity depends linearly in age. For limited variations in age, this should be a

---

<sup>33</sup>The two equations are strictly similar, but for the sake of simplicity we refer to them as two different specifications in which the treatment of endogeneity is not the same.

reasonable assumption.

$$\log Y = \log A + \alpha \log \left( L_a + \sum_{o=10}^{o=15} (\gamma_c + \gamma_{c'} \cdot (o - 15)) L_{co} \right) + (\log X)\beta \quad (1.17)$$

$$\log Y \approx \log A + \alpha \log L_a + \alpha \sum_{o=10}^{o=15} (\gamma_c + \gamma_{c'} \cdot (o - 15)) \frac{L_{co}}{L_a} + (\log X)\beta \quad (1.18)$$

$$\approx \log A + \alpha \log L_a + \alpha \gamma_c \frac{L_c}{L_a} + \alpha \gamma_{c'} \sum_{o=10}^{o=15} (o - 15) \cdot \frac{L_{co}}{L_a} + (\log X)\beta \quad (1.19)$$

where  $L_{co}$  stands for labor time provided by children of age  $o$ . In this specification, the productivity of a 15-year old child relative to an adult is  $\gamma_c$  and this relative productivity decreases by  $\gamma_{c'}$  for each year below 15. From there, we can implement the same strategy as before: control for household fixed effects, village-by-year fixed effects, other covariates, as well as instrumentation of the RHS child labor variables. Given that the data do not reject the hypothesis of exogeneity in the allocation of child labor, we simply assume that the share of child work by age is also exogenous. We therefore simply implement the same identification strategy as before except that we allow for a heterogeneity term by age.

## 1.6.2 Results

Table 1.9 provides the semi-elasticities of production to child days of work, depending on their gender. The first panel is devoted to the specification with the two instruments (eq. 1.15), while the second panel is devoted to the specification where only  $L_c$  is instrumented (eq. 1.16). The first-stages are provided in Table A1.11, in the Appendix. The F-statistics are above 40 but they are notably higher for boys than for girls. It might be that there is more predetermination of the child activity when this child is male. From the IV columns, it is clear that whether instrumenting the two variables or only the child labor leads to exactly the same conclusion: the exogeneity is not rejected and the parameters are strikingly close to each other. We do not reject the exogeneity assumption. When relying on the semi-elasticities obtained from the OLS estimates, we find that boys are markedly more productive than girls. This could be either due to differences in strength, or to the fact that girls spend only a share of their day in the field (because they have

TABLE 1.9 : AVERAGE SEMI-ELASTICITIES WITH GENDER HETEROGENEITY  
(PERFECT SUBSTITUTES)

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS		IV for eq. (1.15)		IV for eq. (1.16)	
Girls	0.00169* (0.00101)	0.00124 (0.00138)	0.00228 (0.00538)	0.000415 (0.00680)	0.000662 (0.00478)	0.00241 (0.00644)
Boys	0.00479*** (0.00142)	0.00418*** (0.00158)	0.00246 (0.00547)	0.00439 (0.00600)	0.00365 (0.00508)	0.00534 (0.00665)
$\hat{\eta}_g$			-0.127 (1.060)	0.166 (1.324)		
$\hat{\eta}_b$			0.406 (1.058)	-0.0319 (1.129)		
$\hat{\eta}$					0.168 (0.960)	-0.224 (1.274)
F-test girls			59.40	43.23		
F-test boys			91.95	93.83		
F-test					54.47	31.14
Households FE	×	×	×	×	×	×
Climatic factors	×		×		×	
Village*year F.E		×		×		×
Simulation sample	1460	1482	1460	1482	1460	1482

Note: The estimation sample includes all households with adult labor and cultivated lands while the simulation sample is constrained to households with adult, child labor and cultivated lands. Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Climatic factors are the precipitation at date t, the number of idiosyncratic and covariate shocks, the mean temperature and the greenness increase in days.

In equation (1.15),  $\frac{L_b}{L_a}$  and  $\frac{L_g}{L_a}$  are separately instrumented while in equation (1.16),  $\frac{L_c}{L_a}$  is instrumented.

to also provide domestic work, for instance) while boys spend their full day on the plot. We do not have the information on number of hours per day spent farming on an annual basis. However, we know the number of hours devoted to each activity in the week before the survey. Figure A1.3 shows that while boys spend more than 5 hours per week on agricultural tasks, girls only spend less than 4 hours. We do not have the information of the number of *days* spent farming in the previous week, so we cannot compute a number of hours per day of work. However, recall that girls spend on average 54.3 days per year conditional on participation against 48.9 days for boys. If we assume that past week information is relevant for inferring children’s activities on an annual basis, then this confirms that girls actually spend less time on the plots than boys for a given number of days and explains the productivity differential.

TABLE 1.10 : MEAN OF SEMI-ELASTICITIES FROM PERFECT SPECIFICATION WITH AGE HETEROGENEITY (EQUATION 1.19)

	(1)	(2)	(3)	(4)	(5)	(6)
Age	10	11	12	13	14	15
Without village-year F.E						
OLS	0.00197 (0.00152)	0.00228** (0.00114)	0.00258*** (0.000880)	0.00288*** (0.000852)	0.00318*** (0.00108)	0.00349** (0.00144)
IV: $\frac{N_c}{N_a}$	0.00223 (0.00656)	0.00264 (0.00643)	0.00305 (0.00638)	0.00345 (0.00642)	0.00386 (0.00655)	0.00427 (0.00677)
F-test				47.47		
$\hat{\eta}$				-0.169 ( 1.259)		
Simulation sample	1460	1460	1460	1460	1460	1460
With village-year F.E						
OLS	0.00149 (0.00194)	0.00185 (0.00137)	0.00221** (0.000978)	0.00258*** (0.001000)	0.00294** (0.00142)	0.00331* (0.00200)
IV: $\frac{N_c}{N_a}$	0.00262 (0.00672)	0.00298 (0.00654)	0.00333 (0.00645)	0.00369 (0.00645)	0.00405 (0.00654)	0.00440 (0.00671)
F-test				31.14		
$\hat{\eta}$				-0.115 (1.249)		
Simulation sample	1482	1482	1482	1482	1482	1482
FE	×	×	×	×	×	×

Note: The estimation sample includes all households with adult labor while the simulation sample is constrained to households with adult and child labor. Standard errors are clustered and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are pesticide and inorganic fertilizer spending, the use of organic fertilizer, the precipitation at date  $t$ , the number of idiosyncratic and covariate shocks, the erosion and the irrigation status, the productive assets, the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the age of the household’s head, the access to the plot GPS information, survey month dummies and year dummies.

Table 1.10 provides the semi-elasticities by child age. We find again that the IV are close to the OLS estimates and there is little change depending on whether we allow for village-by-year fixed effects. We confirm that older children are more productive than younger ones. Fifteen year old children are 63% more productive than 11 year old and the semi-elasticity for 10 year old children is not significantly different from 0. This is consistent with the fact that few children younger than 10 work on farm.

This also informs us of the validity of our instrumentation strategy. Indeed, entry in and exit out of the age range is also associated to changes in the other inputs if a) children younger than 10 actually contribute or b) children older than 15 have a different productivity than adults. Given that children younger than 10 have been shown not to be economically productive, a) is ruled out. We also provide in Table A1.13, in the Appendix, a more flexible specification that confirms the results. We also find that the productivity of 15 years old children is not significantly different from adults productivity, which rules out b). The instruments therefore seem to satisfy the exclusion restriction.

## 1.7 Conclusion

In this paper, we estimate the economic contribution of children as unpaid family workers on Tanzanian farms. Firstly, we exploit the panel dimension of the data to control for unobserved permanent household and time-varying village characteristics. Then, we instrument child labor by the variation in the number of child household members over time.

To test whether these estimations are sensitive to the functional form, we provide three specifications that assume different degrees of substitution between child labor and other agricultural inputs. Results show that one day of labor performed by a child aged 10 to 15 increases the value of output between US \$0.91 and US \$1.6. This informs us about the opportunity costs of children on the farms. To our knowledge, this paper is the first to provide a daily estimate of the children's shadow wage as unpaid family workers on the farm which is the most common form of child labor in African countries. We also explore

the heterogeneity of the children's productivity by gender and by age. It appears that children start being productive from 12 years old and reach an adult's productivity by 15 years old. In addition, the fact that boys spend more time in the fields and contribute less than girls to domestic activities may explain why their daily productivity is higher.

Finally, to compensate the additional days of labor that unenrolled children are able to perform, we argue that the households should receive between US \$2 and US \$3.4 per month. Our findings are especially relevant when assessing the value of conditional cash transfer programs that aim at compensating the opportunity cost of education that can hinder the demand for education.

# Appendix A1

## 1.7.1 Variables definition

**Household's production:** This variable is the sum of the household's production from all the cultivated crops growing on different plots. Each crop's production is the product of the cultivated quantity and the median price. The median price is computed at the district level. The value of this production is estimated in current Tanzanian shillings. (TSH).

**Cultivated area:** Total surface area (in acres) of cultivated lands by household.

**Adult labor:** Number of farming days in the last year performed by adults aged 16 to 65 in the household.

**Child labor:** Number of farming days in the last year performed by children aged 10 to 15 in the household.

**Rainfall t1:** Refer to the datasets part.

**Types of climat:** This variable informs about the type of climat (1.Tropical-warm and semiarid, 2.Tropical-warm and subhumid, 3.Tropical-warm and humid, 4.Tropical-cool and semiarid, 5.Tropical-cool and subhumid and 6.Tropical-cool and humid).

**Temperature:** This variable is the annual mean temperature in celsius times 10.

**Greenness increase:** This variable indicates the average change in greenness (integral of daily EVI values) during growing season.

**Quality of land:** Dummy variables are introduced to precise if the quality of lands is good, mix (depending on the plot), medium or bad.

**Number of idiosyncratic productivity shocks:** Number of productivity shocks (drought or floods, crop disease or crop pests and severe water shortages) that severely affected the household or few households in the village between 2007-2008.



**Number of covariate productivity shocks:** Number of productivity shocks (drought or floods, crop disease or crop pests and severe water shortages) that severely affected most or all households in the community between 2007-2008.

**Land erosion:** Dummy variable informing if the household had an erosion issue for at least one plot during the year of the survey.

**Land irrigation:** Dummy variable informing if the household has at least one irrigated plot.

**Organic fertilizer:** Dummy variable informing if the household uses organic fertilizer for at least one plot.

**Inorganic fertilizer:** Variable indicating the total spending (in Tanzanian Shillings) for inorganic fertilizer during the year of the survey.

**Pesticide spending:** Variable indicating the total spending (in Tanzanian Shillings) for pesticide during the year of the survey.

**Gender of the household's head:** Dummy variable indicating if the household's head is a woman.

**Repartition of crops:** Variables indicating the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton (these crops constitute the ten main crops in Tanzania).

## 1.7.2 Additional tables and figures

TABLE A1.1 : DESCRIPTIVE STATISTICS: CHARACTERISTICS OF AGRICULTURAL HOUSEHOLDS

VARIABLES	(1) N	(2) mean	(3) sd	(4) min	(5) max
$\log^M(\text{cropproduction})$	4993	11 .112	3 .698	0.693	18 .163
$\log^M(\text{cultivatearea})$	4993	1.162	0.965	0.693	6 .405
$\log^M(La)$	4993	4 .881	0.958	0.750	7 .904
$\log^M(Lc)$	4993	0.517	1.984	0.693	6 .469
$N_c$	4993	1 .078	1.192	0	16
$N_a$	4993	3 .588	2 .372	0	28
Organic fertilizer	4993	0.220	0.414	0	1
log( inorg. fertilizer)	4993	1.295	4 .389	0.693	16 .051
log( pesticide)	4993	0.866	3 .726	0.693	13 .764
precipitation t1	4993	0.017	0.809	2 .077	2 .663
NBR idio. shocks	4993	0.065	0.279	0	3
NBR covariate shocks	4993	0.270	0.514	0	3
Mean temperature	4922	227.388	26 .845	146	278
Greenness increase	4951	129.238	18 .758	12 .8	166.000
Productive assets	4993	0.077	1.838	-0.342	113.54
Percentage of eroded lands	4995	16.8	0.374	0	1
Percentage of irrigated lands	499	3.0	0.170	0	1
Age of the household head	4993	48.835	14 .449	19	107
Percentage of maize	4993	0.446	0.380	0	1
Percentage of paddy	4993	0.101	0.232	0	1
Percentage of sorghum	4993	0.050	0.173	0	1
Percentage of cassave	4993	0.019	0.110	0	1
Percentage of sweet potatoes	4993	0.030	0.119	0	1
Percentage of beans	4993	0.136	0.275	0	1
Percentage of cowpeas	4993	0.028	0.127	0	1
Percentage of sunflower	4993	0.030	0.133	0	1
Percentage of groundnuts	4993	0.053	0.159	0	1
Percentage of cotton	4993	0.016	0.101	0	1

Source: LSMS-ISA data and rainfall data.

TABLE A1.2 : NUMBER OF DAYS OF FARM LABOR PER HOUSEHOLD, BY CATEGORY

Year	2008	2010	2012	3 waves
Number of days of child labor (10-15)	18.75	16.46	18.61	17.91
Number of days of total adult labor	215.90	184.58	197.71	199.41
Number of days of household adult labor	200.62	170.36	180.37	183.84
Number of days of non-household adult labor	15.27	14.22	17.35	15.57

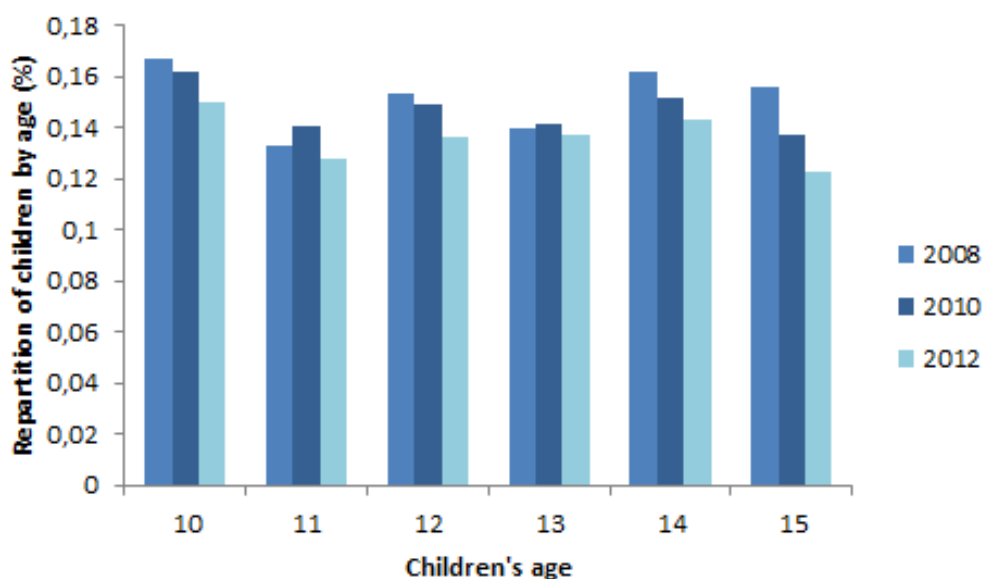
Source: LSMS-ISA data. Sample: Households with at least adult labor and with cultivated plots.

TABLE A1.3 : PERCENTAGE OF CHILDREN WITHOUT PARENT LIVING IN THE HOUSEHOLD.

Age	2008	2010	2012
10	12.0	18.4	13.6
11	10.6	11.1	12.2
12	9.8	11.1	16.1
13	11.7	0.9	14.3
14	9.6	11.1	11.8
15	10.0	13.1	12.1

Sample: households with adult labor and with cultivated lands.

FIGURE A1.1: DISTRIBUTION OF CHILDREN WITHIN THE POOL OF CHILDREN.



Source: LSMS-ISA data (2008, 2010, 2012).

TABLE A1.4 : OLS ESTIMATIONS: PERFECT SUBSTITUTE SPECIFICATION

	(1)	(2)	(3)	(4)
$\log L_a$	0.656*** (0.0712)	0.795*** (0.0933)	0.783*** (0.0903)	0.735*** (0.101)
$\frac{L_c}{L_a}$	0.815*** (0.125)	0.564*** (0.158)	0.527*** (0.156)	0.469** (0.182)
$\log(\text{cultivated area})$	1.088*** (0.0729)	0.695*** (0.103)	0.698*** (0.104)	0.551*** (0.109)
org. fertilizer	0.425*** (0.126)	0.116 (0.140)	0.122 (0.140)	0.250 (0.170)
$\log(\text{inorg})$	0.109*** (0.0112)	0.0658*** (0.0175)	0.0660*** (0.0176)	0.0693*** (0.0202)
$\log(\text{pesticide})$	0.0238 (0.0162)	0.0292** (0.0144)	0.0287** (0.0146)	0.0157 (0.0168)
Land eroded	0.230* (0.118)	-0.0813 (0.123)	-0.0735 (0.124)	-0.0350 (0.150)
Land irrigated	0.634** (0.250)	0.506 (0.391)	0.520 (0.388)	0.345 (0.408)
Productive assets	0.00328 (0.0113)	-0.00534 (0.0117)	0.0106 (0.0124)	0.0350 (0.0603)
NBR idio. shocks	-0.0174 (0.162)	0.00451 (0.207)	0.0453 (0.205)	-0.138 (0.227)
NBR covariate shocks	-0.136* (0.0793)	-0.190** (0.0879)	-0.198** (0.0887)	-0.0108 (0.101)
Rainfall			0.0516 (0.0734)	
mean temperature			0.0412 (0.0438)	
Greenness increase			0.0149** (0.00690)	
R-squared	0.379			
Within HH R-squared		0.233	0.232	0.456
Households F.E		×	×	×
Village*year F.E				×
Observations	4,994	4,994	4,924	4,994
Number of Households		1,837	1,832	1,837

Sample: households with adult labor and with cultivated lands. Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. We also control by the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the gender of the household's head, the access to the plot GPS information, the age of the household head, survey month dummies and year dummies.

TABLE A1.5 : OLS ESTIMATIONS: IMPERFECT SUBSTITUTE SPECIFICATION

	(1)	(2)	(3)	(4)
$\log L_a$	0.652*** (0.0711)	0.793*** (0.0932)	0.781*** (0.0901)	0.735*** (0.101)
$\frac{L_c}{L_a}$	0.516*** (0.145)	0.268* (0.149)	0.242 (0.148)	0.425* (0.218)
$\frac{L_c}{L_a} * (1 - \log \frac{L_c}{L_a})$	0.512** (0.207)	0.555** (0.229)	0.537** (0.229)	0.0846 (0.321)
$\log(\text{cultivated area})$	1.084*** (0.0729)	0.691*** (0.103)	0.694*** (0.104)	0.550*** (0.109)
org. fertilizer	0.420*** (0.126)	0.121 (0.140)	0.127 (0.139)	0.250 (0.170)
$\log(\text{inorg})$	0.109*** (0.0111)	0.0656*** (0.0175)	0.0658*** (0.0176)	0.0693*** (0.0202)
$\log(\text{pesticide})$	0.0242 (0.0162)	0.0290** (0.0144)	0.0287* (0.0146)	0.0157 (0.0168)
Land eroded	0.230* (0.117)	-0.0785 (0.123)	-0.0705 (0.124)	-0.0348 (0.150)
Land irrigated	0.638** (0.249)	0.515 (0.388)	0.529 (0.385)	0.347 (0.406)
Productive assets	0.00336 (0.0113)	-0.00547 (0.0117)	0.0103 (0.0124)	0.0349 (0.0605)
NBR idio. shocks	-0.0174 (0.161)	0.00315 (0.207)	0.0438 (0.205)	-0.138 (0.227)
NBR covariate shocks	-0.140* (0.0793)	-0.195** (0.0881)	-0.203** (0.0889)	-0.0117 (0.101)
Rainfall			0.0516 (0.0732)	
mean temperature			0.0415 (0.0436)	
Greenness increase			0.0147** (0.00688)	
R-squared	0.380			
Within HH R-squared		0.233	0.232	0.456
Households F.E		×	×	×
Village*year F.E				×
Observations	4,994	4,994	4,922	4,994
Number of Households		1,837	1,832	1,837

Note: Sample: households with adult labor and with cultivated lands. Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. We also control by the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the gender of the household's head, the access to the plot GPS information, the age of the household head, survey month dummies and year dummies.

TABLE A1.6 : OLS ESTIMATIONS: TRANSLOG SPECIFICATION

	(1)	(2)	(3)	(4)
log $L_a$	1.184*** (0.370)	1.447*** (0.502)	1.391*** (0.521)	1.507*** (0.528)
log $L_c$	0.646*** (0.120)	0.562*** (0.144)	0.524*** (0.144)	0.552*** (0.156)
log $L_a$ *log $L_a$	-0.0582 (0.0372)	-0.0700 (0.0504)	-0.0651 (0.0523)	-0.0828 (0.0533)
log $L_c$ *log $L_c$	-0.00729 (0.0152)	-0.0280 (0.0184)	-0.0301 (0.0184)	-0.0113 (0.0205)
log $L_a$ *log $L_c$	-0.0990*** (0.0243)	-0.0714*** (0.0272)	-0.0631** (0.0265)	-0.0885*** (0.0304)
log(cultivated area)	1.084*** (0.0729)	0.684*** (0.103)	0.685*** (0.104)	0.540*** (0.110)
org. fertilizer	0.393*** (0.127)	0.102 (0.140)	0.105 (0.139)	0.245 (0.171)
log (inorg)	0.108*** (0.0110)	0.0664*** (0.0177)	0.0665*** (0.0178)	0.0699*** (0.0202)
log (pesticide)	0.0265 (0.0162)	0.0327** (0.0144)	0.0323** (0.0146)	0.0190 (0.0167)
Land eroded	0.231** (0.116)	-0.0917 (0.123)	-0.0840 (0.125)	-0.0419 (0.150)
Land irrigated	0.625** (0.251)	0.539 (0.380)	0.552 (0.378)	0.384 (0.401)
Productive assets	0.00654 (0.0127)	-0.00469 (0.0120)	0.0102 (0.0126)	0.0380 (0.0633)
NBR idio. shocks	-0.0433 (0.160)	-0.0206 (0.207)	.0204 (0.2052)	-0.1589 (0.2268)
NBR covariate shocks	-0.129 (0.0792)	-0.184**-0.1905 (0.0879)	-0.0101 (0.0888)	 (0.1012)
Rainfall			0.0598 (0.0726)	
Mean temperature			0.0466 (0.0447)	
Greenness increase			0.0138** (0.00686)	
R-squared	0.383			
Within HH R-squared		0.239	0.236	0.459
Households F.E		×	×	×
Village*year F.E				×
Observations	4,994	4,994	4,924	4,994
Number of Households	1,837	1,837	1,832	1,837

Note: Sample: households with adult labor and with cultivated lands. Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. We also control by the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the gender of the household's head, the access to the plot GPS information, the age of the household head, survey month dummies and year dummies.

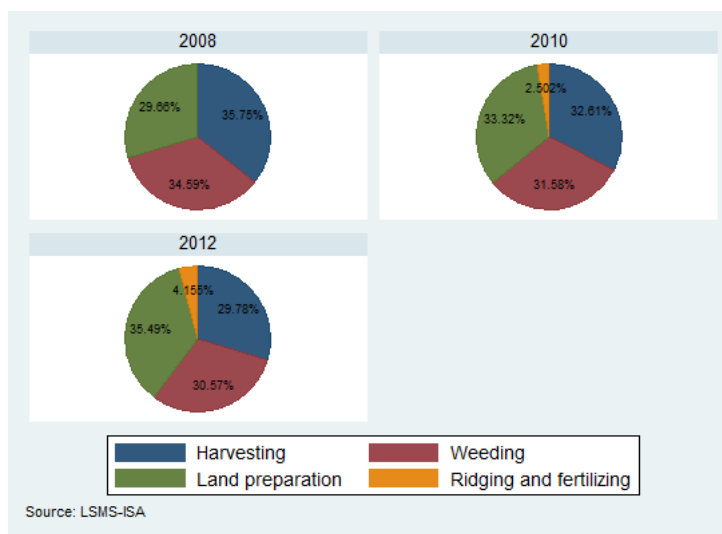
TABLE A1.7 : OLS ESTIMATIONS: TRANSLOG SPECIFICATION WITH INTERACTION TERMS BETWEEN LABOR AND OTHER INPUTS

	(1)	(2)	(3)
$\log L_a$	1.102*** (0.366)	1.399*** (0.497)	1.451*** (0.519)
$\log L_c$	0.677*** (0.119)	0.583*** (0.147)	0.550*** (0.160)
$\log L_a * \log L_c$	-0.102*** (0.0236)	-0.0728*** (0.0279)	-0.0852*** (0.0310)
$\log L_a * \log L_a$	-0.0454 (0.0364)	-0.0617 (0.0498)	-0.0719 (0.0523)
$\log L_c * \log L_c$	-0.0108 (0.0152)	-0.0299 (0.0182)	-0.0141 (0.0205)
$\log L_a * \text{Small Agricultural assets}$	0.00482 (0.0280)	0.0254 (0.0364)	-0.0386 (0.0564)
$\log L_a * \text{Machine assets}$	0.0405** (0.0182)	0.0404*** (0.0150)	0.0451** (0.0219)
$\log L_a * \ln(\text{pesticide})$	-0.0475*** (0.0178)	-0.0330** (0.0159)	-0.0160 (0.0178)
$\log L_a * \ln(\text{fertilizer})$	-0.00124 (0.0118)	-0.00556 (0.0182)	-0.0252 (0.0184)
$\log L_c * \text{Small Agricultural assets}$	-0.00442 (0.0137)	0.00767 (0.0164)	0.0125 (0.0205)
$\log L_c * \text{Machine assets}$	0.00719 (0.0107)	0.000248 (0.0129)	-0.0113 (0.0159)
$\log L_c * \ln(\text{pesticide})$	0.0125** (0.00547)	0.00453 (0.00531)	0.00212 (0.00597)
$\log L_c * \ln(\text{fertilizer})$	-0.0120*** (0.00336)	-0.00881** (0.00365)	-0.00494 (0.00448)
Test interaction terms	6.225	3.987	2.457
Prob>F	2.56e-09	1.82e-05	0.00581
R-squared	0.386		
Within HH R-squared		0.241	0.471
Households F.E		×	×
Village*year F.E			×
Observations	4,990	4,990	4,990
Number of households	1,837	1,837	1,837

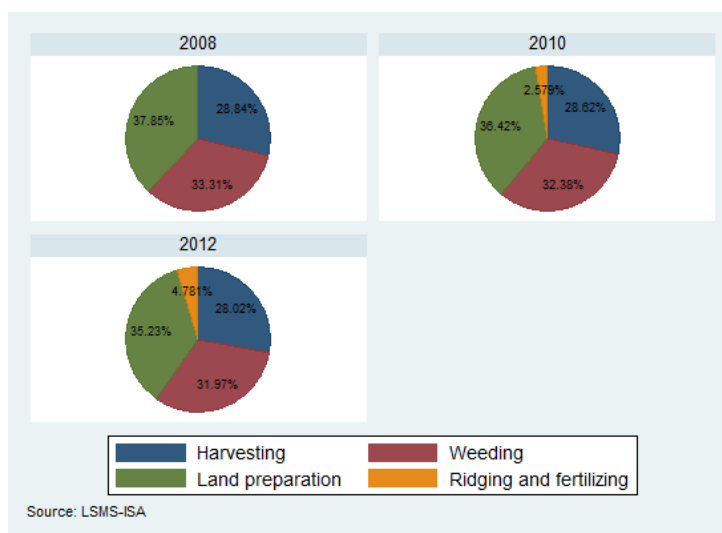
Note: Sample: households with adult labor and with cultivated lands. Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. We also control by the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the gender of the household's head, the access to the plot GPS information, the age of the household head, survey month dummies and year dummies.

FIGURE A1.2: ACTIVITIES IN AGRICULTURAL WORK

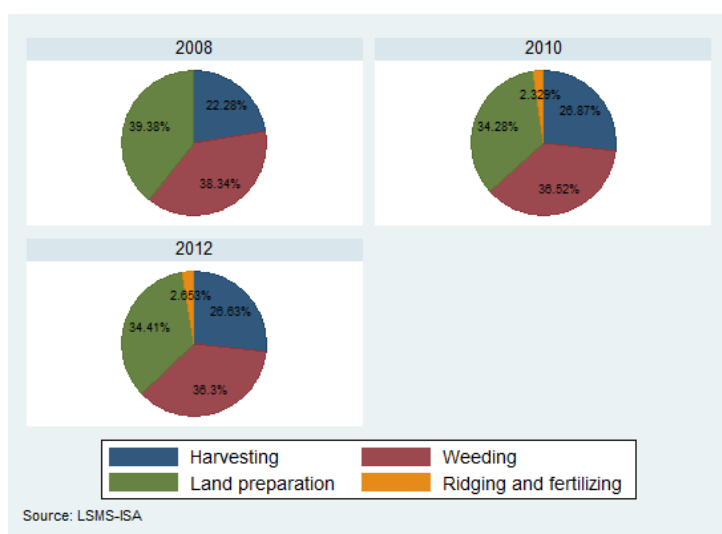
(A) CHILDREN



(B) HOUSEHOLD ADULTS



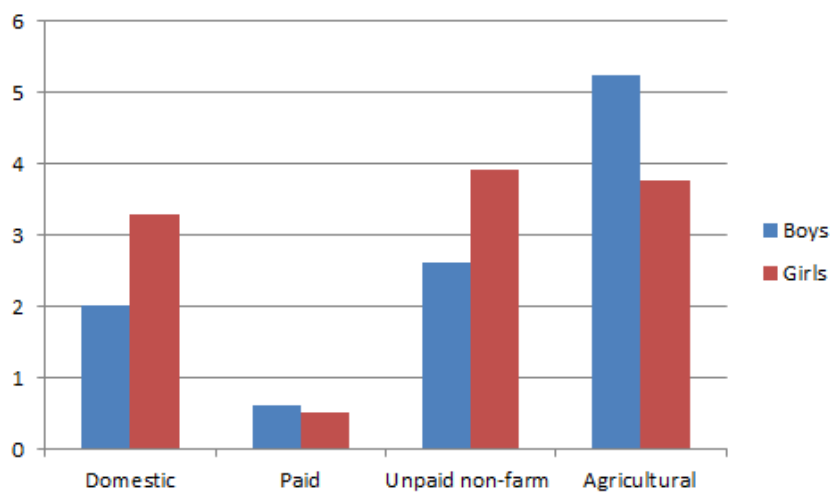
(C) NON-HOUSEHOLD ADULTS



Note: "Ridging and fertilizing" was not an available category in 2008.

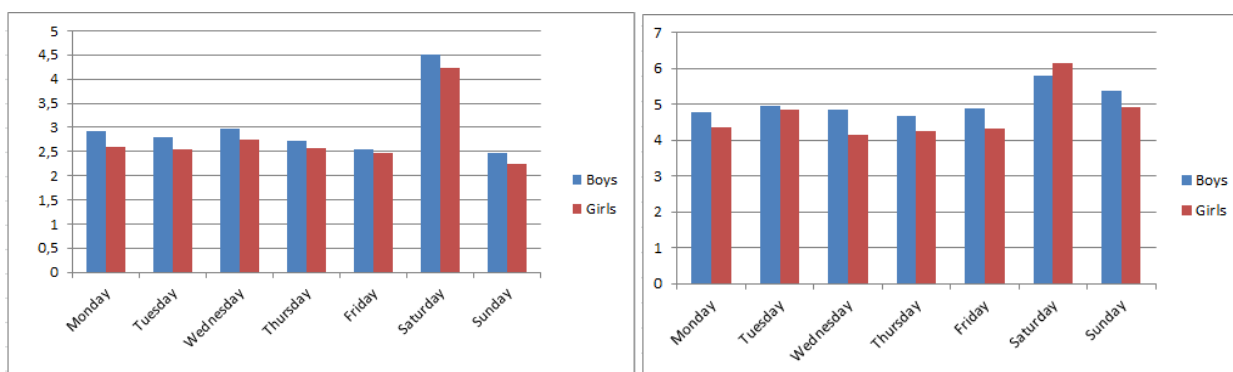


FIGURE A1.3: TIME SPENT AT WORKING BY ACTIVITIES OVER THE LAST WEEEEK



Source: LSMS-ISA data (2008, 2010, 2012). Sample: Active children from households who farm lands.

FIGURE A1.4: TIME SPENT AT WORKING (IN HOURS)

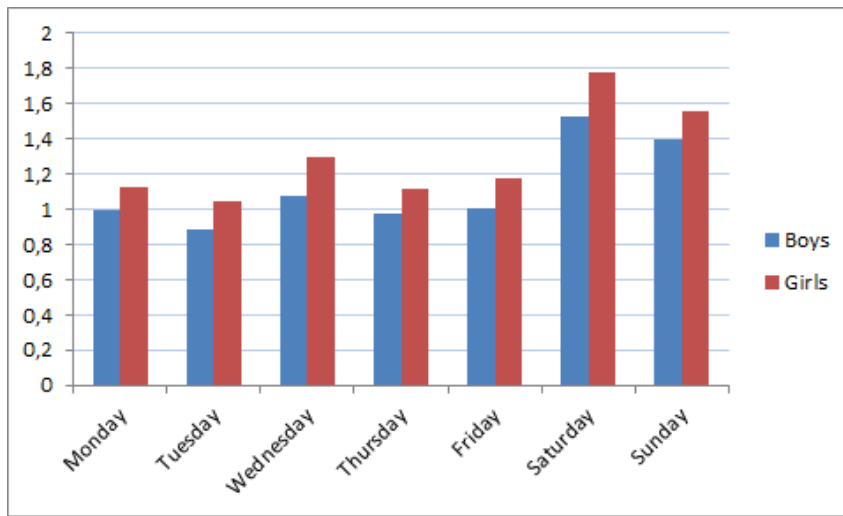


(A) IN THE FIELDS

(B) OUTSIDE THE FIELDS

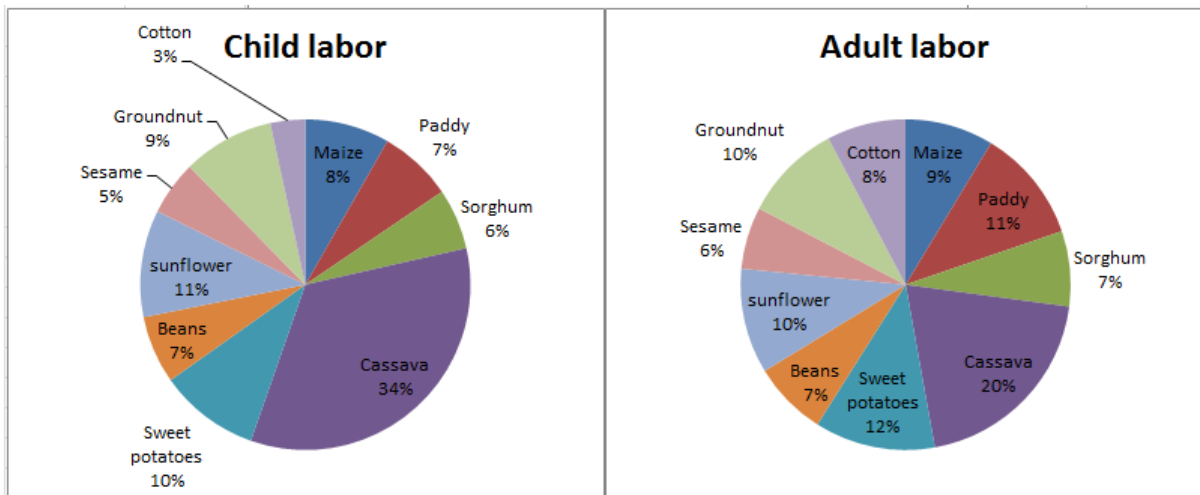
Source: LFS survey, 2014. Sample: Active children from households who farm lands.

FIGURE A1.5: TIME SPENT IN DOMESTIC TASKS (IN HOURS)



Source: LFS survey, 2014. Sample: Active children from households who farm lands.

FIGURE A1.6: LABOR DISTRIBUTION AMONG MAIN CROPS



Source: LSMS-ISA data. Sample: Computation are made from plots without intercropping.

TABLE A1.8 : IV ESTIMATIONS OF THE PRODUCTION FUNCTION

VARIABLES	Perfect		Imperfect		Translog	
	(1)	(2)	(3)	(4)	(5)	(6)
$\log L_a$	0.906*** (0.113)	0.906*** (0.113)	1.039*** (0.132)	1.039*** (0.132)	1.443** (0.566)	1.443** (0.566)
$\frac{L_c}{L_a}$	0.259 (0.845)	0.259 (0.845)	0.151 (0.886)	0.151 (0.886)		
$\frac{L_c}{L_a} * (1 - \log \frac{L_c}{L_a})$			0.243 (0.223)	0.243 (0.223)		
$\log L_c$					0.735*** (0.208)	0.735*** (0.208)
$\log L_a * \log L_a$					-0.0428 (0.0565)	-0.0428 (0.0565)
$\log L_c * \log L_c$					-0.0502*** (0.0171)	-0.0502*** (0.0171)
$\log L_a * \log L_c$					-0.0633*** (0.0235)	-0.0633*** (0.0235)
$\log$ (cultivated area)	0.0694 (0.150)	0.0694 (0.150)	0.0847 (0.139)	0.0847 (0.139)	0.0539 (0.125)	0.0539 (0.125)
org. fertilizer	0.0847 (0.149)	0.0847 (0.149)	-0.00127 (0.136)	-0.00127 (0.136)	-0.0374 (0.143)	-0.0374 (0.143)
$\log$ (inorg)	0.0757*** (0.0187)	0.0757*** (0.0187)	0.0716*** (0.0189)	0.0716*** (0.0189)	0.0713*** (0.0192)	0.0713*** (0.0192)
$\log$ (pesticide)	0.0424** (0.0166)	0.0424** (0.0166)	0.0256* (0.0153)	0.0256* (0.0153)	0.0264 (0.0184)	0.0264 (0.0184)
Land eroded	-0.142 (0.130)	-0.142 (0.130)	-0.191 (0.141)	-0.191 (0.141)	-0.216 (0.136)	-0.216 (0.136)
Land irrigated	0.372 (0.396)	0.372 (0.396)	0.388 (0.398)	0.388 (0.398)	0.449 (0.430)	0.449 (0.430)
Productive assets	0.00759 (0.0595)	0.00759 (0.0595)	0.00623 (0.0678)	0.00623 (0.0678)	0.00151 (0.0502)	0.00151 (0.0502)
NBR idio. shocks	0.0567 (0.196)	0.0567 (0.196)	-0.0210 (0.200)	-0.0210 (0.200)	-0.0700 (0.210)	-0.0700 (0.210)
NBR covariate shocks	-0.134 (0.0950)	-0.134 (0.0950)	-0.102 (0.103)	-0.102 (0.103)	-0.0844 (0.0954)	-0.0844 (0.0954)
Rainfall	0.0475 (0.0881)	0.0475 (0.0881)	0.0794 (0.0878)	0.0794 (0.0878)	0.0886 (0.0755)	0.0886 (0.0755)
mean temperature	0.0233 (0.0472)	0.0233 (0.0472)	0.0504 (0.0476)	0.0504 (0.0476)	0.0625 (0.0476)	0.0625 (0.0476)
Greenness increase	0.00814 (0.00782)	0.00814 (0.00782)	-0.00523 (0.00955)	-0.00523 (0.00955)	-0.00780 (0.00779)	-0.00780 (0.00779)
residual	-0.0460 (0.843)	-0.0460 (0.843)	0.242 (0.939)	0.242 (0.939)	-0.141 (0.156)	-0.141 (0.156)
Observations	5,147	5,147	5,147	5,147	5,070	5,070
Number of HH	1,760	1,760	1,760	1,760	1,761	1,761
Households F.E	×	×	×	×	×	×
Village*year F.E		×		×		×

Note: The estimation sample includes all households with adult labor and cultivated lands while the simulation sample is constrained to households with adult labor and cultivated lands. Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. We also control by the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the gender of the household's head, the access to the plot GPS information, the age of the household head, survey month dummies and year dummies.

TABLE A1.9 : ESTIMATION OF THE SUBSTITUTION  
PARAMETER  $\rho$

	Imperfect		Translog	
	(1)	(2)	(3)	(4)
OLS	1.069 (1.680)	-0.787 (0.820)	-0.0198 (0.0409)	-0.0582 (0.0663)
IV	12.03 (238.6)	-0.83 (0.796)	-0.0189 (0.0412)	-0.0571 (0.0661)
Households FE	YES	YES	YES	YES
Village*year FE	NO	YES	NO	YES
Observations	4,994	4,994	4,994	4,994

Note: Sample: households with adult labor and with cultivated lands. Standard errors are clustered and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are pesticide and inorganic fertilizer spending, the use of organic fertilizer, the precipitation at date t, the number of idiosyncratic and covariate shocks, the erosion and the irrigation status, the productive assets, the gender of the HH head, the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the gender of the household's head, the access to the plot GPS information, survey month dummies and year dummies.

TABLE A1.10 : OLS ESTIMATIONS OF THE PRODUCTION FUNCTION FOR THE UNRESTRICTED SAMPLE (WITH HOUSEHOLDS OBSERVED ONLY AT ONE WAVE).

	(1)	(2)	(3)
$\log L_a$	0.711*** (0.0634)	0.712*** (0.0636)	1.757*** (0.301)
$\log L_c$	0.688*** (0.110)	0.718*** (0.183)	
$\frac{L_c}{L_a}$		-0.0508 (0.228)	
$\frac{L_c}{L_a} * (1 - \log \frac{L_c}{L_a})$			0.422*** (0.104)
$\log L_a * \log L_a$			-0.116*** (0.0301)
$\log L_c * \log L_c$			0.00425 (0.0142)
$\log L_a * \log L_c$			-0.0698*** (0.0208)
$\log(\text{cultivated area})$	0.998*** (0.0611)	0.998*** (0.0611)	1.004*** (0.0607)
org. Fertilizer	0.371*** (0.134)	0.371*** (0.134)	0.383*** (0.135)
$\log(\text{inorg})$	0.142 (0.109)	0.142 (0.109)	0.124 (0.110)
$\log(\text{pesticide})$	0.0872*** (0.0104)	0.0872*** (0.0104)	0.0856*** (0.0103)
Land eroded	0.0154 (0.0135)	0.0154 (0.0135)	0.0188 (0.0134)
Land irrigated	0.0945 (0.0592)	0.0944 (0.0592)	0.101* (0.0588)
Productive assets	-0.124 (0.128)	-0.124 (0.128)	-0.143 (0.128)
NBR idio. shocks	-0.176*** (0.0648)	-0.175*** (0.0649)	-0.177*** (0.0651)
NBR covariate shocks	0.103 (0.101)	0.102 (0.102)	0.105 (0.100)
Rainfall	0.822*** (0.260)	0.822*** (0.261)	0.808*** (0.258)
Mean temperature	0.0110 (0.00977)	0.0109 (0.00977)	0.0158* (0.00947)
Greenness increase	-0.0125*** (0.00263)	-0.0125*** (0.00263)	-0.0119*** (0.00261)
R-squared	0.383	0.383	0.385
Households F.E			
Village*year F.E			
Observations	7,050	7,050	7,050

Note: Sample: households with adult labor and with cultivated lands. Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. We also control by the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the gender of the household's head, the access to the plot GPS information, the age of the household head, survey month dummies and year dummies.

TABLE A1.11 : EFFECT OF THE NUMBER OF CHILDREN ON  
CHILD LABOR BY GENDER (FIRST STAGE)

IV variables	Girls		Boys	
	(1)	(2)	(3)	(4)
$\frac{N_g}{N_a}$	0.262*** (0.0343)	0.257*** (0.0395)		
$\frac{N_b}{N_a}$			0.237*** (0.0219)	0.234*** (0.0238)
R-squared	0.124	0.292	0.118	0.289
F-test	59.40	43.23	91.95	93.83
HH F.E	×	×	×	×
village-year F.E		×		×
Observations	1460	1482	1460	1482

Note: The estimation sample includes all households with adult labor while the simulation sample is constrained to households with adult and child labor. Standard errors are clustered and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are pesticide and inorganic fertilizer spending, the use of organic fertilizer, the precipitation at date t, the number of idiosyncratic and covariate shocks, the erosion and the irrigation status, the productive assets, the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the gender of the household's head, the access to the plot GPS information, survey month dummies and year dummies.

TABLE A1.12 : EFFECT OF THE NUMBER OF CHILDREN BY AGE GROUPS ON CHILD LABOR (FIRST STAGE)

Age-group	perfect and imperfect (1)	translog (2)
10	0.150 (0.186)	0.0954 (0.550)
11	0.439* (0.248)	0.305 (0.463)
12	0.187 (0.154)	-0.396 (0.602)
13	0.304** (0.133)	-0.369 (0.390)
14	0.504*** (0.152)	0.897* (0.461)
15	0.486*** (0.136)	0.807* (0.455)
R-squared	0.063	0.032
F-test	35.76	2.307
HH F.E	×	×
Observations	4,811	4,921

Note: The estimation sample includes all households with adult labor while the simulation sample is constrained to households with adult and child labor. Standard errors are clustered and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are pesticide and inorganic fertilizer spending, the use of organic fertilizer, the precipitation at date  $t$ , the number of idiosyncratic and covariate shocks, the erosion and the irrigation status, the productive assets, the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the gender of the household's head, the access to the plot GPS information, survey month dummies and year dummies.

TABLE A1.13 : MEAN OF SEMI-ELASTICITIES: PERFECT SPECIFICATION WITH AGE HETEROGENEITY (EQUATION (1.19)).

Age	(1)	(2)	(3)	(4)
	OLS		IV: $\frac{N_{5-15}}{N_a}$	
5	0.000665 (0.00348)	0.00209 (0.00451)	-0.000423 (0.00714)	0.00298 (0.00906)
6	0.000937 (0.00305)	0.00216 (0.00392)	-0.000139 (0.00679)	0.00304 (0.00870)
7	0.00121 (0.00261)	0.00222 (0.00333)	0.000146 (0.00645)	0.00309 (0.00838)
8	0.00148 (0.00219)	0.00229 (0.00275)	0.000431 (0.00615)	0.00315 (0.00809)
9	0.00175 (0.00178)	0.00235 (0.00218)	0.000715 (0.00588)	0.00320 (0.00784)
10	0.00202 (0.00139)	0.00242 (0.00165)	0.001000 (0.00564)	0.00326 (0.00764)
11	0.00230** (0.00106)	0.00248** (0.00119)	0.00128 (0.00545)	0.00332 (0.00748)
12	0.00257*** (0.000854)	0.00255*** (0.000932)	0.00157 (0.00530)	0.00337 (0.00737)
13	0.00284*** (0.000857)	0.00262** (0.00103)	0.00185 (0.00520)	0.00343 (0.00732)
14	0.00311*** (0.00107)	0.00268* (0.00141)	0.00214 (0.00515)	0.00348 (0.00732)
15	0.00339** (0.00140)	0.00275 (0.00192)	0.00242 (0.00516)	0.00354 (0.00738)
Adults	0.00207 (0.00179)	0.00254 (0.00250)	0.00330 (0.00644)	0.00157 (0.00905)
Residual			.17734 (1.01085)	.03448 (1.35406)
F-test			20.12	11.61
Households F.E	×	×	×	×
Village*year F.E		×		×
Simulation sample	1,547	1,571	1,547	1,571

Note: The estimation sample includes all households with adult labor and cultivated lands while the simulation sample is constrained to households with adult, child labor and cultivated lands. Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are pesticide and inorganic fertilizer spending, the use of organic fertilizer, the precipitation at date t, the number of idiosyncratic and covariate shocks, the erosion and the irrigation status, the productive assets, the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the age of the household's head, the access to the plot GPS information, survey month dummies and year dummies.



TABLE A1.14 : MEAN OF THE SEMI-ELASTICITY OF CHILDREN AGED 11 TO 16 AND ADULTS AGED 17 TO 65

	(1)	(2)	(3)	(4)	(5)	(6)
	Perfect		Imperfect		Translog	
	OLS	IV	OLS	IV	OLS	IV
Adult	0.00483*** (0.000817)	0.00412 (0.00539)	0.00337*** (0.000877)	0.00604 (0.00584)	0.00445*** (0.000931)	0.00389*** (0.000915)
Child	0.00121*** (0.000444)	0.00177 (0.00430)	0.00391*** (0.000976)	0.00231 (0.00422)	0.00493*** (0.00129)	0.00528 (0.00680)
$\hat{\eta}$		-0.00534 (0.760)		0.334 (0.786)		-0.00145 (0.132)
Test $\alpha_{aa} = \alpha_{cc} = \alpha_{ac} = 0$						4.336
Test $\alpha(1 + \rho)\gamma_c = 0$	0.273					
Households F.E	YES	YES	YES	YES	YES	YES

Note: The estimation sample includes all households with adult labor while the simulation sample is constrained to households with adult and child labor (16). Standard errors are clustered and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are pesticide and inorganic fertilizer spending, the use of organic fertilizer, the precipitation at date t, the number of idiosyncratic and covariate shocks, the erosion and the irrigation status, the agricultural asset index, the percentage of cultivated lands allocated to maize, paddy, sorghum, cassava, sweet potatoes, beans, cowpeas, sunflower, groundnuts and cotton, the gender of the household's head, the access to the plot GPS information, survey month dummies and year dummies.

TABLE A1.15 : MEAN OF THE SEMI-ELASTICITY OF LABOR. THE INSTRUMENTS ARE THE NUMBER OF CHILDREN BY AGE GROUPS  $N_{10} \dots N_{15}$ .

Variables	Perfect		Imperfect		Translog	
	OLS	IV	OLS	IV	OLS	IV
Without village-year F.E						
Adult	0.003** (0.001)	0.004 (0.006)	0.007*** (0.002)	0.008 (0.006)	0.005*** (0.001)	0.005*** (0.001)
Child	0.003*** (0.001)	0.002 (0.005)	0.004*** (0.001)	0.004 (0.005)	0.004*** (0.001)	0.014 (0.022)
$\hat{\eta}$		0.141 (0.952)		0.159 (0.961)		-0.179 (0.401)
$\alpha_{aa} = \alpha_{cc} = \alpha_{ac} = 0$						4.926
$\alpha(1 + \rho)\gamma_c = 0$	5.46					
Prob>F			0.0198	0.0000	0.00237	0.0024
Climatic factors	×	×	×	×	×	×
Households F.E	×	×	×	×	×	×
Simulation sample	1460	1460	1460	1460	1460	1460

Note: The estimation sample includes all households with adult labor and cultivated lands while the simulation sample is constrained to households with adult, child labor and cultivated lands. Standard errors reported in parentheses are clustered at the ward level and are bootstrapped for IV estimates. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Climatic factors are the precipitation at date t, the mean temperature and the greenness increase in days. For the perfect and the imperfect function, the IV is  $\frac{N_c}{N_a}$ . For the translog function, the IV is  $N_c$ .

TABLE A1.16 : COMPARISON OF HOUSEHOLDS WITH AND WITHOUT LIVESTOCK

Variables	(1) With livestock	(2) Without livestock
Age HH head	50,40	49,10
Education HH head	1,84	1,87
Production (TSH)	456627,80	271047,20
Cultivated area (acres)	6,78	4,11
Number of children aged 10 to 15	1,20	0,81
Number of adults	3,22	2,56
Days of child labor (10-15)	25,25	12,04
Days of adult labor (16-65)	209,07	144,28

Sources: LSMS-ISA panel data (2008,2010,2012)

### 1.7.3 Semi-elasticities with function $\log^M$

#### Dependant variable

Equations (1.4), (1.5) and (1.6) describe the functional forms used in this paper (respectively for the specifications perfect, imperfect and translog). These equations follow the form  $\log Y = f(L_c, L_a, X)$ . We estimate these equations with the function  $\log^M$ , as we need to take into account the observations that report an absence of output:  $\log^M Y = f(L_c, L_a, X)$ .

This decision affects marginally the interpretation of the semi-elasticities. Indeed, the monetary unit of Tanzania is small ( $1\$ \approx 1000TSh$ ), and accordingly  $\log^M Y \approx \log Y$  as soon as  $Y > 0$ . Besides, let us write the fundamental model  $Y = u \exp f(L_c, L_a, X)$  where  $u \geq 0$  is a multiplicative error term independant from  $L_c, L_a, X$ .

$$\log^M(Y) \approx f(L_c, L_a, X) + \log u \text{ when } u > 0 \quad (1.20)$$

$$= 0 \text{ when } u = 0 \quad (1.21)$$

$$E(\log^M(Y)|L_c, L_a, X) \approx P(u > 0)E(\log Y|u > 0, L_c, L_a, X) \quad (1.22)$$

The semi-elasticities (1.7) and (1.9) are based on the difference in  $\log Y$  between two situations with and without labor. The corresponding differences in  $\log^M(Y)$  are differences in  $\log Y$  multiplied by the chances to have a production. In other words, we measure an additional production per day in percentage multiplied by the chances to have a production.

#### Explanatory variables

The semi-elasticities (1.7) and (1.9) depend on the expected production for two different levels of  $L_c, L_a$ . When  $\log$  enters the estimated equations, we use the function  $\log^M$  in the estimation and in the simulations following the estimation, hence the simulated function is the same as the estimated function.

## Chapter 2

# The Impact of the Universal Primary Education Program on Labor Market Outcomes: Evidence from Tanzania



## 2.1 Introduction

Education is a cornerstone for economic growth and plays a crucial role in labor markets. As a consequence, governments and non-governmental organizations have put education at the top of their agenda. More specifically, several governments of developing countries have implemented policies to universalize primary education. An extensive body of literature underlines the positive correlation between education and earnings but does not inform about the causality of this relationship. [Card \(2001\)](#) reviews papers that aim to identify the causal impact of education on earnings. To disentangle the ability effect from the education effect on earnings, these selected papers either instrument education based on characteristics of the schooling system, or use family background as a control or an instrument. Among the eleven papers included in Card's survey, only two of them focus on developing countries, the paper of [Duflo \(2001\)](#) where education is instrumented by a school construction program in Indonesia, and the study by [Maluccio \(1998\)](#) where education is instrumented by the distance to school in rural Philippines. Both authors restrict their analysis to individuals who earn a wage, which raises the question of the representativeness of the samples. Indeed, wage-earning individuals are likely to be self-selected and to have specific characteristics. [Maluccio \(1998\)](#) does not deal with this sample selection issue while [Duflo \(2001\)](#) adopts an imputation technique to compute a wage for individuals from the self-employment sector. While this method is suitable for countries with a developed formal sector, it is less adapted to countries that are mainly agriculture-based. Indeed, in agriculture, few individuals are wage-earners, and household level agricultural production is more likely to be more representative of household wealth.

To address this important issue, another strand of the literature estimates the returns to education among agricultural households. [Griliches \(1964\)](#) was the first researcher to measure the impact of the education of the household head on agriculture with a production function. [Lockheed \*et al.\* \(1980\)](#) review papers estimating the impact of education on agricultural production and find very mixed results depending on the country

and the specification of education. However, these papers do not consider the endogeneity of education of the head.

The first contribution of this paper is to assess the efficiency of a massive primary education program. To do so, I evaluate whether the Universal Primary Education program implemented in Tanzania from 1974 to 1978 ensured the expansion of the education system and contributed to reducing inequality in access to education.

The second contribution is to estimate the returns to education in developing countries for the entire population. Since developing countries are often characterized by the large size of both the non-agricultural self-employed sector and the agricultural sector, I use consumption aggregates that are available for all sample households. To account for the potential endogeneity of education, I instrument education of the household head by the intensity of the UPE program, which constitutes a natural experiment. In 1974, educational levels were low at the national level, with wide variation between regions. The introduction of the UPE program led to substantial results: 3.3 million children aged 7 to 13 were enrolled in 1980, compared to 1.2 million in 1974 (Bonini, 2003). To reduce disparities in access to education, the Tanzanian socialist government gave priority to deprived areas, which led the latter to experience higher schooling expansion. Therefore, the exposure to the UPE program varied according to the age of the individual at the time of the reform and to the educational level by regions before the introduction of the program. Thus, the UPE program gives rise to an exogenous variation in education that I exploit to instrument education and to determine the effect of education on consumption. In order to capture variability in the returns to education, I also distinguish between the returns to education for subgroups: the agricultural sector, the non-farm self-employed sector, and wage-work activities.

The third contribution of this paper is to address the effect of education on the labor market organization, more precisely, on the probability of working in each sector of activity. In order to do so, I adopt the same identification strategy and I instrument education by exploiting the nature of the UPE program.

The main findings of this paper suggest that the UPE program reduced inequalities of access to education and that one additional year of education augmented consumption aggregates from 7.4 percent to 10.5 percent. Since the program was directed toward agriculture by providing a specific curriculum with agricultural classes, returns to education were higher in the agricultural sector than in the wage-earner sector. Similarly, education increased the probability of working in agriculture and decreased the probability of working in the non-agricultural self-employment sector.

The remainder of the paper is organized as follows: section 3.3 provides a broad picture of the evolution of education in Tanzania and describes the data and the main variables of the analysis. Section 2.3.1 introduces the identification strategy; section 2.3.2 presents the effect of the UPE program on education; section 2.4.1 and section 2.4.3 respectively, focus on the effect of education on consumption and on labor market participation. Finally, section 2.5 concludes.

## **2.2 The program**

### **2.2.1 Historical background and the UPE program.**

When colonization ended in 1961, access to education in Tanzania was very unequal between regions (Court and Kinyanjui, 1980). These spatial disparities were based on ecological endowments and were exacerbated by colonial activities and transport networks.<sup>1</sup> At this time, the purpose of primary education was to prepare for secondary education and was to encourage a small number of rural students to find white-collar jobs in urban areas (Kinunda, 1975). The arrival in power of the Prime Minister Nyerere in 1964 marked a radical political and economic change. In 1967, the policy of Education for Self-Reliance (ESR) was approved. Education became the mainstay of the Tanzanian socialist economy that would ensure economic growth and primary education was put at the service of the development of rural areas, where 96 percent of the population lived.

---

<sup>1</sup>The most privileged zones were the Arusha-Kilimanjaro-Tanga and the Mwanza-Shinyanga corridors, and the Coast Morogoro-Kigoma (Maro and Mlay, 1979).



This policy should have led to radical changes but, in practice, was slowly enforced. It was only in 1974 that the government committed itself to reach at a forced-march Universal Primary Education (UPE) by 1978. The aim of this program was threefold: i) to improve the equity of access to education, ii) to teach agricultural skills that would be relevant in a rural society, and iii) to offer a political and civic education (Nyerere, 1967).

To achieve the UPE goal, the government made a series of changes. First, it implemented a villagization program to increase access to schools and other social services. From 1968, villagization consisted of constructing community villages commonly called *ujamaa*, but in 1974, households living in remote areas were forced to move (see Table 2.2 ). Most of the time, the distance to their prior dwelling was less than five kilometers. From 1974 to 1977, more than 10 million people were moved and 2,650 *ujamaa* were built (Martin, 1988). Prime Minister Nyerere (1987) considered that gathering the rural population was necessary to develop education, to reduce inequalities and to improve agricultural production. The Tanzanian government invested massively in primary education and concentrated its efforts on deprived areas. Local resources were mobilized for classrooms and a large number of new schools were built. In 1978, expenditure on primary education was three times the amount dedicated to secondary education (Bonini, 2003). Thus, the UPE program combined with villagization greatly reduced distances to schools. Simultaneously, teachers' recruitment and teacher training were restructured. To deal with the growing number of pupils, the government trained 10,000 teachers. Despite this, there was still a shortage of primary school teachers, which may have affected the quality of education, especially in the beginning of the UPE plan (Sabates *et al.*, 2011). The government also made additional adjustments to improve schools' attractiveness. Tuition fees were eliminated, primary education became mandatory, and Swahili, most pupils' mother tongue, was designated as the language of instruction.

The aim of primary school was also to become a self-reliant economy through agriculture: «*kilimo cha kufa na kupona*», *Agriculture for Life and Death*. To fulfill this goal, the examination in the middle of the primary cycle was removed, the starting

TABLE 2.1 : VILLAGES IN TANZANIA

Year	Number of villages	Number of residents
1968	180	58 000
1969	650	300 000
1970	1200	50 000
1971	4484	1 595 240
1972	5556	1 980 862
1973	5631	2 028 164
1974	5008	2 560 474
1975	6944	9 140 229
1976	7658	13 067 220
1978	7768	13 847 000
1979	8200	13 905 000

Source: Shao (1982)

age was postponed from 5 to 7 years old and agriculture classes were introduced in the curriculum. As a result, pupils leaving the primary schools would be old enough and would have acquired the abilities to work in the fields. To encourage people to start working after primary school, access to the secondary cycle was drastically limited by regional quotas (Martin, 1988).<sup>2</sup>

The results of this UPE plan were considerable: From 1974 to 1978, enrolled children aged 7 to 13 rose from 43.1 to 90.4 percent, access to primary education was improved due to the construction of new schools, and disparities among regions were reduced (Bonini, 2003).

## 2.2.2 Data

### 2.2.2.1 Data sets

This study uses three data sources: a census data set, a household panel survey, and administrative data. First, the census data used are a 10 percent IPUMS sample from the 2002 Population and Housing Census in Tanzania. This census was carried out by the National Bureau of Statistics (NBS) and contains basic information on dwelling characteristics, individual demographics and socio-economics for 500, 519 households.

<sup>2</sup>Despite this policy, no significant drop of the secondary enrollment rate is observed.

Secondly, the household panel data are the LSMS-ISA (LSMS-Integrated Surveys on Agriculture) data collected by the World Bank in 2008-2009, 2010-2011 and 2012-2013.<sup>3</sup> The LSMS-ISA data include 3265 households in 2008, 3924 households in 2010 and 5015 households in 2012. The number of households is increasing over the three waves due to the high number of split-off households and to the low attrition rate that does not exceed 5 % over the three rounds. This dataset gives detailed information on labor activities, household consumption, agriculture inputs and outputs and on other individual characteristics. Despite a district reorganization between the dates of the two datasets, both data cover the 26 Tanzanian regions and are representative at the regional and national levels.

Finally, I use administrative data collected by the Ministry of Economic Affairs and Development Planning and recorded in *Jensen et al. (1968)*. These data gather information on the distribution of primary schools and on GDP by regions and districts in 1967 for mainland Tanzania. GDP records are divided in sub-activities such as crops, livestock, mining, manufacturing, construction, public utilities, transport, rent, and other services. These data are unique because they constitute, to the best of my knowledge, the only source of information on primary school provision in Tanzania at this time.<sup>4</sup>

### **2.2.2.2 Measuring household wealth**

Usually, living standards are measured either by income or by consumption. In developing countries where agriculture is widespread, incomes are very sensitive to current shocks and may not be representative of household well-being (*Meyer and Sullivan, 2003*) while consumption can be smooth through formal or informal mechanisms. In this respect, consumption has the advantage of being more representative of long-run well being. The second interest of using consumption stems from the labor organization in developing countries. Individuals mostly work in household activities such as agriculture or self-employed activities where individual incomes are not well identified (*Jolliffe, 1998*).

---

<sup>3</sup>From October 2008 to December 2009 for the first wave, from October 2010 to December 2011 for the second wave, and from October 2013 to December 2013 for the third wave.

<sup>4</sup>The National Bureau of Statistics gives access to the number of schools by region only from 2002.

Additionally, income is not similarly measured between sectors,<sup>5</sup> which calls into question the reliability of comparison between sectors. Last, but not least, consumption is available for all households, which allows one to avoid selection and imputation issues. Thus, these features advocate the use of consumption rather than incomes data in developing countries.

Deaton and Zaidi (2002) propose guidelines to construct a very detailed consumption variable from household survey data. They consist of defining a weighted per capita consumption variable composed of four components, food items, non-food items (electricity, health expenditure, etc), housing consumption (derived from imputing a rent for each household) and consumption from consumer durables. To adjust household consumption for variation in household composition, the consumption variable is divided by an equivalence scale.<sup>6</sup> The Living standard Measurement Study (LSMS) data are particularly well suited for constructing the consumption index since they collect exhaustive information on consumption expenditures. However, the serious limitation of the Deaton and Zaidi (2002) method is such accurate data are costly to collect and are often not included in large datasets. As in most census data, the 2002 Tanzanian census excludes income and expenditures, but records a list of dwelling characteristics and durable goods. A large number of authors ((McKenzie, 2005), (Vyas and Kumaranayake, 2006)) adopt a factor analysis which consists in constructing an asset index based on access to utilities and housing characteristics. This index proves to be useful for measuring inequalities between households and are good proxies for long-term wealth. The additional advantage of these indexes is that they limit measurement errors (Sahn and Stifel (2003), (Filmer and Pritchett, 2001)). Thus, I construct a consumption variable from the Tanzanian LSMS data<sup>7</sup> and a consumption proxy from the 2002 census.

To take advantage of the large sample size of the census data and obtain a monetary value of consumption which eases comparison with the literature (Duflo (2001) and

---

<sup>5</sup>Self-employment income is rarely a wage and agricultural income is measured through production.

<sup>6</sup>The equivalence scale is made from the household's size: every adult represents one unit and each child represents 0.3 units.

<sup>7</sup>The data report detailed spending, except for consumer durables.

Maluccio (1998)), I construct a consumption proxy from the census data by adopting the method developed by (Elbers *et al.*, 2003) and (Tarozzi and Deaton, 2009)) from census data and household survey matching. By following a two step-procedure, I predict household consumption from a set of predictors  $X$  that are common to both data sources.<sup>8</sup> The idea behind this method is first to estimate the joint distribution of  $\log C$  and  $X$  from the household survey:

$$\text{Log}C_i = bX_i + \nu_i \quad (2.1)$$

Where  $\nu_i$  is the error term of household  $i$ . Then, I use the estimated distribution to predict  $\widehat{\log C}$  in the census data. This method is valid only if several assumptions hold. The predictors  $X$  should be similarly measured in both data sets. Questions regarding the predictors have the same wording for the two questionnaires, but as Tarozzi and Deaton (2009) suggest, differences may still persist due to differences in questionnaire length or interviewer training.<sup>9</sup> Table A2.3 gathers coefficients of equation 2.1 in the LSMS data. As predicted, all dwelling characteristics have a positive and significant impact on consumption. The R-squared coefficient is high (0.53), meaning that the predictors have good explanatory power.

Graph A2.3 plots the relationship between the expected utility  $\widehat{\log C}$  and  $\log C$  in the LSMS data.<sup>10</sup> I find a clear positive linear relationship between these two variables. However,  $\widehat{\log C}$  may not capture all the variation in consumption, especially at the ends of the distribution where large consumption gaps can exist between households who have all the basic dwelling characteristics. As a result, the dispersion for extreme values of  $\widehat{\log C}$  is larger, but this effect seems negligible. To account for the artificially low variance of  $\widehat{\log C}$  compared to  $\log C$ , I adopt the method proposed by Barham and Boucher (1998) and Gubert *et al.* (2010). This recommends adding to  $\widehat{\log C}$ , an error term drawn from a normal distribution with the same variance that  $\hat{\nu}_i$  observed in the survey data. To make

---

<sup>8</sup>The number of rooms in the dwelling, whether the household has drinking water, electricity, a phone, a flush toilet, a high quality roof, high quality walls, etc.

<sup>9</sup>To avoid anachronism issues, I do not include in the list of predictors “having a phone” that may have a different meaning across time and across the data.

<sup>10</sup>For each value of  $\widehat{\log C}$ , I compute the average value of  $\log C$  depicted by a dot.

sure that the results are independent from the random draw, this procedure is replicated a large number of time. Thanks to this method, the standard errors can be normally interpreted. However, this assumes that  $\nu_i$  is exogenous, which may not be the case in practise.<sup>11</sup>

### 2.2.3 Measuring intensity of the UPE program

The UPE program was applied during a limited time frame and targeted regions with poor access to education. Hence, exposure to the program is captured by two types of variation, across locations and over time. The spatial variations can be captured by referring to several administrative geographical units. The smaller the unit, the higher will be the accuracy of the instrument. In this respect, the census district is best suited for the analysis. However, census data report only the district of residence. Since individuals could have moved from one place to another during the UPE program, the place of residence might be endogenous. To avoid this endogeneity issue, I use instead the region of birth, which was determined prior to the program.<sup>12</sup>

Since the UPE program seeks to equalize access to education across all locations, the intensity of the program should be a decreasing function of a locality's initial education level. To capture the heterogeneity in the treatment, I first need to obtain information on each locality's education level before the introduction of the UPE program. The UPE program was fully implemented starting in 1974, but pilot programs started in 1968, just after the ESR policy. As a consequence, some regions benefited from financial support and from the villagization procedure between 1968 and 1974. Thus, the age-cohorts that were in school from 1968 to 1974 were likely to be partially treated. Consequently, the education level of region  $j$  in 1967,  $S_{j,1967}$  is the best indicator of the initial education level.<sup>13</sup>

---

<sup>11</sup>In section 2.4.1, consequences of relaxing this hypothesis will be further studied.

<sup>12</sup>Results that use the district of residence as geographical unit are similar and are available upon request.

<sup>13</sup>Given that primary school in Tanzania ends at 13 years old,  $S_{j,1967}$  is computed from the education level of individuals born in 1954.

Figure A2.4a maps this variable, and it is clear that, indeed, education was unequal across regions: Zanzibar West and Kilimanjaro had already reached the maximum years of primary education while the average education in others regions was low. In contrast, education levels were higher and more homogenous by the end of the UPE program in 1978 (see figure A2.5b). Although  $S_{j,1967}$  gives a good picture of access to education by region in 1967, it also reflects the demand for education, and so it might be endogenous. To get around this problem, I also use data on school infrastructure and I construct  $N_{j,1967}$ , the number of primary schools per square kilometer, by region. Figure A2.4c depicts the distribution of  $N_{j,1967}$ . Two striking features are evident. Firstly, the schooling supply was very unequal between regions in 1967 and secondly, the supply of schools seems correlated with the distribution of the education level at this time.

## 2.2.4 Sample and Descriptive statistics

Knowing that the UPE program partially started in 1968 and was fully implemented in 1974, I focus my analysis on two main groups, the pre-treatment group  $T_0$  and the treatment group  $T_{tot}$ . I define  $T_0$  to be household heads not affected by the UPE reform. It is composed of individuals born between 1945 and 1954 who were older than 13 before the introduction of the program in 1968. The UPE program affected several age-cohorts but with varying intensity. I define  $T_{tot}$  to include all treated children and I distinguish three treated sub-groups according to the age of the household head (see Table 2.2). Both the T1 and T2 cohorts were affected by the reform before it ended in 1978. Household heads from T1 (born from 1961 to 1966) were likely to be treated at the beginning of the reform while household heads from T2 (born from 1967 to 1971) were likely to be treated at the end of the reform. After 1978, school attendance started flattening (King, 1984) but children still benefited from the UPE program infrastructure. This last group, denoted by  $T_3$  (born between 1972 to 1978) can be used to test whether the effect of the UPE program was persistent over time. Alternatively, the control group can be considered to be  $T_{pt}$  (born from 1945 to 1960), that consists of household heads that were likely to be

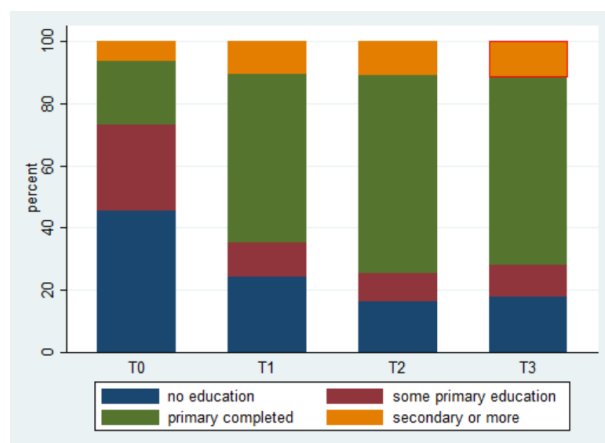
partially treated by the UPE program.

TABLE 2.2 : AGE COHORTS

Age cohorts	Year of birth	Age in 1974	Potential education level during the UPE plan	Obs. IPUMS	Obs. LSMS
$T_b$	1935-1945	29-39	over postsecondary	77,115	1,083
$T_0$	1945-1954	20- 29	postsecondary and over	111,818	1,706
$T_{pt}$	1945-1960	14-19	secondary and postsec.	83,937	1324
$T_1$	1961-1966	8 -13	primary-secondary	113,063	1,555
$T_2$	1967-1971	3-7	no education- primary	103,406	1,408
$T_3$	1972-1978	not born-2	no education	172,232	2,156
$T_{tot}$	1961-1978	not born-13	no education-secondary	388,701	5,119

Figure 2.1 shows the education distribution among these age-cohorts. One notices that the percentage of individuals with no education drastically decreased from  $T_0$  to  $T_1$  (falling from 45 to 24 percent): not only were more people enrolled in primary school, but more people completed primary education. From  $T_1$  to  $T_2$ , the percentage of the population with primary completion kept growing, but the percentage of people with no education remained stable. Although the substantial increase in education between the control and treated groups, about 20% of the population had no education.

FIGURE 2.1: EVOLUTION OF THE EDUCATION ATTAINMENT BY AGE-COHORT.



Source: 2002 census (IPUMS data).

Table A2.1 and A3.1 present descriptive statistics for the LSMS and the census data, respectively. On average, individuals from  $T_{tot}$  appear to be more educated than



individuals from  $T_0$ , and this gap is larger for women. Besides, the head man of the household seems slightly more educated than his partner. Turning to consumption, I notice that  $\widehat{\log C}$  is very close to the  $\log C$  in the LSMS-data (see TableA3.1) and that  $\widehat{\log C}$  is slightly larger in the census data than in the LSMS data. When I compare consumption by age-group, I find no statistical difference between household heads from  $T_0$  and household heads from  $T_{tot}$ . Moving to the sector of activity, I observe that they are not similarly defined in both data sets. The census data distinguish between four categories: i) unemployed or domestic unpaid workers, ii) wage-workers either in the public or the private sector, iii) self-employed workers in non-agricultural activities<sup>14</sup> and iv) self-employed workers in agriculture. The LSMS survey defines the same categories but specifies whether individuals hold several jobs. Agriculture is by far the dominant sector. From  $T_0$  to  $T_{tot}$ , the distribution between sectors was relatively stable, despite the shift from agriculture to self-employed non-agricultural activities.

## 2.3 Empirical strategy

To analyze the effect of education on labor market outcomes, I rely on the following basic equation:

$$Y_i = \theta_1 S_i + \theta X'_i + \varepsilon_i \quad (2.2)$$

where  $Y_i$ , the outcome of interest, denotes the consumption per capita or the occupational status.  $S_i$  is the years of schooling of the head in household  $i$ ,  $X'_i$  is a set of controls and  $\varepsilon_i$  is the error term.

As the literature underlines (Card, 2001), education depends on individual choices that may be endogenous. Thus, unobserved omitted variables may influence both education and the outcome interest and simple OLS estimations would lead to inconsistent estimates of  $\theta_1$ . To address this endogeneity issue, I adopt an instrumental variable approach.

---

<sup>14</sup>This includes non-agricultural business with employees and without employee.

### 2.3.1 Identification strategy

The instrumental variable approach that I use is based on the UPE program. This assumes that being exposed to the program increases the probability of being enrolled in school but is orthogonal to unobserved household characteristics that determine labor market outcomes. Because the UPE program is an exogenous source of education variation, it should be a reasonable instrument.

To capture exposure to the UPE program, I adopt a difference-in-differences strategy based on variations in time and in space. It consists of comparing pre-treatment cohorts ( $T = 0$ ) with treated cohorts ( $T = 1$ ), for whom the intensity of the treatment varied across regions.

Since the program aimed at improving equity of access to education, the intensity should decrease with the education level before the program,  $S_{67}$ . As mentioned above, Zanzibar West was the most educated region in 1967. It is also the region that experienced the lowest increase in years of schooling between 1967 and the end of the program in 1978. Thus, I define the intensity of the program as:

$$I_{j,67} = (S_{Zanzibar\ West,67} - S_{j,67})$$

By construction, Zanzibar is untreated and the intensity of the treatment is a decreasing function of  $S_{j,67}$ . Thus, I instrument education by  $T * I_{j,67}$ , which captures the UPE program's exposure. This variable is a valid instrument (IV) if two conditions are satisfied: i) the IV is correlated with education, and ii) the IV explains the outcome of interest only through education. In such cases, IV estimates correspond to the Local Average Treatment Effect (LATE). Otherwise, IV estimates give inconsistent and biased results. Since the interpretation of IV estimates relies on the quality of the instrument, I now discuss whether the interaction term  $T * I_{j,67}$  is a valid instrument.

The IV variable is a relevant candidate if it is highly correlated with the endogenous variable. To check whether  $S_{j,67}$  explains the education expansion by regions, I plot in

figure A2.6 the education increase from the pre-treatment group  $T_0$  to the program period  $T_1$  according to the regional education attainment in  $T_0$ . Each dot depicts a region of birth. I show that, indeed, there is a clear negative relationship between the initial education level and the education increase: the UPE program was more intense in regions with low schooling enrollment at  $T_0$ . I find the same result in figure A2.7, where I consider the supply of school instead of the education level. Likewise, the same conclusion can be drawn at the district level in figure A2.8: the lower the education in  $T_0$ , the higher is the education increase from  $T_0$  to  $T_1$ .<sup>15</sup> However, this relationship is not necessarily causal. The main concern is that the education expansion between  $T_0$  and  $T_1$  is not exclusively due to the UPE program but to other factors correlated with the instrument and the outcome of interest. Despite the fact that the exclusion restriction could not be tested, I try to identify all potential sources that could discredit this condition, and I provide evidence that the instrument is exogenous.

First, I check whether the education expansion is due to the introduction of the UPE program and not to a convergence phenomenon. In case of convergence, less educated regions could have had a higher education increase in order to catch up with the more educated regions. If this were to be true, this phenomenon would be observed before and after the introduction of the UPE program. Subsection 2.3.2 addresses this question and confirms that during the pre-treatment program, the education progression was not statistically different between educated and non-educated regions, whatever their initial education level. On the contrary, there is a trend reversal during the UPE program period and the education expansion was statistically higher for deprived regions.

The exclusion restriction can still be invalidated if other region's characteristics generate the same trend reversal or are correlated to the outcomes of interest. In order to insure the exogeneity of the instrument, I add a set of controls. Among these control variables, I add the number of children aged 7 to 13 to account for possibility that the education expansion may depend on the size of the cohort. Furthermore, the level of wealth may have non negligible impacts on the development of the schooling supply:

---

<sup>15</sup>Each dot represents a district of residence because the districts of birth are not available.

wealthy regions can have higher needs in skilled labor and invest more in education. In addition, when a region becomes specialized in a given sector of activity, this region becomes more vulnerable to all the specific shocks of that sector. For instance, regions with a developed agricultural sector are more likely to be sensitive to commodity price shocks while mining regions are probably more perturbed by variations in energy prices. These variations in the labor market organization can also bias the results if they affect the demand for education and the returns to education. Hence, if regions are not homogeneously affected by sectoral economic shocks, the exclusion restriction is not valid and IV estimations are inconsistent. One way to ensure the validity of the instrument is to control for heterogeneity in order to capture variations in shocks between regions. In this respect, I add regional GDP by sector of activity interacted with a time trend. Among these sectors, I distinguish between the following economic activities: crops, livestock, mining, manufacturing, construction, and activities from the tertiary sector, including public utilities, transport, rent, and other public services.

In addition, [De Chaisemartin and d Haultfoeuille \(2015\)](#) highlight that IV estimates can be far from returns to education in any location when the homogeneity assumption does not hold. However, they show that difference-in-difference (DID) methods with fuzzy treated groups<sup>16</sup> should provide unbiased estimates without relying on any homogeneity assumption, as long as 1) the common trend assumption is valid; 2) there is a control group for which the treatment does not change overtime. According to the above results, these two assumptions seem to be satisfied. In this study, 2) implies there is at least one region where education has not evolved between the pre-treatment period and the treatment period, which is precisely the case of Zanzibar West.<sup>17</sup> This lack of education increase has two explanations: education has already reached the maximum years of primary education in 1967<sup>18</sup> and access to secondary education was cut at time of the UPE program.

Last but not least, IV estimates are biased if the program has influenced outcomes other than education that explain the level of consumption. Thinking of forced

---

<sup>16</sup>This refers to DID when the intensity of the treatment varies between treated groups.

<sup>17</sup>The education level decreases by 0.1 year between 1967 and 1978 which is negligible.

<sup>18</sup>Zanzibar, independent in 1964, benefited from a better access to education.

villagization, this is very likely to happen. Among the possible channels, the program could have changed the access to other social services and the living conditions. Nonetheless, this should not call into question the validity of the instrument in this particular case. Indeed, the specificity of the villagization program is that all individuals were concerned since entire families were asked to move. As a consequence, the treatment group and the control group were similarly affected by these changes. On the contrary, the education policy consisting in education reforms was beneficial only for the treatment group and had no reason to affect outcomes other than education.

### 2.3.2 The impact of the UPE program on education expansion

Since education may be endogenous, I adopt a two-stage procedure, the first stage of which is :

$$S_{ijt} = \alpha + \beta_j + \beta_t + \gamma T * I_{j,1967} + \delta t * X_{j,1967} + \mu_{ijt} \quad (2.3)$$

$\beta_j$  and  $\beta_t$  are region-of-birth fixed effects and birth-cohort fixed effects to account for permanent differences across regions and over time.  $T$  is a dummy taking the value 0 for people belonging to  $T_0$  and 1 for people belonging to either  $T_1$ ,  $T_2$ ,  $T_3$  or  $T_{tot}$ .  $I_{j,1967} * T$  captures the intensity of the UPE program and  $X_{j,1967}$  is a set of region characteristics. It includes the log of population aged 7 to 13 and regional GDP by sectors of activity in 1967. These controls are interacted with a time variable  $t$ . The coefficient of interest,  $\gamma$ , represents the effect of the UPE program on education (years of schooling). When  $I_{j,1967}$  increases, the education expansion between  $T_0$  and the treatment groups are expected to be larger.

Table 2.3 reports the results of equation 2.3. I distinguish the effect for the whole treatment group  $T_{tot}$  and by age cohorts  $T_1$ ,  $T_2$  and  $T_3$ . To consider the possible serial correlation in errors, I cluster standard errors at the regional level (Bertrand *et al.*, 2004). Without GDP controls, I find that when the predicted intensity  $I_{j,67}$  is raised by one additional year between the control group  $T_0$  and  $T_{tot}$ , the education expansion increases by 0.34 between the two periods. This result is consistent with the idea that the UPE

TABLE 2.3 : EFFECT OF THE PROGRAM ON EDUCATION:  $\gamma$  COEFFICIENTS OF EQUATION (2.3)

Dependant variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$T_{tot}$		T1		T2		T3	
Years of education	0.369*** (0.064)	0.336*** (0.060)	0.277*** (0.057)	0.226*** (0.053)	0.418*** (0.050)	0.373*** (0.078)	0.435*** (0.065)	0.443*** (0.089)
R-squared	0.276	0.273	0.318	0.315	0.339	0.337	0.298	0.297
F-test	33.38	31.83	23.70	18.60	70.14	23.34	45.55	25.97
Primary completion	0.038*** (0.013)	0.030** (0.013)	0.029** (0.013)	0.015 (0.013)	0.045*** (0.013)	0.031* (0.017)	0.042*** (0.013)	0.036** (0.017)
R-squared	0.238	0.238	0.289	0.291	0.329	0.332	0.283	0.285
F-test	8.940	5.314	5.181	1.467	12.89	3.367	10.22	4.472
Sample	T0 +T1+T2+T3		T0 +T1		T0 +T2		T0 +T3	
Cohort FE	yes	yes	yes	yes	yes	yes	yes	yes
Regiont FE	yes	yes	yes	yes	yes	yes	yes	yes
GDP Controls	no	yes	no	yes	no	yes	no	yes
Observations	440,676	423,419	199,450	190,148	190,466	182,454	247,502	238,723

Source: the 2002 census. Notes: Standard errors are clustered at the region of birth level and are reported in parentheses. \*\*\*, \*\*, \* mean respectively that the coefficient are significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are the population aged 7 to 13 in 1967, the household size and the principal sector of activity of the household head.

program targeted regions with low initial education attainment and contributed to the equalization of access to education among regions. We can deduce from columns (3), (5) and (7) that the effect of this program was progressive: when  $I_{j,1967}$  increases by one year, the education expansion increases by 0.28 years of education from T0 to T1, by 0.42 years of education from T0 to T2 and by 0.44 years of education from T0 to T3. These results are in line with the expectations: the effect becomes larger when the exposure to the treatment increases. For all estimations, F-test values are large. When I introduce GDP controls interacted with a time trend (columns (2), (4), (6) and (8)). The effect of  $I_{j,1967}$  is still positive and significant but the coefficient is slightly lower. It is worth noting that the introduction of this control changes the sample of the analysis since the regional GDP variables are only available only for mainland Tanzania. The bottom panel of Table 2.3 indicates whether the UPE plan had fully reached its goal by convincing people not only to enroll in school but also to complete primary education. It is not obvious since primary education lasts seven years and the UPE plan was strictly implemented for four years. When I control by regional GDP, reported result for  $T_{tot}$  indicates that  $I_{j,1967}$  significantly increases the primary completion by 3.1 percent age points. However, F-test values are lower, suggesting that  $I_{j,1967}$  is a much stronger predictor for the number of years of education than for primary completion.

As a robustness check, I also test the impact of the density of schooling infrastructures  $N_{j,1967}$ . Since I do not have information of the number of schools in Zanzibar, I choose Kilimanjaro as the reference<sup>19</sup> and the intensity is rewritten as a decreasing function of the number of schools  $I'_{j,67} = (N_{Kilimanjaro,67} - N_{j,67})$ . As a result, the first-stage equation becomes:

$$S_{ijt} = \alpha + \beta_j + \beta_t + \gamma T * I'_{j,1967} + \delta_t X_{j,1967} + \mu_{ijt} \quad (2.4)$$

Table A2.4 presents the estimates for mainland Tanzania. Coefficients are positive and significant for all treatment groups. When the intensity increases by one additional school per square kilometer, the education expansion between the control and the treatment groups is raised by 0.05 year to 0.08 years of education, depending on the specification.

I also estimate equations 2.3 and 2.4 using the LSMS data. Results, gathered in Table A2.5, provide a less clear message. First, F-test values are much lower, and second, the effects are not robust to the introduction of GDP controls and to the change of instrument. This can be explained by the smaller sample sizes of the sub-samples.<sup>20</sup> Thereafter, I use only the 2002 census to study the impact of UPE program on education.

I also estimate a more flexible regression that allows the effect of the UPE program to vary according to the time exposure to the program:

$$S_{ijt} = \alpha + \beta_j + \beta_t + \sum_{t=1945}^{1954} \gamma_t I_{j1967} + \sum_{t=1961}^{1978} \gamma_t I_{j1967} + \delta_t X_{j1967} + \mu_{ijt} \quad (2.5)$$

In this equation,  $\gamma_t$  indicates age-cohort coefficients. It measures the effect of the reform by age-cohort. The difference between  $\gamma_t$  and  $\gamma_{t+1}$  represents the education expansion between t and t+1 generated by the education level in 1967. For the pre-treatment group,  $I_{j,67}$  should have no impact on education expansion and  $\gamma_t$  values should be close to 0, while for treated groups, regions with a high predicted intensity  $I_{j,67}$  should benefit from a

---

<sup>19</sup>Kilimanjoro is the second most educated region in 1967, and it also experienced a negligible education expansion from 1967 to 1978.

<sup>20</sup>Since few observations are available by region and by year, I cannot capture any effect.

larger education progression and  $\gamma_t$  values should be increasing . This is precisely what is

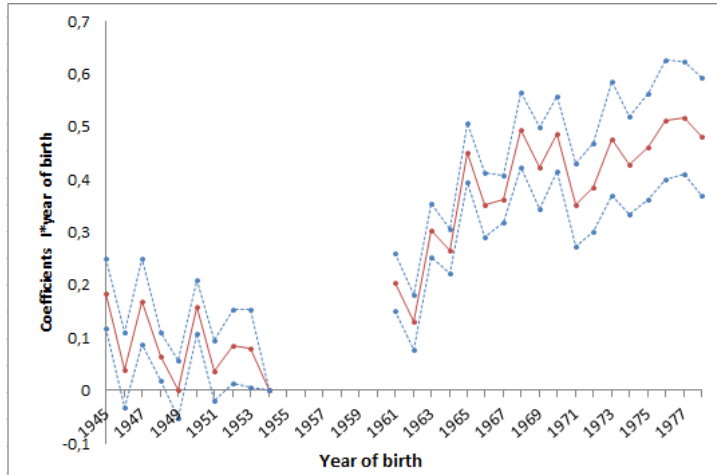


FIGURE 2.2:  $\gamma_t$  COEFFICIENTS OF THE INTERACTION BETWEEN AGE-COHORTS AND EDUCATION LEVEL BY REGION IN 1967.

Source: 2002 census (IPUMS data).

shown in figure 2.2. Each coefficient in this table corresponds to  $\gamma_t$  coefficients of equation 2.5. The reference year is the year before the introduction of the UPE program, in 1967, which corresponds to the age-cohort born in 1954. As expected, most of the coefficients in the pre-treatment program were not statically different from 0. From 1961 to 1978,  $\gamma_t$  coefficients steadily increased by 0.4 point. Cohorts born after 1968 were still exposed but the slope declines afterward. This graph confirms that the identification strategy is reasonable: the trend was not present before the program and the UPE program had a significant impact on education for the treated cohorts (all coefficients are significant at 1% level). Thus, if no regional time-varying characteristics correlated with the program's intensity are omitted, these fuzzy difference-in-difference results should correctly estimate the impact of the UPE program.

## 2.4 Results

This section presents the main results. The first sub-section is devoted to the returns to education for the entire population. To better understand the distribution of these



returns, I also compare in sub-section 2.4.2 the returns to education across different sectors of activity. Education may also have the advantage of increasing the probability that individuals can obtain work in sectors that are better paid. Then, subsection 2.4.4 investigates whether the returns to education come from the returns “within sectors” or from a “distribution effect”.

### 2.4.1 The returns to education

I measure the returns to education by looking at the effect of education  $S_{ijt}$  of household head  $i$  born in region  $j$  at year  $t$  on current consumption  $C_{ijt}$ . The main equation is:

$$\text{Log}(C_{ijt}) = \alpha + \beta_j + \beta_t + \theta S_{ijt} + \delta_t X_j + \epsilon_{ijt} \quad (2.6)$$

where  $\beta_j$  and  $\beta_t$  are, respectively, region-of-birth and year-of-birth fixed effects. Regional controls  $X_j$  are also included. For the sake of comparison with my earlier results, this equation is estimated separately for treatment groups  $T_{tot}$ , T1, T2 and T3.

I first ignore the potential endogeneity of education and run OLS regressions. Several conclusions can be drawn from Table A2.7. In the top panel, the estimates indicate that education increases the log consumption aggregate (constructed using the Deaton and Zaidi method) by 7%. Comparing the middle and bottom panels, the estimates from the consumption proxy  $\widehat{\text{Log}C_{ijt}}$  are very similar for the LSMS data and for the census data and, around 4.2%. Notwithstanding, there is an important gap between consumption estimates and consumption proxy estimates. To correctly interpret these results, it is necessary to clarify under which conditions the proxy for consumption is a relevant outcome.

In traditional settings,  $\theta$  in equation 2.6 captures the causal impact provided that  $S_{ijt}$  is not correlated with  $\epsilon_{ijt}$ . In contrast, using the proxy for consumption introduces an additional assumption:  $\hat{b}$  and  $\widehat{\nu}_{ijt}$  estimates from equation 2.1 should be unbiased. Recall that  $\widehat{\nu}_{ijt}$  was drawn from a normal distribution and was assumed to be exogenous.

Yet,  $\nu_{ijt}$ , the consumption part not explained by households’ dwelling characteristics, results from households’ preferences and is probably influenced by the education of the

households' heads. For instance, households with educated household heads may be more willing to spend money for the education or health of their children. If so, there is a remaining endogenous part of the residual  $\nu''_{ijt}$  that is not captured by  $\widehat{\nu}_{ijt}$ . Thus, equation 2.6 can be rewritten:

$$\widehat{b}X_{jt} + \widehat{\nu}_{ijt} = \alpha + \beta_j + \beta_t + \theta S_{ijt} + \delta_t X_j + \epsilon_{ijt} - \nu''_{ijt} \quad (2.7)$$

And  $\widehat{\theta} = \theta + \frac{\text{cov}(\epsilon_{ijt}, S_{ijt})}{V(S_{ijt})} - \frac{\text{cov}(\nu''_{ijt}, S_{ijt})}{V(S_{ijt})}$ .

The positive correlation between education and  $\epsilon_{ijt}$  leads to the traditional upward bias, while the positive correlation between education and  $\nu''_{ijt}$  causes downward bias in the coefficient of interest. In conclusion, if  $\nu_{ijt}$  is not purely exogenous, using the proxy for consumption adds an additional source of bias. To obtain consistent estimates of  $\theta$ , I instrument education and I rely on the first-stage equations 2.3 and 2.4. Instead of using a dummy treatment variable, I also distinguish the effect of the treatment by age-cohort. I refer to equation (2.5) but I impose that each  $\gamma_{jt}$  equals 0 for the pre-treatment cohorts:

$$S_{ijt} = \alpha + \beta_j + \beta_t + \sum_{t=1961}^{1978} \gamma_t S_{j1967} + \delta_t X_j + \mu_{ijt}$$

In this equation,  $\gamma_t$  identifies the effect of the UPE program by age-cohort in comparison with the pre-program period  $T_0$ .

If the instrument is not correlated with  $\nu$  nor with  $\epsilon$ , and if the relevance condition is valid, IV estimates identify the causal impact of education. Even though this first condition cannot be empirically tested, the intensity of the treatment has no obvious reason to be correlated with  $\nu$  and  $\epsilon$ , except through education.

Table 2.4 reports the 2-SLS estimates for the 2002 census. When I consider the entire treatment period,  $T_{tot}$ , results are robust to the instrument and to specifications. When I add controls for GDP by sectors of activity, I find that one additional year of education of household head increases the log of household consumption between 7.1 and 9.3 %. F-statistics are high, which suggests that the instruments have strong predictive power.

TABLE 2.4 : ESTIMATIONS OF EDUCATION ON  $\widehat{consumption}$

	$T_{tot}$		T1		T2		T3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
OLS estimates								
	0.042***	0.042***	0.045***	0.045***	0.046***	0.046***	0.043***	0.043***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
R-squared	0.574	0.575	0.601	0.603	0.595	0.595	0.581	0.582
IV estimates with $T * I_{j,1967}$								
$S_{j67} * T$	0.075***	0.082***	0.077***	0.067**	0.061***	0.071***	0.079***	0.088***
	(0.019)	(0.022)	(0.028)	(0.029)	(0.020)	(0.017)	(0.016)	(0.019)
R-squared	0.455	0.446	0.510	0.523	0.511	0.505	0.459	0.443
F-test	33.38	31.83	23.70	18.60	70.14	23.34	45.55	25.97
IV estimates with $\sum_{t=1961}^{1978} \gamma_t * i_{j,1967}$								
	0.068***	0.073***	0.065**	0.049**	0.060***	0.067***	0.078***	0.087***
	(0.017)	(0.016)	(0.026)	(0.019)	(0.021)	(0.017)	(0.015)	(0.019)
R-squared	0.468	0.463	0.527	0.537	0.517	0.513	0.464	0.450
F-test	188.7	131.6	24.18	17.73	35.21	9.094	21.86	7.266
Observations	440,676	423,419	199,450	190,148	190,466	182,454	247,502	238,723
IV estimates with $I'_{j,1967} * T_{tot}$								
$N_{j67} * T$	0.076***	0.093***	0.0916***	0.0887***	0.068***	0.078***	0.0721***	0.081***
	(0.020)	(0.025)	(0.0185)	(0.0265)	(0.017)	(0.023)	(0.0207)	(0.026)
R-squared	0.232	0.189	0.305	0.312	0.326	0.312	0.259	0.242
F-test	69.16	21.55	42.31	35.36	97.80	27.36	82.75	20.75
IV estimates with $\sum_{t=1961}^{1978} \gamma_t * I'_{j,1967}$								
	0.067***	0.076***	0.082***	0.072***	0.066***	0.073***	0.072***	0.081***
	(0.019)	(0.021)	(0.016)	(0.018)	(0.018)	(0.021)	(0.021)	(0.026)
R-squared	0.247	0.233	0.326	0.344	0.327	0.318	0.258	0.242
F-test	37.90	39.75	23.67	12.52	36.92	6.640	18.05	7.303
Observations	423,419	423,419	190,148	190,148	182,454	182,454	238,723	238,723
Sample	T0 +T1+T2+T3		T0 +T1		T0 +T2		T0 +T3	
Cohort FE	yes	yes	yes	yes	yes	yes	yes	yes
Region FE	yes	yes	yes	yes	yes	yes	yes	yes
GDP Controls	no	yes	no	yes	no	yes	no	yes

Source: the 2002 census. Notes: Standard errors are clustered at the birth region level and are reported in parentheses. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are the population aged 7 to 13 in 1967, the household's size and the sector of activity.

In comparison with OLS estimates, coefficients are larger. With regard to the ability bias, this result is counter-intuitive. If educated individuals have higher abilities,  $\theta$  captures both the education and the ability effect and OLS estimates are likely to be over-estimated. However, the opposite effect can be observed when education is measured with error (Griliches, 1977) and when returns to education are heterogenous<sup>21</sup> (Card, 2001). The most plausible explanation in this framework is that instrumenting education removes the negative bias from the correlation between education and  $\nu$ . In addition, is worth emphasizing that these IV estimates are close to those in the existing literature (Maluccio (1998), Duflo (2001)).

To test whether IV estimates are unbiased, I implement a series of robustness checks. First, I test whether results are robust when I use the primary education level by region in 1967,  $P_{j,1967}$  instead of  $S_{j,1967}$ .<sup>22</sup> Table A2.8 show that IV estimates are not very sensitive to this choice: the returns are about 1% lower, but the difference is not statically significant. Finally, I test whether the introduction of partial treated individuals  $T_{pt}$  in the treatment group changes the results and I deduce from Table A2.9 that the estimates are very similar and are not statically different.

Heretofore, standard errors are clustered at the level at which the instrument is defined, in other words, at the regional level. Yet, a small number of clusters can lead to over-rejection of standard asymptotic tests (Cameron *et al.*, 2008). To check whether I underestimate the standard errors, I run equation 2.6 by instrumenting education with the intensity  $I_{j',67}$  defined at a lower scale, the district of residence  $j'$ .<sup>23</sup> Since I construct this variable from the district of residence, this method provides biased estimates in case of selective migration.<sup>24</sup> Table A2.10 shows that results are very similar. Coefficients are close and are still significant at the 1% level. This entails that selective migration and

---

<sup>21</sup>When the instrument affects the education choices of less-educated subgroups, which have high marginal returns to education, IV estimates are upward-biased. Regarding the UPE program that focused on individuals with restricted access to primary schools, IV estimates may over-estimate the average marginal returns to education of the entire population.

<sup>22</sup> $P_{j,1967}$  represents the education level but cannot be above 7, the length of primary education.

<sup>23</sup>In 2012, there were 31 regions against 169 districts in Tanzania.

<sup>24</sup>I construct the instrument from the district of residence instead of the district of birth, which is not available.

over-rejecting issues are negligible.

#### 2.4.1.1 Quality bias

The analysis of these IV estimates is controversial if the UPE program affected both the quantity and the quality of education (Duflo, 2001). If so, the UPE program's results would confound these two entangled effects. This massive education program could have lowered the quality of education if it had led to overcrowded rooms (Little, 2008) or if the quality of teachers and the status of teaching had fallen (Towse *et al.*, 2002). To test whether it is true, I compare the returns to education from different treated age-cohorts (T1, T2 and T3) that were affected by the UPE program with different intensities. For instance, the T1 cohort was exposed to the program at the end of their education while the T2 cohort was exposed to the program from the beginning of their education. At this period, primary schools had to face up to an increasing number of pupils. From 1974 to 1978, the number of students in grade 1 increased from 200,000 to 901,770 students and these rapid changes could have lowered the quality of education (King, 1984). On the contrary, the T3 cohort was only indirectly treated by the program and primary enrollment started flattening out for the T3 age-cohort (King, 1984) (see Table A2.11). In the meantime, primary education was exposed to structural changes<sup>25</sup> and quality of education was defined as the new priority (Bonini, 2003). Then, one would expect that quality of education would be lower for T1 and T2 and would be higher for T3. If so, returns to education in T1 and T2 would constitute a lower bound while returns to education in T3 would constitute a higher bound. This is precisely what is observed in Table 2.4. Whatever the instrument and the specification, returns to education are higher for the T3 cohort. However, these differences are small and are almost never significant. This comparison from different age-cohorts suggests that the UPE program may have lowered quality of education but to a small extent.

---

<sup>25</sup>The structural changes started in 1986 when Tanzania signed agreements with the IMF and the World Bank

## 2.4.2 Returns to education by sector of activity

So far, returns to education have been estimated for the whole population. However, they can vary from one sector of activity to another. In this subsection, I investigate this question and I estimate the consumption equation for each sector:

$$\text{Log}(C_{iajt}) = \alpha_a + \beta_{aj} + \beta_{at} + \theta_a S_{ijt} + \delta_{ta} X_j + \epsilon_{iajt} \quad (2.8)$$

where the subscript “a” depicts the household head main activity and indicates whether he: 1) does not work or is unpaid, 2) works in agriculture, 3) works in non-farm self-employed activities, or 4) is a wage-worker.

TABLE 2.5 : IV ESTIMATES OF THE RETURNS TO EDUCATION BY SECTORS

Activity	Don't paid Don't work (1)	agri. (2)	self employed (3)	wage- work (4)	Don't paid Don't work (5)	agri (6)	self employed (7)	wage- work (8)
OLS estimates								
	0.030*** (0.006)	0.033*** (0.002)	0.062*** (0.003)	0.055*** (0.002)	0.029*** (0.006)	0.033*** (0.002)	0.064*** (0.003)	0.055*** (0.002)
R-squared	0.545	0.415	0.505	0.457	0.543	0.416	0.511	0.464
IV estimates: $T * I'_{j,1967}$								
	0.491 (0.871)	0.074*** (0.024)	0.096*** (0.023)	0.098*** (0.024)	0.116 (0.119)	0.097*** (0.028)	0.103*** (0.025)	0.048** (0.024)
F-test	0.400	25.75	72.65	37.01	3.842	8.315	15.69	65.22
IV estimates with sample selection correction (IV for occupation equation: $T * I'_{j,1967}$ )								
	0.312 (0.437)	0.075*** (0.024)	0.095*** (0.026)	0.087*** (0.025)	0.190 (0.183)	0.096*** (0.028)	0.110*** (0.029)	0.048* (0.027)
Mills no work	-0.008				-0.010 (0.008)			
Mills agri.		-0.011*** (0.010)				-0.010*** (0.003)		
Mills self.			0.004 (0.003)				0.004* (0.003)	
Mills wage				0.025** (0.012)				0.020 (0.014)
F-test	0.547	25.85	56.73	28.88	6.677	8.569	13.21	51.74
Cohort FE	yes	yes	yes	yes	yes	yes	yes	yes
Region FE	yes	yes	yes	yes	yes	yes	yes	yes
GDP control	no	no	no	no	yes	yes	yes	yes
Observations	3,518	277,747	81,396	60,536	3,518	277,747	81,396	60,536

Source: the 2002 census. Notes: Standard errors are clustered at the region of birth level and are reported in parentheses. In IV estimations, standard errors are bootstrapped. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are the population aged 7 to 13 in 1967, the household's size and the sector of activity.

The first panel of Table ?? presents the OLS results. It shows that returns to education

are much lower in agriculture than in the non-farm self-employment activities and in the formal (wage work) sector. However, 2SLS estimates (Table ?? and A2.12 ) do not lead to the same conclusion: returns to education are higher in agriculture and in non-farm self-employed activities than in the wage-activities. Differences between agriculture and self-employed activities are not statistically different, but the returns to education are statistically higher in agriculture than in wage-work activities when instrumenting education with  $I'_{j,67}$ .

By comparison, IV estimates are three times, 32% and 15% larger than OLS estimates in agriculture, non-farm self-employed activities and wage-workers activities, respectively. These ratio of IV to OLS estimates illustrates the size of the bias. These discrepancies of bias between sectors have several possible explanations.

First, this may come from changing correlations between education  $S_{ijt}$  and the error terms  $\epsilon_{ijt}$  and  $\nu_{ijt}$ . For instance, if working in the formal sector requires higher abilities, the correlation between  $S_{ijt}$  and  $\epsilon_{ijt}$  may be higher and entails a larger upward bias for wage-workers. Besides  $\nu_{ijt}$ , the determinants of households' consumption unexplained by dwelling characteristics are likely to be correlated to education of the household's head  $S_{ijt}$ . However, there is no clear reason for differences in these correlations across sectors of activity.<sup>26</sup> Last, but not least, these differences of bias can be explained by the nature of the UPE program itself. IV estimates probably capture an estimate of LATE, representative of the specificities of the UPE program and of the schooling curriculum at that time. Indeed, the aim of this program was to improve agricultural skills and to boost rural productivity through agricultural classes (Kinunda, 1975). This could explain why individuals who benefited from this program have higher returns to education when they work in agriculture.

---

<sup>26</sup>Figures A2.1 and A2.2 depict the distribution of education spending and food items. Expenditures are slightly lower in agriculture but the distributions look similar.

### 2.4.2.1 Sample selection bias

Contrary to other studies, I do not need to adopt an imputation method because consumption, the outcome of interest, is available for every household. Despite this, I still may encounter selection issues when I estimate the returns to education for non-random sub-samples such as sectors of activity. To address this possible sample selection bias, I adopt the two-stage model proposed by Wooldridge (2010) that deals with the endogeneity of education and the selection issue by using an exclusion restriction variable in the first stage equation and in the selection equation (see appendix 2.5.1 for more details). Since the UPE program was expected to influence both access to education and the labor market organization, I use the intensity of the UPE program, as the excluded variable for both equations.

Results with sample selection correction are reported at the bottom panel of Table ?? and A2.12 . The introduction of sample selection corrections does little to change the 2SLS estimates: the returns are still much higher in the agricultural sector and in the self-employed sector while they are lower in the formal sector.<sup>27</sup> Furthermore, there is little evidence of sample selection bias. Coefficients of the Mill's ratio are close to 0 and are not statistically significant except in the agricultural sector. Thus, estimating the consumption by sectors of activity does not seem to generate significant sample selection issues.

### 2.4.3 Effect of education on the labor market participation

Education can also ease the access to sectors that require skilled labor. In a rural country where the government promoted education to increase the agricultural productivity, it could be interesting, both from a macroeconomic and a microeconomic perspective, to identify the effect of education on the distribution between the sectors of activity. I estimate a multinomial logit model where  $A_{ijt}$  is the sector of activity, taking the value of

---

<sup>27</sup>When instrumenting education with  $I'_{j,67}$ , the difference between agriculture and wage-work activities is still statistically significant.



1 if the individual does not work or are unpaid, 2 if the individual works in the agricultural sector, 3 if the individual is self-employed in non-farm activities and 4 if he has a wage-work. The activity equation has the following functional form:

$$A_{ijt} = \alpha + \beta_j + \beta_t + \theta S_{ijt} + \delta_t X_j + \epsilon_{ijt} \quad (2.9)$$

To avoid endogeneity issues, I instrument education by the exposure to the UPE program and I follow a two-step Control Function approach (Wooldridge, 2014). After obtaining the predicted residual from the first stage equations, I plug it into equation (2.9). This predicted residual is also used to test the endogeneity of education.

TABLE 2.6 : AVERAGE MARGINAL EFFECT OF EDUCATION ON THE PROBABILITY OF WORKING IN EACH SECTOR OF ACTIVITY (MULT. LOGIT)

Activity	Don't paid Don't work (ref) (1)	agri. (2)	self employed (3)	formal (4)	Don't paid Don't work (ref) (5)	agri. (6)	self employed (7)	formal (8)	
OLS	-0.001*** (0.000)	-0.003* (0.002)	-0.004*** (0.001)	0.007*** (0.000)	-0.001*** (0.000)	-0.003* (0.002)	-0.004** (0.002)	0.007*** (0.000)	
Observations	443,541				426,261				
Instrument: $I'_{j1967} * T_{tot}$	0.000 (0.001)	0.017*** (0.004)	-0.020*** (0.005)	0.002 (0.003)	-0.000 (0.001)	0.015** (0.007)	-0.010* (0.006)	-0.005 (0.004)	
$\hat{\mu}_{ijt}$		0.045 (0.074)	0.336*** (0.080)	0.312*** (0.076)		0.059 (0.158)	0.300* (0.160)	0.374** (0.157)	
F-test		33.25				30.96			
Observations	443,541				426,261				
Instrument: $I'_{j1967} * T_{tot}$	0.001 (0.001)	0.010** (0.004)	-0.013*** (0.004)	0.003* (0.002)	0.001 (0.001)	0.006* (0.003)	-0.007** (0.003)	-0.001 (0.002)	
$\hat{\mu}_{ijt}$		0.095 (0.059)	0.284*** (0.071)	0.287*** (0.069)		0.111 (0.074)	0.3066*** (0.103)	0.328*** (0.088)	
F-test		69.53				48.14			
Observations	426,261				426,261				
Cohort FE	yes	yes	yes	yes	yes	yes	yes	yes	
Region FE	yes	yes	yes	yes	yes	yes	yes	yes	
GDP control	no	no	no	no	yes	yes	yes	yes	

Sources: 2002 census. Notes: Standard errors, reported in parentheses, are bootstrapped and clustered at the birth region level. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%. CF-IV: IV estimates with control function method. Additional controls are the population aged 7 to 13 in 1967 and the household's size.

Results are reported Table 2.6 . From OLS estimates, I observe that education decreases the probability of being unemployed, of working in agriculture and in non-agriculture self-employed activities while it increases the probability of having a wage

work. However, IV estimates show a completely different picture. Education raises the probability of working in agriculture and reduces the probability of being self-employed in non-farm activities. These effects, robust to the instrument used, probably identify the LATE explained by the specificity of the UPE program. In most estimations, the predicted residuals are statistically different from 0 which confirms the importance of dealing with the endogeneity of education.

#### 2.4.4 Decomposition of the education effects

To investigate the relative impact of education on consumption and on the probability of working in each sector of activity, I base my analysis on the following expected consumption:  $E(C) = \sum_{a=1}^n P_a * C_a$ , where  $P_a$  is the probability of working in the sector of activity a and  $C_a$  is the consumption level of individuals working in activity a. If education impacts both the choice of the sector of activity and the level of consumption in these respective sectors, I can decompose the education effect in two parts:

$$\frac{\delta E(C)}{\delta S} = \overbrace{\sum_{a=1}^n \frac{\delta P_a}{\delta S} * C_a}^{\text{distribution effect}} + \overbrace{\sum_{a=1}^n P_a * \frac{\delta C_a}{\delta S}}^{\text{intra sector effect}} \quad (2.10)$$

The first term represents the monetary benefit of education due to the change in the distribution between sectors and the second term corresponds to the returns to education within sectors. More specifically,  $\frac{\delta P_a}{\delta S}$  is the effect of education on the probability of working in activity a (see Table 2.6 ) and  $\frac{\delta C_a}{\delta S}$  is the return to education by activity (see Table A2.12 and Table ??).  $C_a$  and  $P_a$  are approximated by the predicted values of  $\hat{C}_a$  and  $\hat{P}_a$  from equation 2.8 and equation 2.9, respectively.

Table 2.7 provides results from equation 2.10. OLS estimates show that both the distribution and the intra-return effects are positive and significant. I conclude from this decomposition that total returns to education come only from the intra-returns effect. IV estimates suggest that the intra-returns effect is the main effect while the distribution effect is much small and slightly negative. Since education increases the probability of

TABLE 2.7 : THE CUMULATIVE EFFECT OF EDUCATION

Model	(1)	(2)	(3)	(4)	(5)	(6)
	OLS		IV: $S_{j,1967} * T_{tot}$		IV: $N_{j,1967} * T_{tot}$	
Distribution effect	0.0018*** (0.001)	0.002*** (0.001)	-0.0023*** (0.003)	-0.002*** (0.003)	-0.002*** (0.005)	-0.002*** (0.001)
Intra sector effect	0.042*** (0.010)	0.042*** (0.010)	0.081*** (0.0029)	0.093*** (0.014)	0.081*** (0.007)	0.091*** (0.014)
Cohort FE	yes	yes	yes	yes	yes	yes
Region FE	yes	yes	yes	yes	yes	yes
GDP Control	no	yes	no	yes	no	yes
Observations	440,452	423,191	440,452	423,191	423,191	423,191

Source: the 2002 census. Notes: Standard errors are clustered at the region of birth level and are reported in parentheses. Since results are produced from a multi-stage procedure, standard errors are bootstrapped. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are the population aged 7 to 13 in 1967, the household's size and the sector of activity.

working in the agricultural sector (see section 2.4.3), and average consumption is lower in this sector, this effect is not surprising and illustrates the specificity of the UPE program.

## 2.5 Conclusion

This paper studies the benefits of education in Tanzania and considers two particular aspects, household consumption and the choice of the sectors of activity. To instrument education of household heads, I exploit variation in time and in space of the Universal Education Program, supported by the villagization from 1974 to 1978.

I find that this massive primary education program contributed to a reduction of inequalities among regions. After this program ended, its effects persisted for the next age-cohorts. Despite the controversial means of villagization, the Tanzanian government fulfilled its goals by improving access to basic education, even in remote areas. Unfortunately, several changes were implemented at the same time, which prevents one from identifying the relative efficiency of each policy.

By using a household survey, census data, and records on the number of schools, I find that education increases household consumption between 7.3 and 9.3 percent, depending on the specification and the instrument. This analysis has the advantage of focusing on the entire population, instead of wage-workers, who are in the minority in most developing

countries and are very likely to be self-selected. I also compare the returns to education between sectors of activity. I find that the returns to education are higher in agriculture and in non-farm self-employed activities than in wage-work activities. This conclusion, at first sight surprising, is consistent with the Tanzanian governmental policy that aimed to put education at the service of agriculture by teaching agricultural skills. Compared to the few studies on the benefits of primary education in agriculture in African countries that find low returns (Appleton *et al.*, 1996; Jolliffe, 2004), I argue that returns to education in agriculture are positive, provided that the curriculum at school is consistent with agriculture. This gets closer to Foster and Rosenzweig (1996)'s results suggesting that returns to education are positive only during specific contexts such as during technological changes, when education helps farmers to adopt new technologies.

This suggests that the introduction of agricultural classes could help households to escape poverty by increasing the farmers productivity. In terms of public recommendations, this result is all the most relevant in a context where the large majority of individuals work in rural activities and where the government has a limited range of intervention tools to support farmers.

## Appendix A2

### 2.5.1 Measuring the effect of education by sector of activity with the Heckman selection model

To overcome endogeneity issues and selection issues, I follow the Wooldridge (2010) 's method. It consists in estimating three different equations :

$$\begin{aligned}\log(C_{iajt}) &= \alpha_{1a} + \beta_{1aj} + \beta_{1at} + \theta_{1a}S_{ijt} + \delta_{1ta}X_j + \epsilon_{1isjt} \\ A_{ijt} &= \alpha_{2a} + \beta_{2aj} + \beta_{2at} + \theta_{2a}I_{j,67} * T + \gamma_{2a}N_{ijt} + \delta_{2ta}X_j + \epsilon_{2iajt} \\ S_{ijt} &= \alpha_{3a} + \beta_{3aj} + \beta_{3at} + \theta_{3a}I_{j,67} * T + \delta_{3ta}X_j + \epsilon_{3iajt}\end{aligned}$$

The first equation is the equation of interest, the consumption equation by sector of activity a. The second equation is the selection equation.  $A_{ijt}$  is the main sector of activity of households head (unemployed or domestic unpaid workers, wage-workers , self-employed workers in non agricultural activities and self-employed workers in agriculture). The third equation represents the endogenous education equation of household i.

$I_{j1967}$ , is the instrument for both  $S_{ijt}$  and for  $A_{ijt}$ . To obtain unbiased estimates of the impact of education on consumption, I compute the inverse Mills ratios  $\hat{\lambda}_{ia}$  with from the predicted probabilities in the selection equation. Then, I introduce the inverse Mills ratios into the consumption equation :

$$\text{Log}(C_{iajt}) = \alpha_{1a} + \beta_{1aj} + \beta_{1at} + \theta_{1a}S_{ijt} + \delta_{1ta}X_j + \gamma_{1a}\hat{\lambda}_{ia} + \epsilon_{1iajt}$$

This equation is estimated by 2SLS, using  $I_{j,67}$  as an instrument. Standard errors have to be bootstrapped to account that it is a two-step procedure. The sample selection issue can be tested by checking whether  $\gamma_{1a}$  is significantly different from 0. As Wooldridge (2010) underlines, if the same instruments are used for the occupational equation and for the consumption equation, the introduction of the Mills ratio generates collinearity that may affect performance in the case of small samples. Since, sub-samples' size are very large in this analysis, this collinearity issue should be limited.

## 2.5.2 Sample and statistic descriptives

TABLE A2.1 : DESCRIPTIVE STATISTICS  
FROM THE 2002 CENSUS

	$T_0$	$T_{tot}$
Edu. main man	4.29	5.98
Edu. main woman	2.89	4.96
$\widehat{\log consumption}$	14.185	14.238
Activity of main man (%)		
Does not work/unpaid	6.4	6.1
Agriculture	63.1	55.9
Self-employed	11.9	22.1
Wage worker	18.5	15.8
Activity of main woman (%)		
Does not work/unpaid	19	23.6
Agriculture	69.7	58.6
Self-employed	6.6	11.5
Wage worker	4.6	6.2
same activity	34.7	34.8
both actif	43.6	42.5
rural	61.5	53.9
HH size	5.039	4.023
Obs.	111,818	388,701

Sources: The 2002 census (IPUMS data).

TABLE A2.2 : DESCRIPTIVE STATISTICS FROM THE LSMS-ISA  
DATA

Year	2008	2010	2012
Composition of samples			
Number of districts	126	128	131
Number of wards	87	103	108
Number of HH	3265	3921	5004
Number of children	4512	5239	6236
Household characteristics			
Number of adults	3.208	3.425	3.511
Number of children	3.315	3.321	3.305
Household production (TSH)	246957.4	372482.4	708005.2
cultivated area (acres)	4.822	4.548	5.759
Children characteristics			
Child's age	11.407	11.370	11.335
Child is female	0.507	0.503	0.507
Child is enrolled in school	0.815	0.883	0.858
Number of completed grade	4.295	4.388	4.324
Child dropout school this year	0.043	0.050	0.061
Child has repeated a grade this year	0.123	0.117	0.129
Child reads and writes	.	0.733	0.720
Child works last week	0.114	0.422	0.262
Child works last year	0.116	0.444	0.382

### 2.5.3 Construction of the proxy for consumption

TABLE A2.3 : EFFECT OF DWELLING CHARACTERISTICS ON CONSUMPTION

VARIABLES	log (consumption)
Solid wall	0.148*** (0.015)
Housing water	0.124*** (0.019)
Flush toilet	0.040** (0.016)
Electricity	0.388*** (0.019)
Permanent floor	0.379*** (0.017)
Solid roof	0.478*** (0.055)
Nb. of bedrooms	0.093*** (0.005)
Age HH head	-0.002*** (0.000)
Gender HH head	-0.107*** (0.014)
Nb. child aged 5-15	0.092*** (0.004)
Nb. adult aged 16-65	0.157*** (0.004)
Constant	12.566*** (0.041)
R-squared	0.532
Observations	12,178

Sources: The three pooled waves of the LSMS data. Notes: additional controls: Regions dummies, survey year dummies. Standard errors are reported in parentheses. \*\*\*, \*\*, \* mean respectively that the coefficient are significantly different from 0 at the level of 1%, 5% and 10%.



FIGURE A2.1: DISTRIBUTION OF EDUCATION SPENDINGS.

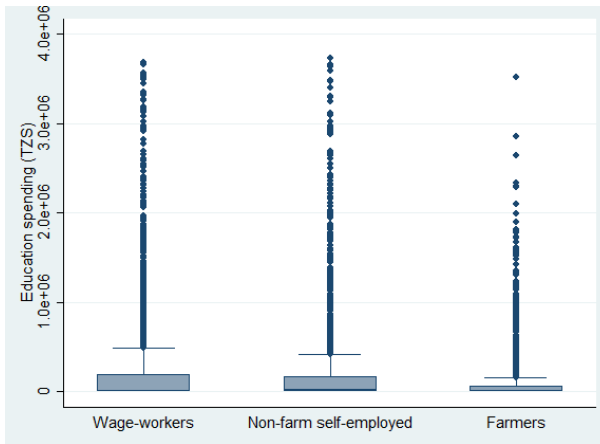
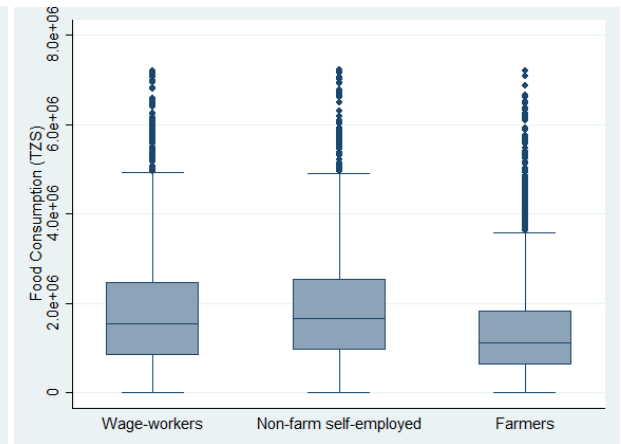
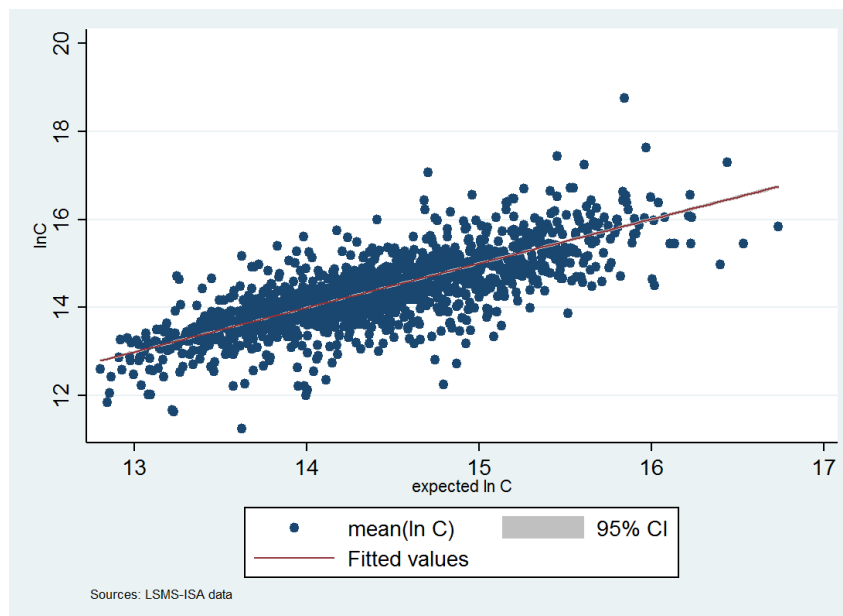


FIGURE A2.2: DISTRIBUTION OF FOOD SPENDINGS.



Sources: The LSMS data (2008, 2010, 2012).

FIGURE A2.3: RELATIONSHIP BETWEEN THE EXPECTED CONSUMPTION  $\widehat{\ln C}$  AND  $\ln C$ .



Source: LSMS data (2008, 2010, 2012)

## 2.5.4 First stages

TABLE A2.4 : EFFECT OF THE PROGRAM ON THE EDUCATION LEVEL:  $\gamma$  COEFFICIENTS OF 2.4

Outcomes	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$T_{tot}$		T1		T2		T3	
years of education	0.052*** (0.006)	0.045*** (0.010)	0.041*** (0.006)	0.038*** (0.006)	0.060*** (0.006)	0.058*** (0.011)	0.063*** (0.007)	0.065*** (0.014)
R-squared	0.272	0.273	0.315	0.315	0.336	0.337	0.296	0.297
F-test	69.53	21.66	42.44	34.62	99.29	28.42	82.55	20.73
Primary completion	0.006*** (0.001)	0.006*** (0.001)	0.005*** (0.001)	0.005*** (0.001)	0.008*** (0.001)	0.008*** (0.001)	0.007*** (0.001)	0.008*** (0.002)
R-squared	0.238	0.239	0.290	0.291	0.332	0.333	0.285	0.286
F-test	30.45	19.10	15.48	32.71	35.75	30.60	28.03	17.83
Cohort FE	yes	yes	yes	yes	yes	yes	yes	yes
Region FE	yes	yes	yes	yes	yes	yes	yes	yes
GDP Control	no	yes	no	yes	no	yes	no	yes
Observations	423,419	423,419	190,148	190,148	182,454	182,454	238,723	238,723

Source: the 2002 census. Notes: Standard errors are clustered at the birth region level and are reported in parentheses. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are the population aged 7 to 13 in 1967, the household's size and the sector of activity.

TABLE A2.5 : EFFECT OF THE PROGRAM ON THE EDUCATION LEVEL FROM THE HOUSEHOLD SURVEY ( $\gamma$  COEFFICIENTS OF 2.3).

Instrument	(1)	(2)
$I_{j67} * T_{tot}$	0.198*	0.370
	(0.103)	(0.239)
R-squared	0.203	0.198
F-test	3.690	2.399
Observations	5,820	4,982
$I'_{j67} * T_{tot}$	0.00700	0.0281
	(0.0179)	(0.0182)
R-squared	0.176	0.178
F-test	0.152	2.395
Observations	4,982	4,982
Cohort FE	yes	yes
Regiont FE	yes	yes
GDP Control	no	yes

Source: the pooled LSMS survey (2008, 2010, 2012). Notes: Standard errors are clustered at the region of birth level and are reported in parentheses. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are the population aged 7 to 13 in 1967, the household's size and the sector of activity.

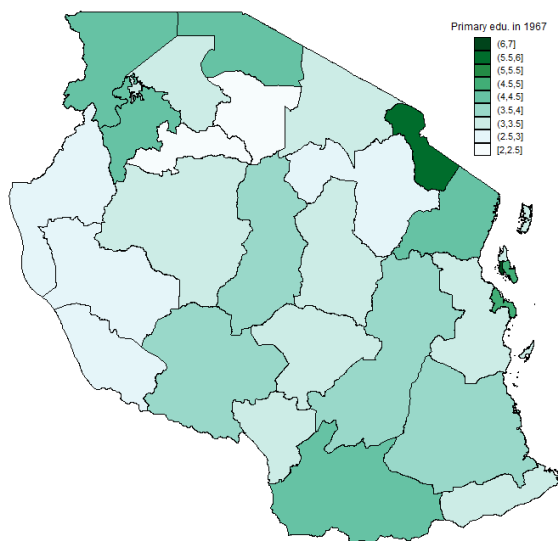
TABLE A2.6 : EVOLUTION OF EDUCATION ATTAINMENT BY PERIOD AND REGION GROUPS

Age-cohort	Region -	Region +	Difference (region+ - region-)
Tb	1.536 (.0119)	2.806 (.0227)	1.270 (.0270)
T0	2.571 (.0128)	4.249 (.0221)	1.677 (.0255)
T1	4.746 (.0128)	6.265 (.0186)	1.518 (.0226)
T2	4.524 (.0119)	6.675 (.0167)	1.150 (.0205)
T3	5.392 (.008)	6.529 (.0119)	1.137 (.0144)
Difference (T0-Tbaseline)	1.035 (.0175)	1.443 (.0316)	.408 (.0362)
Difference (T1-T0)	2.175 (.0181)	2.016 (.0288)	-.158 (.0340)
Difference (T2-T0)	2.952 (.0175)	2.426 (.0276)	-.527 (.0328)
Difference (T3-T0)	2.821 (.0152)	2.280 (.0251)	-.541 (.0293)

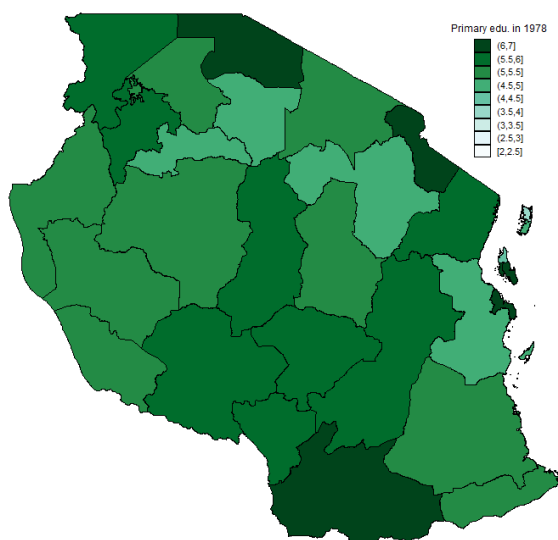
Source: IPUMS data, 2002. Notes: Standards errors are in parenthesis. Region + represents regions with education in 1958 higher than 3 years of education, the average primary education level by this time. Region - represents regions with education in 1968 lower than 3 years of education, the average primary education level by this time. Tb: individuals born between 1935 and 1945.

FIGURE A2.4: ACCESS TO EDUCATION IN TANZANIA

(A) EDUCATION LEVEL IN 1967



(B) EDUCATION LEVEL IN 1978



(C) DISTRIBUTION OF PRIMARY SCHOOL BY REGION IN 1967

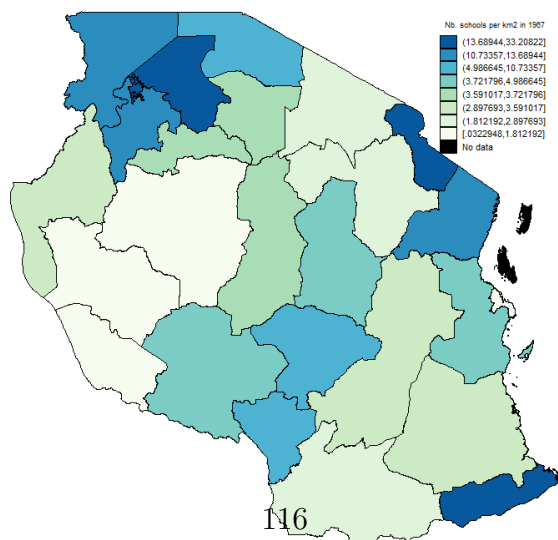
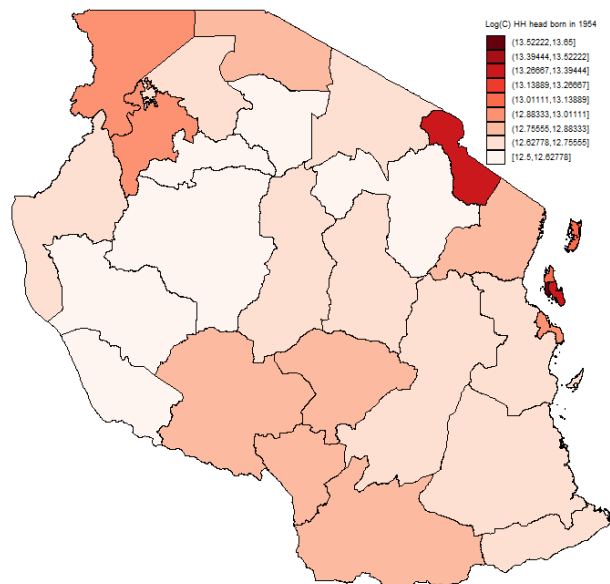
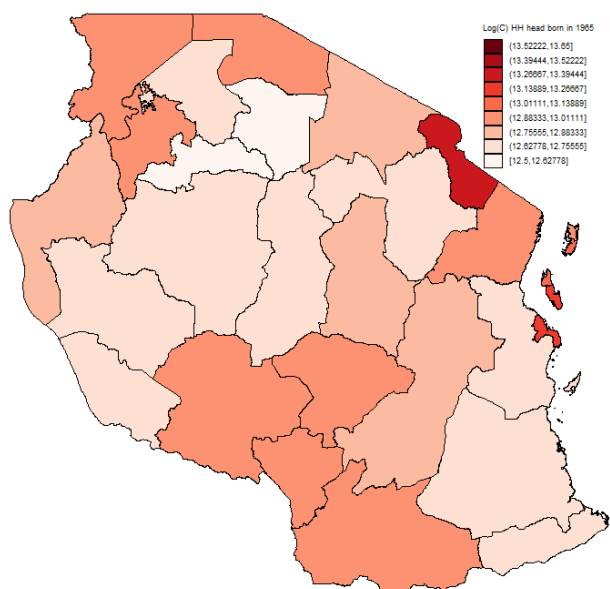


FIGURE A2.5: CONSUMPTION LEVEL OF HOUSEHOLD HEADS

(A) CONSUMPTION OF HOUSEHOLD HEADS  
BORN IN 1967



(B) CONSUMPTION OF HOUSEHOLD HEADS  
BORN IN 1978



Source: LSMS-ISA data (2008, 2010, 2012).

## 2.5.5 Robustness Checks :

TABLE A2.7 : OLS ESTIMATES OF THE RETURNS TO EDUCATION

	$T_{tot}$		T1		T2		T3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LSMS: Deaton and Zaidi consumption aggregate $\log(C)$								
	0.070***	0.071***	0.069***	0.069***	0.070***	0.070***	0.072***	0.073***
	(0.004)	(0.004)	(0.005)	(0.005)	(0.007)	(0.007)	(0.005)	(0.005)
R-squared	0.425	0.425	0.478	0.480	0.443	0.447	0.463	0.464
LSMS: $\widehat{\log(C)}$								
	0.041***	0.041***	0.043***	0.043***	0.044***	0.044***	0.042***	0.0427***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.00325)
R-squared	0.570	0.569	0.633	0.634	0.604	0.606	0.583	0.584
Observations	5,820	4,982	2,699	2,282	2,556	2,138	3,215	2,816
IPUMS: $\widehat{\log(C)}$								
	0.042***	0.042***	0.045***	0.045***	0.046***	0.046***	0.043***	0.043***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
R-squared	0.574	0.575	0.601	0.603	0.595	0.595	0.581	0.582
Sample	T0 +T1+T2+T3		T0 +T1		T0 +T2		T0 +T3	
Cohort FE	yes	yes	yes	yes	yes	yes	yes	yes
Region FE	yes	yes	yes	yes	yes	yes	yes	yes
GDP Control	no	yes	no	yes	no	yes	no	yes
Observations	440,676	423,419	199,450	190,148	190,466	182,454	247,502	238,723

Notes: Standard errors are clustered at the birth region level and are reported in parentheses. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are the population aged 7 to 13 in 1967, the household size and the sector of activity of the household head.

TABLE A2.8 : EFFECT OF EDUCATION ON CONSUMPTION : IV ESTIMATIONS WITH EDUCATION LEVEL AND PRIMARY EDUCATION LEVEL.

	$T_{tot}$		T1		T2		T3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
IV estimates with $I_{j67} = S_{Zanzibar\ West,67} - S_{j67}$								
	0.075***	0.082***	0.077***	0.067**	0.061***	0.071***	0.079***	0.088***
	(0.019)	(0.022)	(0.028)	(0.029)	(0.020)	(0.017)	(0.016)	(0.019)
R-squared	0.455	0.446	0.510	0.523	0.511	0.505	0.459	0.443
F-test	33.38	31.83	23.70	18.60	70.14	23.34	45.55	25.97
IV estimates with $P_{Zanzibar\ West,67} - P_{j67}$								
	0.071***	0.075***	0.073**	0.059*	0.059***	0.065***	0.077***	0.082***
	(0.020)	(0.023)	(0.030)	(0.031)	(0.021)	(0.017)	(0.016)	(0.020)
R-squared	0.241	0.233	0.340	0.358	0.334	0.329	0.251	0.239
F-test	27.08	26.14	17.98	13.43	51.24	19.18	34.04	22.15
IV estimates with $\sum_{t=1961}^{1978} \gamma_t * I_{j1967}$								
	0.068***	0.073***	0.065**	0.049**	0.060***	0.067***	0.078***	0.087***
	(0.017)	(0.016)	(0.026)	(0.019)	(0.021)	(0.017)	(0.015)	(0.019)
R-squared	0.468	0.463	0.527	0.537	0.517	0.513	0.464	0.450
F-test	188.7	131.6	24.18	17.73	35.21	9.094	21.86	7.266
IV estimates with $\sum_{t=1961}^{1978} \gamma_t * (P_{Zanzibar\ West,67} - P_{j67})$								
	0.064***	0.065***	0.061**	0.040*	0.056**	0.060***	0.075***	0.079***
	(0.018)	(0.017)	(0.028)	(0.021)	(0.022)	(0.017)	(0.016)	(0.019)
R-squared	0.251	0.250	0.354	0.363	0.335	0.334	0.255	0.245
Additional CC	YES	YES	YES	YES	YES	YES	YES	YES
F-test	297.0	174.3	27.39	24.90	41.74	7.430	31.61	7.827
Sample	T0 +T1+T2+T3		T0 +T1		T0 +T2		T0 +T3	
Cohort FE	yes	yes	yes	yes	yes	yes	yes	yes
Region FE	yes	yes	yes	yes	yes	yes	yes	yes
GDP Controls	no	yes	no	yes	no	yes	no	yes
Observations	440,676	423,419	199,450	190,148	190,466	182,454	247,502	238,723

Source: the 2002 census. Notes: standard errors are clustered at the birth region level and are reported in parentheses. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are the population aged 7 to 13 in 1958, the percentage of people living in rural areas in 1958, the household's size and the sector of activity.

TABLE A2.9 : EFFECT OF EDUCATION ON CONSUMPTION: IV ESTIMATIONS WITH PARTIALLY TREATED.

	$T_{tot}$		$T_{tot}$ and partial treated	
	(1)	(2)	(3)	(4)
$I_{j67} * T$	0.075***	0.082***	0.083***	0.097***
	(0.019)	(0.022)	(0.018)	(0.031)
R-squared	0.234	0.220	0.234	0.196
F-test	33.38	31.83	26.34	13.50
Sample	$T_0 + T_1 + T_2 + T_3$		$T_0 + T_{pt} + T_1 + T_2 + T_3$	
Cohort FE	yes	yes	yes	yes
Region FE	yes	yes	yes	yes
GDP Control	no	yes	no	yes
Observations	440,676	423,419	516,061	495,082

note: Source: the 2002 census. Standard errors are clustered at the birth region level and are reported in parentheses. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are the population aged 7 to 13 in 1967, the household's size and the sector of activity.



TABLE A2.10 : IV ESTIMATIONS AT DIFFERENT GEOGRAPHICAL SCALES

Instrument	(1)	(2)
$I_{j,1967} * T$	0.075*** (0.019)	0.082*** (0.022)
R-squared	0.455	0.446
F-test	33.38	31.83
$I_{j',1967} * T$	0.074*** (0.014)	0.073*** (0.020)
R-squared	0.243	0.262
F-test	52.81	74.00
Sample	$T_0 + T_1 + T_2 + T_3$	
Cohort FE	yes	yes
Region FE	yes	yes
GDP Control	no	yes
Observations	440,676	423,419

Source: the 2002 census. Notes: standard errors are clustered at the birth region level and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are the population aged 7 to 13 in 1967, the household's size and the sector of activity.

FIGURE A2.6: EVOLUTION OF EDUCATION ATTAINMENT BY REGION FROM T0 TO T1 ACCORDING TO THE EDUCATION LEVEL IN T0.

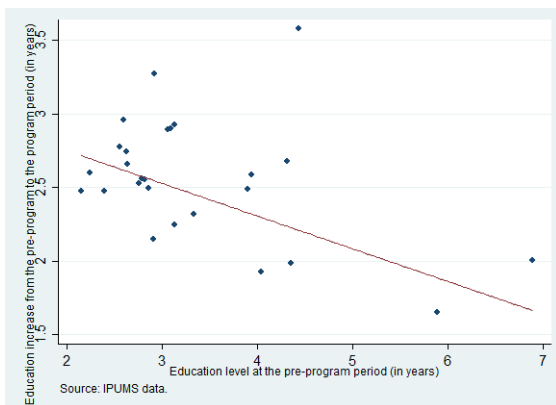


FIGURE A2.7: EVOLUTION OF EDUCATION ATTAINMENT BY REGION FROM T0 TO T1 ACCORDING TO THE NUMBER OF SCHOOLS IN 1967.

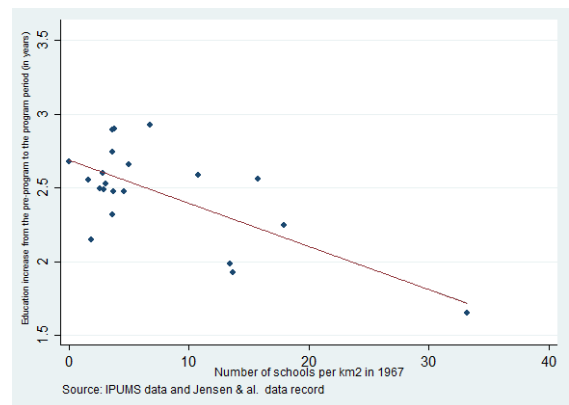
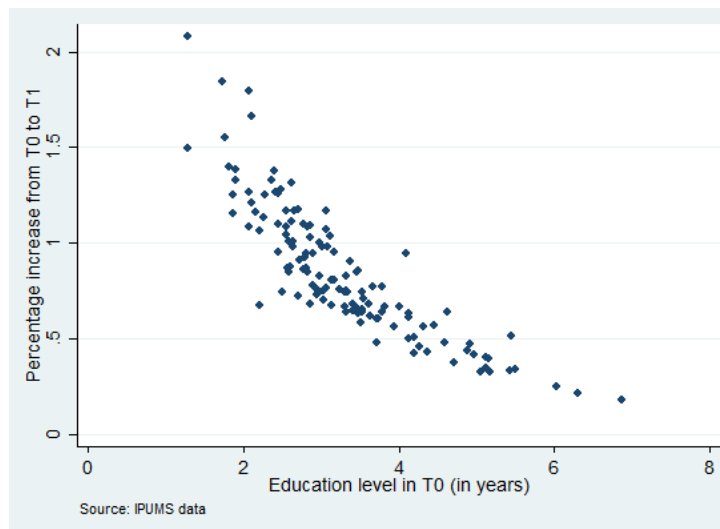


TABLE A2.11 : EFFECT OF  
EDUCATION ON THE WEALTH INDEX  
(INCOME INDEX CONSTRUCTED FROM  
THE 2002 CENSUS DATA)

	(1)	(2)
OLS estimates		
	0.051***	0.051***
	(0.003)	(0.003)
R-squared	0.470	0.465
IV estimates		
$I_{j67} * T$	0.105***	0.128***
	(0.0158)	(0.0263)
R-squared	-0.007	-0.066
F-test	33.37	31.82
$\sum_{t=1961}^{1978} \gamma_t * I_{j1967}$	0.096***	0.116***
	(0.017)	(0.023)
R-squared	0.011	-0.033
F-test	195.8	132.7
Sample	T0 +T1+T2+T3	
Cohort FE	yes	yes
Region FE	yes	yes
GDP Control	no	yes
Observations	440,683	423,426

Source: the 2002 census. Notes: standard errors are clustered at the birth region level and are reported in parentheses. \*\*\*,\*\*,\* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are the population aged 7 to 13 in 1967, the household's size and the sector of activity.

FIGURE A2.8: EVOLUTION OF THE EDUCATION ATTAINMENT BY DISTRICT  
ACCORDING TO THE EDUCATION LEVEL IN 1967.



Sources: The 2002 census.

TABLE A2.12 : IV ESTIMATES OF THE RETURNS TO EDUCATION BY SECTOR

Activity	Don't paid Don't work (1)	agri. (2)	self employed (3)	formal (4)	Don't paid Don't work (5)	agri (6)	self employed (7)	formal (8)
OLS estimates								
	0.030*** (0.006)	0.033*** (0.002)	0.062*** (0.003)	0.055*** (0.002)	0.029*** (0.006)	0.033*** (0.002)	0.064*** (0.003)	0.055*** (0.002)
R-squared	0.545	0.415	0.505	0.457	0.543	0.416	0.511	0.464
IV estimates : $T * I_{j,1967}$								
	0.173 (0.306)	0.086*** (0.019)	0.095*** (0.024)	0.089*** (0.027)	0.100 (0.110)	0.101*** (0.026)	0.085** (0.033)	0.064* (0.034)
F-test	0.181	18.63	38.89	21.74	1.619	10.11	21.66	25.08
IV estimates with sample selection correction (IV for occupation and first-stage equation: $T * I_{j,1967}$ )								
	0.179 (0.313)	0.088*** (0.019)	0.100*** (0.025)	0.087*** (0.026)	0.103 (0.109)	0.103*** (0.026)	0.086** (0.034)	0.068** (0.034)
Mills no work	-0.011* (0.006)				-0.011 (0.009)			
Mills agri		-0.009*** (0.003)				-0.010*** (0.003)		
Mills self.			0.005** (0.002)				0.003 (0.003)	
Mills wage				0.026** (0.011)				0.022 (0.014)
F-test	0.181	19.05	33.21	22.18	1.625	10.18	20.80	23.79
Cohort FE	yes	yes	yes	yes	yes	yes	yes	yes
Region FE	yes	yes	yes	yes	yes	yes	yes	yes
GDP control	no	no	no	no	yes	yes	yes	yes
Observations	3,576	284,127	87,582	65,167	3,518	277,747	81,396	60,536

Source: the 2002 census. Notes: standard errors are clustered at the region of birth level and are reported in parentheses. In IV estimations standard errors are bootstrapped. \*\*\*,\*\*,\* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%. Additional controls are the population aged 7 to 13 in 1967, the household's size and the sector of activity.

FIGURE A2.9: EDUCATION LEVEL BY DISTRICT IN 1967.

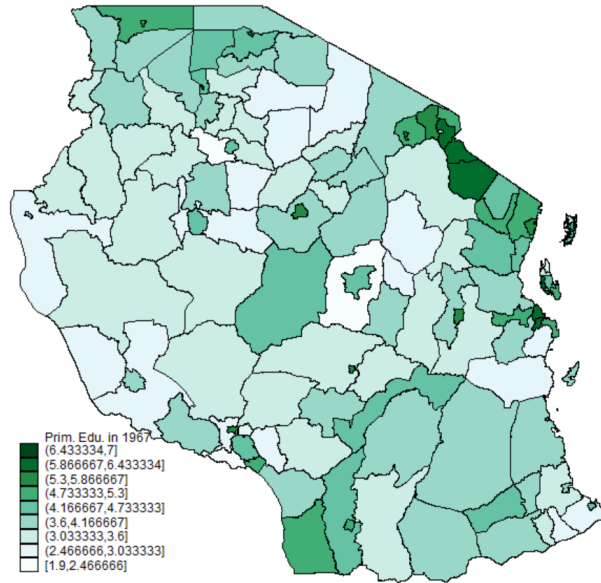
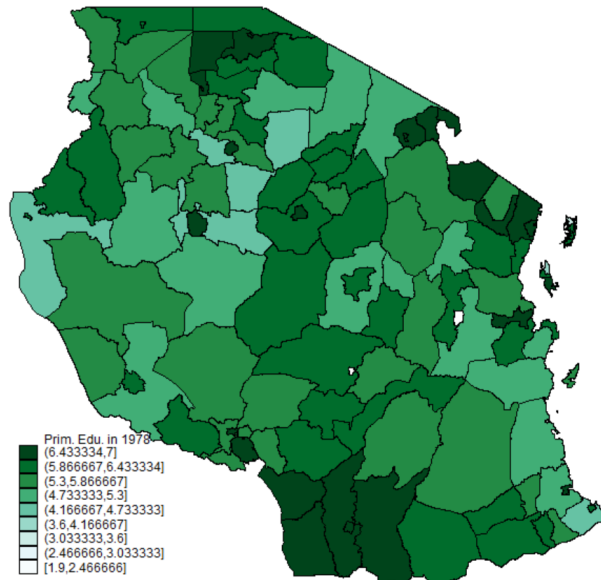
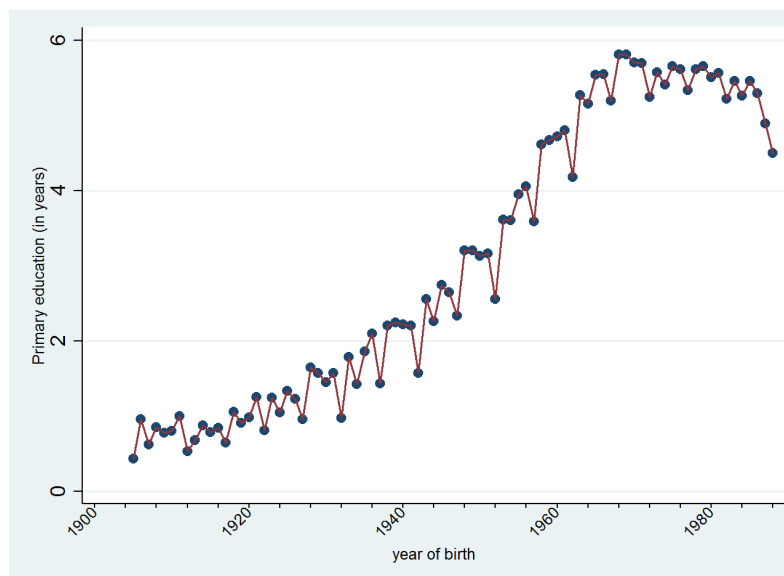


FIGURE A2.10: EDUCATION LEVEL BY DISTRICT IN 1978.



Sources: The 2002 census.

FIGURE A2.11: PRIMARY EDUCATION LEVEL (IN YEARS) BY AGE COHORTS.



Sources: The 2002 census.

## Chapter 3

Good or bad timing? The  
pro-cyclical and counter-cyclical  
effect of shocks on education  
investment and on schooling  
performance.



## 3.1 Introduction

The United Nations Environment Programme (UNEP) records that natural disasters such as droughts and floods have been multiplied by two over the last 25 years. Concurrently, the Food Agriculture Organization (FAO) highlights that commodity prices have become more volatile. This increasing number of shocks, which sparks politicians' and academics' interest, strikes most sub-Saharan African countries, where agriculture is the dominant activity. To face these productivity shocks in an environment where formal markets are imperfect, households may use child labor and cut back on education investment (Jacoby and Skoufias, 1997b). This raises concerns about transmission of inequalities from one generation to the next and perpetuation of poverty. In this paper, I focus on climate and price shocks to investigate the impacts of productivity shocks on education decisions and on schooling performance.

Theoretically, this relationship is not straightforward. Productivity shocks induce an income effect and a substitution effect that move in opposite directions. The income effect results from the change in the available income for education, and is expected to vary with households financial constraints and access to formal insurance.<sup>1</sup> In contrast, the substitution effect which results from the change in the labor productivity is likely to vary with access to the labor market<sup>2</sup> (?).

Ferreira and Schady (2009) conduct a literature review on the relationship between productivity shocks and education and show that negative productivity shocks worsen education investments in developing countries – in other words, education is a procyclical outcome. To capture the causal effect of negative shocks on education, several authors focus on transitory crop shocks that reflect agricultural crises. Whether they consider pests, rodents, birds and locusts (Gubert and Robilliard, 2007), adverse weather shocks (Jensen, 2000), or drastic falls in cash-crop prices (Cogneau and Jedwab, 2012), they find

---

<sup>1</sup>A large body of the literature finds that when households have large assets or have access to credit markets, they do not need to call on marginal workers such as children to cope with shocks (?).

<sup>2</sup>To take advantage of a labor productivity increase, households can use hired labor instead of increasing child labor.



that these unanticipated income shocks reduce school enrollment. These results indicate that the income effect outweighs the substitution effect.

With this consideration in mind, a closely related question is whether the effects of productivity shocks are symmetric, so that positive shocks encourage education investments. More recently [Shah and Steinberg \(2017\)](#) argue the opposite and show that higher rainfalls in India lower education achievement. Similarly, [Beegle \*et al.\* \(2006b\)](#) and [Boozer and Suri \(2001\)](#) also find that larger rainfalls increase child labor, reducing school year and enrollment. By comparison, [Kruger \(2007\)](#) looks at the effects of coffee price increase in Brazil. All these papers have in common that they look at positive shocks rather than negative ones. Taken together, these empirical findings point out that the relationship between productivity shocks and schooling decisions is not linear.

The first contribution of this paper is to precisely check whether positive and negative productivity shocks hinder education. To investigate this question, I consider two main outcomes, education decisions and schooling performance. These two dimensions are complementary and allow one to capture different changes in education investment.<sup>3</sup>

Second, this paper aims to assess the effect of repetition of shocks. Although households can develop strategies in the short-run to cope with income shocks, they may not necessarily manage to protect the education of their children when the shocks are recurrent. Looking at the effect of the number and the length of shocks is all the more relevant when we consider schooling performance, which is the result of a cumulative process.

Finally, the relationship between productivity shocks and education can be analyzed with respect to the child's age at which the shock occurs. A large literature focuses on the fetal origins hypothesis, claiming that nutritional shocks in early life have severe and permanent consequences on human capital accumulation ([Almond and Currie, 2011](#); [Currie and Vogl, 2013](#)). [De Vreyer \*et al.\* \(2014\)](#) take the locust plague in Mali as a natural experiment and find that children born during the plague have a lower educational attainment. However, the heterogeneity of these effects along children's age has been less

---

<sup>3</sup>Productivity shocks may change the time spent at school without changing enrollment.

studied. Yet, [Shah and Steinberg \(2017\)](#) highlight that the substitution effect may be substantially different if the opportunity cost of time evolves as children age. In this respect, the third contribution of this paper is to provide a dynamic picture by looking at the effects of shocks from birth to date.

Based on a simple human capital model with two periods, I first identify the underlying mechanisms behind productivity shocks in early life and in school age. Then, I use the LSMS-ISA panel survey (2008 to 2012) and the Uwezo cross section survey (2010 to 2014) in Tanzania to study the relationship between productivity shocks and education outcomes. To capture exogenous productivity variation, I consider weather shocks and variation in cash-crop prices that are expected to increase available incomes for education and to change the opportunity cost of children's time. I conduct a geographically disaggregated analysis where the identification strategy exploits variation in the intensity of shocks across geographical areas and over time.

The main findings suggest that current productivity shocks affect education decisions but have little effect on schooling performance. In contrast, when schooling shocks are recurrent, schooling performance significantly drops. In line with [Shah and Steinberg \(2017\)](#), I also find that this relationship depends on the age at which the shock occurs. In early-life, the relationship is procyclical since positive productivity shocks are favorable to future schooling performance. When children are of school age and can work, the relationship becomes counter-cyclical.

The remainder of this paper is organized as follows: section 2 outlines the conceptual model, section 3 describes the background, the data and the shocks variables, section 4 presents the empirical strategy and the results and section 5 introduces some robustness checks .

## **3.2 Framework**

In this section, I provide a simple human capital model to understand how parents allocate their children's time when there is a labor productivity shock.

I assume unitary households in which rational agents maximize their utility over two time periods. In the first period,  $t_1$ , children are too young to go to school and to work, while in the second period,  $t_2$ , children can do both activities. The parents' utility is a function of their consumption in the two periods, and of child's cognitive skills  $A$ :

$$U = U(C_1, C_2, A; X) \quad (3.1)$$

I consider that  $U$  is an increasing strictly quasi-concave function in  $C_1$ ,  $C_2$ , and  $A$ , and add  $X$  a vector of households characteristics. Parents care about their child's cognitive skills for two possible reasons: either they have pure preferences for education, or they anticipate that education will make them better off later in life.<sup>4</sup> These cognitive skills are acquired according to the following production function:

$$A = \alpha A(C_1, C_2, E_2) \quad (3.2)$$

Where  $\alpha$  depicts the child's learning efficiency, which depends on the child's innate ability, the child's motivation, and the parents' motivation (Glewwe, 2002).  $A$  is an increasing function of the time spent at school  $E_2$ , of consumption in early childhood  $C_1$ , and of current consumption  $C_2$ . The nutrition-learning nexus assumption is supported by the World Health Organization (WHO), which emphasizes that stunting has long-lasting consequences on the health and education of children.<sup>5</sup> For the sake of simplicity, I assume that cognitive skills do not depend on households' spending for education.<sup>6</sup> In the Tanzanian setting, this assumption seems plausible since very few children go to preschool, and 93.5% of enrolled children aged 7 to 16 go to free public schools. In  $t_2$ , parents decide to allocate total child's time  $T_2$  between schooling attendance  $E_2$  and labor  $L_{2c}$ <sup>7</sup>:

---

<sup>4</sup>If the returns to education are positive, educated children will be able to send larger transfers to their parents in the future.

<sup>5</sup>Stunting is defined as height-for-age is being minus two standard deviations below the mean.

<sup>6</sup>This assumption can be relaxed without changing the results.

<sup>7</sup>Child leisure is neglected but this assumption does not change the model's interpretations. Results are available upon request.

$$T_2 = E_2 + L_{2c} \quad (3.3)$$

In the two periods, households spending correspond to the available income  $I$  and are expressed as follows:

$$C_1 = w_1 L_{1a} (1 - \Delta) = I_1(w_1, \Delta) \quad (3.4)$$

$$C_2 = w_2 L_{2a} + \gamma w_2 L_{2c} + \Delta w_1 L_{1a} \quad (3.5)$$

Equation (3.5) can be rewritten as:

$$C_2 + \gamma w_2 E_2 = w_2 L_{2a} + \gamma w_2 T_{2c} + \Delta w_1 L_{1a} = I_2(w_1, w_2, \gamma, \Delta)$$

where  $w_1$  and  $w_2$  denote labor productivity on the farm, commonly called the shadow wage.  $L_{1a}$  and  $L_{2a}$  stand for adult labor in the two periods, and  $\gamma \in [0, 1]$  is the relative productivity of child labor compared to adult labor. Based on the literature's findings in developing countries, I assume that credit, saving and labor markets are imperfect (Jacoby, 1993; Skoufias, 1994; Chavas *et al.*, 2005; Le, 2009). Although households have no access to formal markets, I suppose they can cope with income shocks by informally saving a fraction of their income  $\Delta \in [0, 1]$  from  $t_1$  to  $t_2$ .

For a sake of simplicity, I assume that  $\Delta$  is exogenous and does not vary with current labor productivity. This hypothesis can be relaxed without changing interpretations of the model.<sup>8</sup>

By substituting (3.2) in (3.1), I express household utility as a direct function of consumption and education:

$$U = U(C_1, C_2, A; X) = \tilde{U}(C_1, C_2, E_2; X) \quad (3.6)$$

---

<sup>8</sup>The amount of transfer between the two periods becomes  $\Delta w_1 L_{1a} + \frac{\partial \Delta}{\partial w_1} w_1 L_{1a}$ . It is unclear whether parents increase or decrease savings with respect to labor productivity, but interpretations of the model remain the same unless parents stop saving.

Parents maximize their utility by choosing  $E_2$  and  $L_{2c}$  subject to the budget constraints (3.4) and (3.5) with respect to  $C_1$ ,  $C_2$ , and  $E_2$  given  $w_1$ ,  $w_2$ ,  $\gamma$ ,  $\Delta$ ,  $X$  and  $T_2$ .<sup>9</sup> The Marshallian demand functions, which depend on the relative prices and the available income  $I_2(w_1, w_2, \gamma, \Delta, u)$ , are written:

$$C_1 = C_1(w_1, \Delta, I_1(w_1, \Delta, u); X) = C_1(w_1, \Delta, w_1(1 - \Delta)L_{1a}, u); X) \quad (3.7)$$

$$E_2 = E_2(w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta, u); X) = E_2(w_2, \gamma, \Delta, \Delta w_1 L_{1a} + w_2(\gamma T_2 + L_{2a}); X) \quad (3.8)$$

$$C_2 = C_2(w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta, u); X) = C_2(w_2, \gamma, \Delta, \Delta w_1 L_{1a} + w_2(\gamma T_2 + L_{2a}); X) \quad (3.9)$$

The corresponding Hicksian demand functions that minimize the total expenditure to maintain a fixed level of utility  $u$  are expressed:

$$C_1 = C_1^*(w_1, \Delta, u; X) \quad (3.10)$$

$$E_2 = E_2^*(w_2, \gamma, \Delta, u; X) \quad (3.11)$$

$$C_2 = C_2^*(w_2, \gamma, \Delta, u; X) \quad (3.12)$$

Based on this basic framework, I analyze how early life and current productivity shocks affect the demand for education and the schooling performance.

---

<sup>9</sup>I assume that, in the short-run, children's education does not change children's productivity.

### 3.2.1 The effect of shocks which occur during schooling

To estimate the effect of productivity shocks which occur during schooling, I compute the partial derivatives of the Marshallian demand (3.8) and (3.9) with respect to  $w_2$ :

$$\frac{\partial E_2(w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta))}{\partial w_2} = \frac{\partial E_2(w_2, \gamma, \Delta, I_2)}{\partial w_2} + \frac{\partial E_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} \frac{\partial I_2(w_1, w_2, \gamma, \Delta)}{\partial w_2} \quad (3.13)$$

$$\frac{\partial C_2(w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta))}{\partial w_2} = \frac{\partial C_2(w_2, \gamma, \Delta, I_2)}{\partial w_2} + \frac{\partial C_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} \frac{\partial I_2(w_1, w_2, \gamma, \Delta)}{\partial w_2} \quad (3.14)$$

From equations (3.11) and (3.12), and the Shepherd's lemma, I obtain the Slutsky equations that allow to decompose the effects of  $w_2$ :

$$\frac{\partial E_2(w_2, \gamma, \Delta, I_2)}{\partial w_2} = \frac{\partial E_2^*(w_2, \gamma, \Delta)}{\partial w_2} - \frac{\partial E_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} \gamma E_2 \quad (3.15)$$

$$\frac{\partial C_2(w_2, \gamma, \Delta, I_2)}{\partial w_2} = \frac{\partial C_2^*(w_2, \gamma, \Delta)}{\partial w_2} - \frac{\partial C_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} \gamma E_2 \quad (3.16)$$

Then, I substitute (3.15) and (3.16) into (3.13) and (3.14) respectively, to express the total effect of  $w_2$  on the current demand for education and consumption:

$$\frac{\partial E_2(w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta))}{\partial w_2} = \frac{\partial E_2^*(w_2, \gamma, \Delta)}{\partial w_2} + \frac{\partial E_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} (\gamma L_{2c} + L_{2a}) \quad (3.17)$$

$$\frac{\partial C_2(w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta))}{\partial w_2} = \frac{\partial C_2^*(w_2, \gamma, \Delta)}{\partial w_2} + \frac{\partial C_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} (\gamma L_{2c} + L_{2a}) \quad (3.18)$$

The first terms of the RHS stand for the substitution effects arising from the change of the relative prices between  $C_2$  and  $E_2$  when the purchasing power remains the same. The quasi-concavity of  $U$  entails that this first term is negative in (3.17) and positive in (3.18): when the labor productivity  $w_2$  gets larger, education demand decreases, while consumption becomes relatively cheaper.

The particularity of this framework is that a change in the labor productivity  $w_2$  generates two income effects. The first income effect is induced by the increase of

the opportunity cost of education ( $\frac{\partial E_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} E_2$  and  $\frac{\partial C_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} E_2$ ), and the second income effect is induced by the endowment reevaluation ( $\frac{\partial E_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} (\gamma T_{2c} + L_{2a})$  and  $\frac{\partial C_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} (\gamma T_{2c} + L_{2a})$ ). On one hand, education becomes more expensive and reduces the available income and on the other hand, children and adults working in the fields become more productive and increase the available income. The second terms of the RHS denotes the sum of these two income effects. As a result, the total effect of a change in  $w_2$  has an ambiguous effect on the education demand  $E_2$ , but has a positive effect on current consumption  $C_2$ .

Using (3.2), I can also deduce the effect of productivity shocks on cognitive skills:

$$\begin{aligned} \frac{\partial A(w_1, w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta))}{\partial w_2} &= \frac{\partial A(w_1, w_2, \gamma, \Delta, I_2)}{\partial C_2} \frac{\partial C_2(w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta))}{\partial w_2} + \\ &\quad \frac{\partial A(w_1, w_2, \gamma, \Delta, I_2)}{\partial E_2} \frac{\partial E_2(w_2, \gamma, \Delta, I_2((w_1, w_2, \gamma, \Delta))}{\partial w_2} \\ &= \frac{\partial A(w_1, w_2, \gamma, \Delta, I_2)}{\partial C_2} \left( \frac{\partial C_2^*(w_2, \gamma, \Delta)}{\partial w_2} + \frac{\partial C_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} (\gamma L_{2c} + L_{2a}) \right) + \\ &\quad \frac{\partial A(w_1, w_2, \gamma, \Delta, I_2)}{\partial E_2} \left( \frac{\partial E_2^*(w_2, \gamma, \Delta)}{\partial w_2} + \frac{\partial E_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} (\gamma L_{2c} + L_{2a}) \right) \end{aligned} \quad (3.19)$$

Productivity shocks have a direct positive effect on  $A$  through  $C_2$  and an indeterminate indirect effect on  $A$  through  $E_2$ . The relative weight of the effects hinges on the form of the cognitive skills production function. In conclusion, variation in contemporaneous productivity has an indeterminate effect on education and on children's cognitive skills. It depends on the relative size of the substitution effects and the income effects, that are likely to vary with access to markets.

### 3.2.2 The effect of shocks which occur in early life

An increase in the labor productivity  $w_1$  has a clear positive effect on the available income in  $t_1$ :

$$\frac{\partial C_1(w_1, \Delta, I_1(w_1, \Delta))}{\partial w_1} = (1 - \Delta)L_{1a}$$

Thus, the saving should increase the available income in  $t_2$  and encourage parents to send their children to school:

$$\frac{\partial E_2(w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta))}{\partial w_1} = \frac{\partial E_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} \frac{\partial I_2(w_1, w_2, \gamma, \Delta)}{\partial w_1} = \frac{\partial E_2}{\partial I_2} \Delta L_{1a} \quad (3.20)$$

Based on this last expression and the functional form of  $A$ , I express the effect of early-life shocks on cognitive skills as follows:

$$\begin{aligned} \frac{\partial A(w_1, w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta))}{\partial w_1} &= \frac{\partial A(w_1, w_2, \gamma, \Delta, I_2)}{\partial C_1} \frac{\partial C_1(w_1, \Delta, I_1(w_1, \Delta))}{\partial w_1} + \\ &\quad \frac{\partial A(w_1, w_2, \gamma, \Delta, I_2)}{\partial E_2} \frac{\partial E_2(w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta))}{\partial w_1} \quad (3.21) \\ &= \frac{\partial A(w_1, w_2, \gamma, \Delta, I_2)}{\partial C_1} (L_{1a}(1 - \Delta)) + \frac{\partial A(w_1, w_2, \gamma, \Delta, I_2)}{\partial E_2} \frac{\partial E_2(w_2, \gamma, \Delta, I_2)}{\partial I_2} \Delta L_{1a} \end{aligned}$$

I deduce from this equation that early-life shocks support the educational achievement through two channels. The first channel stems from the nutrition-learning nexus: when labor productivity increases in early-life, children benefit from a better nutrition, which eases the development of cognitive skills in the long run. The second channel stems from the fact that positive productivity shocks increase the available income for education through larger transfers.

Thus, in early-life, the effect of positive productivity shocks on education outcomes is clear and is not counter-balanced by any substitution effect.



### 3.2.3 Heterogeneity of the effect of shocks

In this sub-section, I point out potential sources of heterogeneity that can be empirically tested.

#### 3.2.3.1 By access to markets

To analyze the impacts of shocks, the first dimension to account for is the access to formal insurance markets (Jacoby and Skoufias, 1997b). When insurance markets are perfect, households are able to perfectly smooth their consumption over time. Thus, shocks do not affect resources for education and no income effect arises from them. In contrast, a substitution effect may still stem from productivity shocks when labor markets are imperfect. Credit and savings market also allow households to smooth their consumption, but the life-cycle income still declines. Credit markets should protect education against negative productivity shocks which reduce the available income, while labor markets protect education should protect against positive productivity shocks which increase the labor productivity and the opportunity cost of education. In most developing countries, including Tanzania (?), access to labor and credit markets remains limited and households are not able to fully protect themselves against productivity shocks.

#### 3.2.3.2 By households characteristics

Household income can also explain variations in the effect of productivity shocks, but the effect is ambiguous. On one hand, the substitution effect should be larger for poor households,<sup>10</sup> but on the other hand, the income effect is larger for rich households who own more lands and assets. Thus, it is not clear whether rich households react more strongly to productivity shocks than poor households.

To analyze the effect of shocks in  $t_2$  on education, I focus on children aged 7 to 16 years old. To go one step further and check whether children react differently by age,

---

<sup>10</sup>If the utility is a strictly positive concave function of  $C_1$  and  $C_2$ , poor households have a higher marginal utility of consumption and have larger incentive to drop their children out of school when education becomes more expensive.

I split school-aged children in two sub-groups, primary-aged children, and secondary-aged children. To determine whether there is heterogeneity in the effects between these groups, it is necessary to highlight their differences. Firstly, older children should be more productive and have a larger work capacity. If so, the substitution effect will be larger for the older age-group. Second, primary school is free, while tuition fees are charged in secondary schools.<sup>11</sup> According to UNESCO (2013), secondary education fees were between 30,000 and 40,000 TSH in 2009, which amounts to half of the average Tanzanian monthly wage. This indicates that households sending their children to secondary schools are likely to be richer. Finally, contrary to primary education, secondary education is not mandatory. Therefore, parents keeping their children in secondary schools may have stronger preferences for education and be less reactive to productivity shocks. In conclusion, it is not clear whether the substitution effect will be larger for the older cohort but this discrepancy can be empirically tested.

### **3.2.3.3 By shock characteristics**

Turning to the effect of shocks, I also investigate whether the effect of shocks depends on their frequency and their length. To test this assumption, I examine the effect of the recurrence and the length of shocks that occur since the beginning of children's schooling. Indeed, if cognitive skills are the result of a cumulative learning process, they should not strictly depend on current shocks, but also on past shocks.

## **3.3 Data**

I bring together geo-referenced data from different sources to analyze the relationship between income shocks and education. In this section, I present the data sources and describe the construction of the main variables.

---

<sup>11</sup>The secondary fees were later removed in November 2015 with the implementation of the Education and Training Policy.

### 3.3.1 Education and child labor data

I exploit two sources of data to measure education outcomes and child labor. The first dataset is the Tanzanian LSMS-ISA (LSMS-Integrated Surveys on Agriculture) panel data, which consist of three rounds (2008-2009, 2010-2011 and 2012-2013). The survey was designed to be nationally representative and covers the entire country. The precise location of households is given by GPS coordinates. Due to the high split-off and the low attrition rate, 3,265 households were interviewed in 2008, 3924 in 2010 and 5,015 in 2012. These data are relevant for this analysis because they gather information on children's activities at the individual level. To examine the effect of productivity shocks on child labor, I look at two variables: whether children have ever worked in the 12 months preceding the survey and the number of days of labor performed in the fields over this period. To understand time allocation decisions, I also consider the number of hours by activity over the last week. Figure A3.1 shows that children mostly work in agriculture and that labor intensity increases with age. Girls work slightly less in agriculture but this is largely compensated by their higher involvement in domestic activities.

The LSMS datasets also provide appropriate information on current education decisions (enrollment and dropping out) and on education outcomes resulting from successive education decisions (grade achievement, whether the child is overage, whether the child has ever been enrolled). Figures A3.2 displays enrollment rate by age. It reaches a maximum at age 11 and decreases afterward. In order to capture education outputs, I also look at the ability of children to read and write. However, this subjective variable is reported by the household head and has to be interpreted with caution. Table A3.1 presents more descriptive statistics on education and child labor from 2008 to 2012.

To better understand the relationship between productivity shocks and cognitive skills, I rely on the Uwezo dataset. The survey is a repeated cross section from 2010 to 2014 and is representative at the district level.<sup>12</sup>

---

<sup>12</sup>I do not include the first round in the analysis because it was a pilot program which covered less than one third of the Tanzanian districts.

The Uwezo program launched by the Twaweza organization seeks to collect test scores for children aged 7 to 16. By including about 100,000 children from more than 50,000 households spread over 4000 villages at each round, these data constitute a national assessment of learning. They have the strong advantage of providing test scores not only for enrolled children but also for children that have never been to school or have dropped out of school. This is not the case for most of the existing data on learning outcomes, which are available only for enrolled children. The questionnaire gathers basic information on children, households and tests scores (Table A3.2 presents some descriptive statistics).

These tests scores are constructed following the Pratham model<sup>13</sup> and are divided into two modules, the literacy test and the numeracy test.<sup>14</sup> All children take the same tests that assess competencies of Standard 2 (Grade 2), a level by which children should have acquired basic reading and numeracy skills. Figure A3.4 shows that, in practice, very few children of standard 2 age validate these skills, instead most children learn them when they grow older. For each test, several competencies of gradual difficulties are assessed and the computed score corresponds to the highest validated competency. The competencies in literacy are 1) letter recognition, 2) word recognition, 3) ability to read a sentence, 4) ability to read a paragraph and 5) text comprehension, while the competencies tested in numeracy are 1) counting, 2) number recognition, 3) ability to rank two numbers, 4) addition, 5) subtraction and 6) multiplication. At each round, these scores are missing for about 2% of children. Since the percentage of missing scores remains negligible, I drop these observations.<sup>15</sup> Thereafter, I standardize the tests<sup>16</sup> and use them as a proxy for cognitive skills. Figure A3.5 depicts inequalities of test scores across districts in Tanzania in 2011.

---

<sup>13</sup>The Pratham model, developed by an Indian NGO, establishes a methodology to evaluate learning outcomes of young children.

<sup>14</sup>Swahili is the official language at school. However, to be able to compare the literacy test across countries where the survey was conducted (Uganda and Kenya), children also take an English test.

<sup>15</sup>I may encounter a selection bias if missing scores are not random. To address this issue, I impute a score and I find that results are not sensitive to the inclusion of children with imputed test scores. Results are available upon request.

<sup>16</sup>I compute the deviation from the mean at each wave.

### 3.3.2 Climate data

Tanzania is an agriculture-based country where 80 % of the population lives in rural areas and where agriculture constitutes half of the GDP. This dependency makes Tanzania vulnerable to many production shocks. Among these shocks, climate shocks constitute one of the main risks that farmers face. The Center for Global Development classifies countries according to their climate change vulnerability, and ranks Tanzania at the 20th most vulnerable in the world out of 55 countries (Wheeler, 2011). The low diffusion of irrigation systems (FAO, 2009) makes households even more sensitive to weather variations over time. Basalirwa *et al.* (1999) delineated 15 homogenous groups in Tanzania based on climatic conditions and topographic features. With this high number of agro-ecological areas, the magnitude and type of climate shocks are also expected to vary across geographical areas.

To investigate the impact of climate shocks, I complement the dataset with monthly data from the Standardized Precipitation Evapotranspiration Index (SPEI), gridded by longitude and latitude lines with a degree of precision of 0.5.<sup>17</sup> In the literature, most authors focus on the Standardized Precipitation index (SPI) (Rosenzweig and Udry, 2014; Jensen, 2000). This indicator, based only on precipitation data, assumes that droughts are particularly sensitive to temporal precipitation variations and that other climate variables are stationary. As a consequence, the SPI neglects the effects of global warming on production even though temperature has severe consequences on the drought intensity. As Vicente-Serrano *et al.* (2012) underlined, it implies that rainfall data are not necessarily suitable to predict crop yield. Indeed, the growing cycle of a plant does not depend only on the rainfall quantity but, most importantly, on the evapotranspiration of water.<sup>18</sup> This evapotranspiration varies with the temperature and explains why the same quantity of rainfall can have a different impact on the severity of droughts.

#### Crop seasons and climatic shocks

---

<sup>17</sup>These data have been developed by Vicente-Serrano *et al.* (2010).

<sup>18</sup>The evapotranspiration occurs through two mechanisms : the evaporation of water from the soil and the transpiration of crops.

I compute climate indicators for the time period that matters the most for the plants' growing cycle (Harari and La Ferrara, 2014). According to Kubik and Maurel (2016), weather conditions from March to May constitute the most relevant period at explaining Tanzanian crop production.<sup>19</sup> An alternative is to use the average value of SPEI from January to June.<sup>20</sup> To capture different types of droughts, I construct several SPEI variables from March to May: 6-month  $SPEI_{6mm}$  and 12-month  $SPEI_{12mm}$  (see appendix 3.7 for more details). As a robustness check, I also exploit the traditional rainfall data from NOAA.<sup>21</sup> To examine the non-linearity of climatic conditions, I also define positive and negative climate shocks. I consider that there is a drought when the SPEI is lower than 0.5 standard deviations, and that there is a positive rainfall shock when the SPEI is larger than 0.5 standard deviations.<sup>22</sup> This construction implies that positive values of the SPEI stand for better productivity conditions.<sup>23</sup> This hypothesis is empirically confirmed (see Table 3.7).

### 3.3.3 Price data

Incomes from agricultural activities should also be responsive to international price variations. From the Arusha Declaration in 1967 to 1980, prices were centrally controlled by the government (Msambichaka *et al.*, 1983). But from 1980 onwards, the market was liberalized and deregulated (Msambichaka L and O, 2006). The objective of this policy was to ensure a competitive, efficient and equitable market. Today, this food market deregulation implies that international price volatilities influence prices at which farmers sell their commodities on the local market. Notwithstanding, most crops are

---

<sup>19</sup>This period, called “Masika”, corresponds to the long rainy season in bimodal areas and to the rainy months in unimodal areas.

<sup>20</sup>Results are very similar and are available upon request.

<sup>21</sup>[http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.FEWS/.Africa/.DAILY/.ARC2/.daily/.est\\_prdp/datafiles.html](http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCEP/.CPC/.FEWS/.Africa/.DAILY/.ARC2/.daily/.est_prdp/datafiles.html)

<sup>22</sup>By taking these thresholds, about 20 % of the LSMS and Uwezo household samples are affected by droughts and about 20 % are affected by positive rainfall shocks.

<sup>23</sup>Since the SPEI is standardized with respect to local historical trends, positive values do not mean that there is an excess of water but only that rainfalls are larger than the historical trends.

still exclusively produced for self-consumption.<sup>24</sup> Consequently, the transmission channel between international prices and producers should exist only for producers of cash-crop commodities. Thus, I consider only price variations of the main cash-crops produced in Tanzania (cotton, coffee, coconut, tobacco, tea, sugar and palm-oil). Since Tanzania holds a small share of the market for these crops, international prices should be exogenous and independent of the Tanzania's production.

To measure volatility in international prices, I exploit the data from the World Bank Commodities Price Data and I use annual prices expressed in 2010 US \$ per kg.<sup>25</sup> Graph A3.8a pictures standardized price variations from 1980 to 2014. Prices of cash-crops have parallel trends and follow a U-shaped trend. To remove long-run changes in prices, I adopt the Hodrick-Prescott (HP) filter. This filter allows one to separate the price  $p_{c,y}$  of commodity  $c$  at time  $y$  into two components, the trend component  $T_{c,t}$ , and the cyclical component. Graph A3.8b depicts the cyclical components of cash-crop commodities. To obtain an aggregate price index, the FAO computes a weighted value of food prices. In order to come up with a price index which is representative of the Tanzanian market, I adopt the same strategy and construct a price index  $P_{j,y}$  based on the main Tanzanian cash-crop commodities. Since geographical areas do not produce the same commodities and are not similarly affected by price variations (Dube and Vargas, 2013; Imbert *et al.*, 2016), I weight price variations by the hectares of land allocated to cash-crop  $c$  in location  $j$  in 2000  $S_{c,j,2000}$ :

$$P_{j,y} = \sum_{c=1}^n \frac{(p_{c,y} - T_{c,y})}{T_{c,y}} * S_{c,j,2000}$$

$\frac{(p_{c,t} - T_{c,t})}{T_{c,t}}$  is the deviation from the trend in percentage. The agricultural intensity,  $S_{c,j,2000}$ , is computed from the geo-coded EarthStat data that combine satellite land cover data and agricultural census. These data provide the size of lands allocated to each crop with a 10km by 10 km resolution.<sup>26</sup> As a result, the price index varies over time and across

<sup>24</sup>For instance, 95 % of Sorghum and Millet are consumed by producers and trade is often limited to local exchanges (Rohrbach and Kiriwaggulu, 2007).

<sup>25</sup>The FAO provides the real prices of food and the IMF has indices of nominal prices, but they only go back to 1990.

<sup>26</sup>2000 is the most recent year for which these data are available with this level of precision.

locations.<sup>27</sup> To test the non-linearity in prices, I define negative and positive price shocks by referring to the first and the last quintiles, respectively.

## 3.4 Estimation strategy and results

In this section, I present the estimation strategies that identify the impact of productivity shocks on education outcomes. I deduce from these reduced forms whether the substitution effect or the income effect empirically prevails.

### 3.4.1 Effect of current productivity shocks

The effect of current shocks on education outcomes can be estimated with the following specification:

$$E_{ijty} = \beta_0 + \beta_1 P_{j,y-1} + \beta_2 SPEI_{j,y-1} + \gamma X_{ijy} + \delta_j + \mu_t + \nu_y + \epsilon_{ijty} \quad (3.22)$$

where  $i$  denotes the child of age  $t$  living in location  $j$  during the survey year  $y$ . Since households geo-coordinates are available only in the LSMS data,  $j$  designates the geographical units (50km\* 50km) where the household lives in the LSMS estimations, while  $j$  designates the district of residence in the Uwezo estimations. The parameters  $\delta$ ,  $\mu$  and  $\nu$  are location, age and year fixed-effects, respectively. The error term  $\epsilon_{ijty}$  is clustered by location  $j$ , and  $X_{ijy}$  is a set of household controls such as the number of adults and children in the household, the number of boys among siblings, and age and education of the household head.  $E_{ijty}$  is a large set of education outcomes that measures education decisions and educational achievement. I regress current education outcomes  $E_{ijty}$  on the lagged climate variable  $SPEI_{j,y-1}$  and on the lagged aggregated price index  $P_{j,y-1}$ .

By adding region and year fixed effects, this estimation strategy compares children from the same location in different rounds of the survey. It captures the causal effect

---

<sup>27</sup>This index can be constructed at the household level. However, the area allocated to each crop may be endogenous at the household level, while at a larger scale (community or district level)  $S_{c,j,2000}$  is representative of the geographical conditions suitable for different crop's cultivations.



of productivity shocks on education outcomes if several assumptions are satisfied. First,  $SPEI_{j,y-1}$  and  $P_{j,y-1}$  should change the labor productivity (see sub-section 3.4.4). Second, the shocks should be purely exogenous (see sub-section 3.3.2 and 3.3.3 for further discussion) and finally, they should not be correlated with unobserved variables that would explain education outcomes. This question will be addressed later in section 3.5. To estimate the effect of productivity variations on children’s education and activities (whether the child works, whether the child is enrolled in school, whether the child has dropped out of school and what is the highest grade achieved), I use the LSMS data and I restrict the sample to school aged children. In Table A3.17, I observe that continuous price and climate variables have no significant impact on either education or work decisions.

TABLE 3.1: EFFECT OF POSITIVE SHOCKS ON CHILDREN’S ACTIVITIES

	Work (1)	Enrolled (2)	Dropout (3)	Grade (4)
Positive Price Shock $_{t-1}$	0.058* (0.033)	-0.035** (0.017)	0.004 (0.011)	-0.063 (0.082)
Positive Rainfall Shock $_{t-1}$	0.084** (0.033)	0.001 (0.014)	0.014* (0.008)	-0.124*** (0.045)
Negative Price Shock $_{t-1}$	-0.013 (0.025)	-0.004 (0.014)	-0.006 (0.009)	0.006 (0.074)
Negative Rainfall Shock $_{t-1}$	0.006 (0.028)	0.009 (0.017)	-0.004 (0.008)	-0.034 (0.045)
R-squared	0.167	0.154	0.084	0.297
Observations	12,677	11,625	11,230	10,588
Localities F.E	×	×	×	×
Year F.E	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Note: Standard errors, clustered by geographical units ( $0.5^{\circ} \times 0.5^{\circ}$  of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the number of boys among siblings, the age and the education of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

Moving to the non-linearity of productivity shocks, I observe that positive rainfall shocks increase the probability of working, increase the probability of dropping out of school, and lowers the grade achievement by 0.11 years (see Table 3.1). Although all coefficients are not significant, these results go in the same direction and suggest

that contemporaneous productivity shocks have a counter-cyclical impact on education decisions ( $\frac{\partial E(w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta))}{\partial w_2} < 0$ ). In contrast, negative productivity shocks have no significant impact on education decisions. According to the model's predictions, these results indicate that the substitution effect dominates the income effect and that children are encouraged to work more and to decrease their demand for education when they become more productive.

Based on the theoretical framework, these effects are also expected to vary with households' wealth. To explore this heterogeneity, I compute the household consumption following the guideline of [Deaton and Zaidi \(2002\)](#).<sup>28</sup> Table A3.9 shows that results are very similar for rich and poor households. Most coefficients are not significantly different, and suggest that positive productivity shocks are detrimental to education decisions at all wealth levels.

I can also examine whether the effects are heterogeneous across children's age and across gender. Table A3.10 shows that the effects of positive shocks on labor and education decisions are very close for boys and girls and are not statistically different. To consider heterogeneity by age, I define the 7-13 years old group as children of primary education age and the 14-16 years old group as children of lower secondary age. As pointed out in section 3.2, it is unclear whether the effects of productivity shocks will be more pronounced for the younger cohort or the older cohort. I see from Table A3.11, that older children are more likely to work and less likely to pursue their education, meaning that the counter-cyclical relationship between productivity shocks and education decisions is strengthened when children get older. This result is consistent since the substitution effect should be larger when children become more productive.

Thus, the LSMS results suggest that positive productivity shocks increase child labor and are unfavorable to education achievement. If these shocks provoke erratic attendance, they should also decrease children's cognitive skills. To test this hypothesis, I use the

---

<sup>28</sup>This consumption variable is composed of four sub-aggregates, food items, non-food items, housing consumption and consumer durables. In order to create a consumption variable independent from current shocks, I exclude all current consumption items such as food consumption and current non-food items that could have been affected by productivity shocks.

Uwezo data and I regress test scores on price and climate shocks.

TABLE 3.2: EFFECT OF CONTEMPORANEOUS SHOCKS ON TEST SCORES

	Swahili (1)	Maths (2)	Swahili (3)	Maths (4)
Positive Price Shock $_{t-1}$	-0.007 (0.016)	-0.012 (0.018)	-0.015 (0.016)	-0.020 (0.020)
Positive Rainfall Shock $_{t-1}$	-0.029* (0.017)	-0.036* (0.020)	-0.023 (0.018)	-0.032 (0.022)
Negative Price Shock $_{t-1}$	-0.022 (0.028)	0.019 (0.026)	-0.020 (0.024)	0.010 (0.021)
Drought $_{t-1}$	-0.001 (0.013)	0.003 (0.016)	0.006 (0.013)	0.010 (0.015)
R-squared	0.321	0.293	0.321	0.287
Observations	328,948	328,948	286,250	286,250
District F.E	×	×	×	×
Year F.E	×	×	×	×
Attend school			×	×

Sources: Uwezo data from 2011 to 2014. Note: Standard errors are clustered at the district level and are reported in parentheses. Controls are years dummies, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the age and the education of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

Table 3.2 presents the results and shows that only positive rainfall shocks decrease Swahili and maths scores, and they do so by 0.03 standard deviations. These effects are no longer significant when I restrict the sample to enrolled children, probably because children who stay at school during positive shocks are positively selected.

In conclusion, households take advantage of a labor productivity increase by calling on child labor. This decision interacts with education enrollment and achievement but has very little effect on schooling performance.

### 3.4.2 Effect of the length and frequency of school-aged

To study whether positive productivity shocks are more detrimental to education when they become more frequent and when their length extends, I compute for each child the length of shocks (the maximum number of consecutive shocks) and the total number of

positive productivity shocks from the beginning of primary education (at 7 years old) to the year of the survey:

$$E_{ijty} = \beta_0 + \beta_1 \sum_{i=7}^y R_{j,i} + \beta_2 \sum_{i=7}^y PP_{j,y} + \gamma X_{ijy} + \delta_j + \mu_t + \nu_y + \epsilon_{ijty} \quad (3.23)$$

$\sum_{i=7}^y R_{j,i}$  and  $\sum_{i=7}^y PP_{j,y}$  depict the number (or length) of positive rainfall shocks and the number (or length) of positive price shocks from age 7 to the year of the survey. The specification is similar than equation (3.22) except that I am interested in the repetition of shocks. Since negative current shocks have no impact of education outcomes, I focus the following analysis on positive shocks.<sup>29</sup> Then, I consider education outcomes that explain current decisions as well as prior decisions (whether the child is overage, whether the child can read and write and the grade achievement).

TABLE 3.3: EFFECT OF SHOCKS DURING SCHOOLING ON EDUCATION DECISIONS.

	Work (1)	Overage (2)	Grade (3)	Read and write (4)
Number Positive Price Shocks	0.015 (0.016)	0.016* (0.009)	-0.037 (0.037)	-0.025** (0.012)
Number Positive Rainfall Shocks	0.045*** (0.012)	0.019 (0.012)	-0.057 (0.042)	0.014 (0.013)
R-squared	0.166	0.247	0.694	0.230
Length Positive Price Shocks	0.023 (0.019)	0.019* (0.011)	-0.059 (0.045)	-0.019* (0.011)
Length Positive Rainfall Shocks	0.056*** (0.015)	0.000 (0.015)	-0.002 (0.049)	-0.020 (0.017)
R-squared	0.166	0.247	0.694	0.230
Observations	10,322	8,717	8,717	6,748
Localities F.E	×	×	×	×
Year F.E	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Note: Standard errors, clustered by geographical units (0.5\*0.5° of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the number of boys in the sibling, the age and the education of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

Table 3.3 reports the main results and suggests that an increase of the number or

<sup>29</sup>The effect of cumulative negative shocks has been tested and no significant effect has been found. Results are available upon request.

length of positive rainfall shocks raises the probability of working by 5%. In addition, the length and the number of positive price shocks increase the probability of being overage and reduce the probability of being able to read and write. Taken together, these estimates support the hypothesis that positive productivity shocks are prejudicial to education achievement.

To go further in the analysis, I look at the effect of shocks by household consumption and I notice that as current shocks, the frequency of shocks affects child labor and education decisions of both poor and rich households. However, the grade progression is slowed down for poor households only (see Table A3.12). Moving to the heterogeneity across age, I compare the effect of the number of shocks that occurred in primary and in secondary education.<sup>30</sup> Results presented in Table A3.13 show that positive productivity shocks increase the probability to work and reduce educational achievement of both age cohorts.<sup>31</sup> In contrast, the comparison of the effect by gender suggests that positive productivity shocks are detrimental for education of boys only (see Table A3.14). Although households may have preferences for boys' education rather than girls' education, boys are more likely to drop out of school if they are more productive (Andre *et al.*, 2017).<sup>32</sup>

To better understand the effect of shocks on cognitive skills, I turn to the effect on test scores. Results presented in Table 3.4 give evidence that positive climatic and price shocks during school age reduce cognitive skills in maths and Swahili. The total number of shocks and the maximum number of consecutive shocks have the same effects. This test scores decline does not seem to be driven by a change in enrollment because the effects remain significant for children enrolled in school (columns 3 to 4). Thus, the results advocate that educational achievement is probably reduced due to lower time investment in school

---

<sup>30</sup>To construct the number of shocks during the two schooling cycles, I sum the number of shocks from the beginning of each cycle (7 year old for primary education and 14 year old for secondary education).

<sup>31</sup>One additional positive productivity shock decreases the probability that young children read and write and decreases the grade achievement of the older cohort. It is worth underlying that most children aged 14 to 16 can read and write which may explain why the effect on the probability of being able to read becomes insignificant for the older cohort.

<sup>32</sup>If they are more productive, the substitution effect will be larger.

or in doing homework.<sup>33</sup> This emphasizes the need for data that gather test scores and detailed information on children’s time allocation.

TABLE 3.4: EFFECT OF SHOCKS DURING SCHOOLING ON TEST SCORES

	Swahili (1)	Maths (2)	Swahili (3)	Maths (4)
Number Positive Price Shocks	-0.026*** (0.007)	-0.032*** (0.007)	-0.027*** (0.007)	-0.030*** (0.007)
Number Positive Rainfall Shocks	-0.027*** (0.010)	-0.019 (0.012)	-0.031*** (0.011)	-0.019 (0.012)
R-squared	0.323	0.295	0.322	0.289
Lenght Pos. Price Shocks	-0.024*** (0.007)	-0.028*** (0.007)	-0.024*** (0.007)	-0.025*** (0.007)
Lenght Pos. Rainfall Shocks	-0.035*** (0.007)	-0.032*** (0.009)	-0.040*** (0.007)	-0.034*** (0.009)
R-squared	0.323	0.295	0.322	0.289
Observations	328,948	328,948	294,521	294,521
District F.E	×	×	×	×
Year F.E	×	×	×	×
Attend School			×	×

Sources: Uwezo data from 2011 to 2014. Note: Standard errors are clustered at the district level and are reported in parentheses. Controls are years dummies, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the age and the education of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

Whether I consider heterogeneity across age groups or gender, I find that positive shocks decrease test scores of each group and that the coefficients are not statistically different between groups. To insure that differences between the younger and the older cohorts are not explained by the construction of the test scores,<sup>34</sup> I standardize the tests with respect to age. Table A3.21 discredits this hypothesis by showing that, even with age-standardization, productivity shocks affect the test scores of both age groups.

Three main conclusions can be drawn from this sub-section. First, the number of schocks which occur during schooling clearly increase child labor. Although they affect education decisions in a small extent, they have a substantial adverse effect on schooling performance ( $\frac{\partial A(w_1, w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta))}{\partial w_2} < 0$ ).

<sup>33</sup>This last channel cannot be tested since the time spent at school is not available in the datasets.

<sup>34</sup>Test scores evaluate skills of the Standard 2 level, the second year of school.

### 3.4.3 Effect of early shocks

In order to shed light on early life shocks' consequences, I estimate the following equation:

$$E_{ijty} = \beta_0 + \beta_1 SPEI_{j,utero} + \dots + \beta_7 SPEI_{j,age6} + \beta_8 P_{j,utero} + \dots + \beta_{14} P_{j,age6} + \gamma X_{ijt} + \delta_j + \mu_t + \nu_y + \epsilon_{ijty} \quad (3.24)$$

I look at the effect of price and climate variations from in utero until 6 years old to identify the critical period in early life. In this purpose, I check whether early life shocks have long-lasting consequences on education of children that are currently of school age.<sup>35</sup>

As the conceptual framework asserts, shocks in early childhood affect education through two possible channels. The first channel is the nutrition-learning nexus, which suggests that positive productivity shocks are likely to improve children's nutrition in early life, stimulate children's growth, and have long-lasting impact on children's cognitive skills. Second, a better labor productivity allows parents to transfer larger savings in absolute value, which can be used to protect education against current income shocks.

To investigate the channel between productivity shocks and children's health, I construct a z-score of height for age based on the 2006 WHO child growth standards (Leroy, 2011). This index measures the prevalence of stunting among children from 0 to 5. Table A3.8 shows that, consistently with the model's expectations, children appear in better health when the labor productivity is improved: at birth, an increase of the climate variable by one standard deviation raises the Z-score of height for age by 0.4 point. Similarly, at one year old, an increase of the climate and the price variable by one standard deviation raises the Z-score by 0.2 and 0.05, respectively. When children are older than 2 year old, these effects become insignificant.

Turning to the effect of early-life shocks on education decisions, it is worthwhile highlighting that early productivity shocks change the attractiveness of location and can generate selective migration. If the household characteristics that drive migration also

---

<sup>35</sup>Since test scores of children aged 0 to 6 are not available, short-run effects of early life shocks cannot be estimated. This being so, long-run effects constitute a lower bound of short-run effects if early-life effects fade over time.

influence education decisions, estimations will be biased. To test whether self-migration significantly biases the results, I compare results of equation (3.24) by including and excluding migrant households.<sup>36</sup> Table 3.5 presents the main results and shows that labor productivity variation in early ages have little effect on current education status (have ever been enrolled, being overage, grade achievement). Yet, children who experienced larger rainfalls in utero to age one benefit from a higher education progression,<sup>37</sup> and prices at 6 years old raises the grade achievement by 0.5 year.

---

<sup>36</sup>Migration information are available in the LSMS data only.

<sup>37</sup>Larger rainfalls in utero have a positive impact on the grade achievement, while they decrease the probability of being overage. In the same vein, larger rainfalls at birth increase the probability of being ever enrolled in school.



TABLE 3.5: EFFECT OF EARLY LIFE SHOCKS ON CHILDREN'S ACTIVITIES(BETA COEFFICIENTS)

	Ever edu		Grade		Overage	
	(1)	(2)	(3)	(4)	(5)	(6)
SPEI-6 March-May <sub>age vitro</sub>	-0.000 (0.004)	-0.003 (0.005)	0.044* (0.023)	0.063 (0.029)	-0.013* (0.007)	-0.020** (0.008)
SPEI-6 March-May <sub>age birth</sub>	0.009** (0.004)	0.007 (0.005)	0.027 (0.024)	0.020 (0.029)	-0.007 (0.007)	-0.006 (0.008)
SPEI-6 March-May <sub>age 1</sub>	0.006 (0.006)	0.008 (0.007)	0.050* (0.026)	0.062* (0.037)	-0.008 (0.007)	-0.014 (0.010)
SPEI-6 March-May <sub>age 2</sub>	0.003 (0.006)	0.005 (0.008)	0.005 (0.033)	0.001 (0.042)	-0.002 (0.008)	0.000 (0.010)
SPEI-6 March-May <sub>age 3</sub>	-0.004 (0.006)	0.000 (0.007)	0.020 (0.036)	0.027 (0.049)	-0.012 (0.009)	-0.015 (0.012)
SPEI-6 March-May <sub>age 4</sub>	-0.005 (0.006)	-0.004 (0.007)	0.018 (0.040)	-0.003 (0.040)	0.012 (0.010)	0.017 (0.011)
SPEI-6 March-May <sub>age 5</sub>	0.001 (0.006)	0.003 (0.006)	-0.020 (0.041)	0.000 (0.047)	-0.010 (0.010)	-0.017 (0.011)
SPEI-6 March-May <sub>age 6</sub>	-0.002 (0.005)	-0.004 (0.005)	-0.032 (0.029)	-0.024 (0.034)	0.004 (0.010)	-0.002 (0.011)
$P_{j,age vitro}$	0.012 (0.026)	-0.005 (0.027)	-0.079 (0.193)	-0.144 (0.216)	0.047 (0.051)	0.079 (0.060)
$P_{j,age birth}$	-0.016 (0.024)	-0.003 (0.026)	0.203 (0.177)	0.248 (0.201)	-0.057 (0.048)	-0.085 (0.057)
$P_{j,age 1}$	0.027 (0.030)	0.006 (0.032)	-0.266 (0.208)	-0.319 (0.237)	0.072 (0.060)	0.104 (0.072)
$P_{j,age 2}$	-0.036 (0.045)	-0.001 (0.049)	0.333 (0.313)	0.429 (0.360)	-0.091 (0.091)	-0.133 (0.106)
$P_{j,age 3}$	0.062 (0.082)	0.007 (0.087)	-0.482 (0.566)	-0.642 (0.641)	0.121 (0.159)	0.207 (0.184)
$P_{j,age 4}$	-0.084 (0.092)	-0.027 (0.097)	0.603 (0.646)	0.738 (0.728)	-0.095 (0.178)	-0.186 (0.204)
$P_{j,age 5}$	0.072 (0.082)	0.029 (0.083)	-0.720 (0.580)	-0.824 (0.649)	0.125 (0.156)	0.210 (0.178)
$P_{j,age 6}$	-0.026 (0.037)	-0.018 (0.037)	0.473* (0.272)	0.485 (0.297)	-0.055 (0.070)	-0.089 (0.077)
R-squared	0.063	0.074	0.610	0.604	0.198	0.206
Observations	9,697	7,756	8,267	6,612	8,267	6,612
District F.E	×	×	×	×	×	×
Year F.E	×	×	×	×	×	×
With migrant HH	×		×		×	

Sources: LSMS-ISA from 2008, 2010 and 2012. Note: Standard errors, clustered by geographical units (0.5\*0.5° of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the number of boys in the sibling, the age and the education of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

I argue that the first effects are explained by nutrition-education nexus, while the effect at 6 year old probably involves other mechanisms. Indeed, positive shocks at 6 year old may either encourage parents to send their children to school earlier, or allow parents to save more money in order to protect children's education against future income shocks. In addition, I observe that excluding migrant households do not plague the results, which implies that the selection migration bias is negligible.

Consistently with these last results, Table 3.6 shows that early productivity shocks from birth to four years old are pro-cyclical and have positive significant impacts on Swahili and math scores ( $\frac{\partial A(w_1, w_2, \gamma, \Delta, I_2(w_1, w_2, \gamma, \Delta))}{\partial w_1} > 0$ ). An improvement in climate conditions by one standard deviation increases test scores between 1.4% and 2.6%. In contrast, prices have very little effect on test scores except at 4 and 6 years old. Surprisingly, I also find that an increase in prices in utero has a significant negative effect.<sup>38</sup>

In conclusion, these results are consistent with the literature (Almond and Currie, 2011; Currie and Vogl, 2013) which finds that positive early life shocks have long-lasting consequences on schooling performance. Following in Shah and Steinberg (2017)'s footsteps, I show that productivity shocks matter for cognitive skills, especially when they occur before 4 year old.

---

<sup>38</sup>These results have no theoretical ground, but might be driven by a lack of accuracy on birth dates.

TABLE 3.6: EFFECT OF EARLY LIFE SHOCKS ON SCHOOLING OUTCOMES (BETA COEFFICIENTS)

	Swahili (1)	Math (2)	Swahili (3)	Math (4)
SPEI-6 March-May <sub>age in utero</sub>	0.002 (0.005)	0.006 (0.005)	0.004 (0.004)	0.007 (0.005)
SPEI-6 March-May <sub>age birth</sub>	0.006 (0.007)	0.014* (0.007)	0.010 (0.006)	0.015** (0.007)
SPEI-6 March-May <sub>age 1</sub>	0.008 (0.009)	0.022** (0.009)	0.011 (0.008)	0.023*** (0.008)
SPEI-6 March-May <sub>age 2</sub>	0.014 (0.010)	0.024** (0.010)	0.018* (0.009)	0.026*** (0.009)
SPEI-6 March-May <sub>age 3</sub>	0.012 (0.009)	0.026*** (0.009)	0.013 (0.009)	0.026*** (0.009)
SPEI-6 March-May <sub>age 4</sub>	0.006 (0.010)	0.020* (0.010)	0.005 (0.010)	0.018* (0.010)
SPEI-6 March-May <sub>age 5</sub>	0.001 (0.009)	-0.000 (0.009)	0.002 (0.010)	0.003 (0.008)
SPEI-6 March-May <sub>age 6</sub>	0.005 (0.008)	-0.002 (0.007)	0.007 (0.008)	0.000 (0.007)
$P_{j,in utero}$	-0.021*** (0.006)	-0.015*** (0.006)	-0.016*** (0.005)	-0.011** (0.005)
$P_{j,birth}$	0.013 (0.011)	0.004 (0.008)	0.012 (0.011)	-0.001 (0.008)
$P_{j,age 1}$	0.000 (0.010)	0.006 (0.008)	-0.001 (0.010)	0.011 (0.008)
$P_{j,age 2}$	0.010 (0.018)	0.010 (0.016)	0.015 (0.019)	0.004 (0.017)
$P_{j,age 3}$	-0.004 (0.017)	-0.014 (0.016)	-0.004 (0.015)	-0.004 (0.016)
$P_{j,age 4}$	0.021 (0.015)	0.032* (0.017)	0.022 (0.016)	0.025 (0.021)
$P_{j,age 5}$	-0.013 (0.009)	-0.012 (0.008)	-0.011 (0.009)	-0.005 (0.011)
$P_{j,age 6}$	0.013* (0.008)	0.011 (0.009)	0.011 (0.008)	0.010 (0.011)
R-squared	0.274	0.247	0.282	0.251
Observations	279,855	279,855	252,471	252,471
District F.E	×	×	×	×
Year F.E	×	×	×	×
Attend school			×	×

Note: Standard errors are clustered at the ward level and are reported in parentheses. \*\*\*, \*\*, \* means respectively that the coefficient is significantly different from 0 at the level of 1%, 5% and 10%.

### 3.4.4 Mechanisms

To understand how households take their education decisions when they are exposed to climate and price shocks, I study the underlying channels at play with the following specification:

$$Y_{h jy} = \beta_0 + \beta_1 P_{j,y} + \beta_2 SPEI_{j,y} + \gamma X_{h jy} + \delta_j + \nu_y + \epsilon_{h jy} \quad (3.25)$$

Where the subscript h depicts the household, j the location, and y the year of the survey. To test whether climate and price shocks translate into productivity shocks, I examine the effect of shocks on two outcomes  $Y_{h jy}$ , household production and household labor decisions.

TABLE 3.7: EFFECTS OF PRODUCTIVITY SHOCKS ON THE LOG OF HOUSEHOLD PRODUCTION (BETA COEFFICIENTS).

	(1)	(2)	(3)
SPEI-6 months March-May	0.266*** (0.094)		0.327*** (0.099)
$P_{jy-1}$		0.114** (0.049)	0.154*** (0.049)
$P_{jy-1} * \text{cash} - \text{crop prod.}$		0.146 (0.434)	0.134 (0.432)
R-squared	0.174	0.174	0.177
Observations	12,180	12,183	12,180
Localities and Times F.E	×	×	×
Households F.E	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Note: Production and Consumption are computed in Tanzanian shillings (TZS). Standard errors, clustered by geographical units (0.5\*0.5° of precision), are reported in parentheses. Controls are survey month dummies, cultivated lands, the number of days of labor in the field and the age of the household head. \*\*\*,\*\*,\* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

Table 3.7 presents the beta coefficients and shows that both climate and price variables significantly raise the household production. When rainfalls increase by one standard deviation, household production is raised by 33%, and when prices increase by one

standard deviation, the value of the production is raised by 15 % for the non cash-crop producers and by 28 % for the cash-crop producers.<sup>39</sup> At first glance, it is surprising to see non cash-crop producers affected by cash-crop prices, but this result is probably explained by the high correlation between international prices. Indeed, figure A3.8b and A3.9 show that subsistence and cash-crop price deviations have parallel trends : they are negative from 2002 to 2011 and positive afterwhile. As a result, cash-crop price index may confound the effect of cash-crop prices, and the effect of other agricultural prices.

To analyze the effect of climate and price shocks on production, it is necessary to remind a substantial difference: climate shocks probably increase the quantity of harvested crops, while prices raise the value of the production without necessarily changing the quantity produced.<sup>40</sup>

TABLE 3.8: EFFECT OF CLIMATE AND AGGREGATE PRICE VARIABLES ON DAYS OF LABOR IN THE FIELD (BETA COEFFICIENTS).

	Hired labor	Adult labor	Child labor	Hired labor	Adult labor	Child labor
SPEI-6 months March-May	6.414*	27.95***	35.38***	6.299*	28.00***	35.09***
	(3.607)	(10.09)	(10.44)	(3.544)	(9.960)	(10.37)
$P_{jy-1}$	-0.551	2.388	7.503*			
	(0.843)	(5.185)	(4.204)			
$P_{jy-1} * cash - crop prod.$	2.835	1.391	33.65			
	(5.269)	(23.88)	(21.89)			
$P_{jy-1} S.R.$				-0.619	2.995	7.504*
				(0.812)	(5.349)	(4.266)
$P_{jy-1} S.R. * S.R. cash - crop prod.$				0.924	-2.303	33.62
				(5.085)	(24.38)	(22.21)
Within R-squared	0.091	0.099	0.100	0.090	0.098	0.100
Observations	3,365	7,394	5,256	3,365	7,394	5,256
Localities and Time F.E	×	×	×	×	×	×
Household F.E	×	×	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: Standard errors, clustered at the household level, are reported in parentheses. Controls are survey month dummies, cultivated lands, the number of adults and the number of children in the household and the age of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

To take opportunity of a labor productivity increase, parents may be encouraged to increase child labor. Table 3.8 presents the effect of price and climate variations on the yearly household labor supply. I find that, when rainfalls and cash-crop prices are raised

<sup>39</sup>Table A3.3 compares the effects of the traditional SPI with the SPEI-6 months and SPEI-12 months. I observe that all climatic variables have a positive impact on production but that the effect of the SPEI-6 months is larger. This is why, SPEI-6 months is the reference in this analysis.

<sup>40</sup>The quantity of harvested cash-crops increases only if households decide to allocate more labor.

by one standard deviation, children perform 35 days and 7.5 days more, respectively. Several reasons can explain why the price effect is five times lower than the climate effect. First, Table 3.7 shows that prices have lower effect on production than rainfalls. Second, households probably take more time to perceive and react to price variations than climate variations. Finally, households who produce cash-crops with longer cycles (coffee, coconut, tea, palm oil and sugar) are not able to change the size of cultivated lands in the short-run and have little room to adapt their labor allocation.<sup>41</sup> When I look at the effect of price variations on subsistence crops (see Table A3.4), I find no effect on labor, which is consistent. Indeed, when subsistence crops become more expensive, the value of the households production increase but this does not translate in an opportunity loss since products are self-consumed. To go one step further, I compare the effect of productivity shocks by children's age (see Table A3.5), but I find no significant difference. I also exploit the LSMS data on the weekly allocation of labor and I find that consistently, rainfalls and prices increase the time spent in agricultural activities (see Table A3.7).

Thus, these results lead to think that rainfall and cash-crop prices constitute productivity shocks.

---

<sup>41</sup>In the LSMS data, this problem should be limited because only 0.26% of agricultural households produce these goods in the sample. This explains why I find that the effect of cash crops and the effect of short-run cash-crops are not statistically different.

## 3.5 Discussion

In this analysis, climate and price shocks affect education decisions and schooling performance by changing households resources and the opportunity cost of children's time. However, the causal impact is identified only if the shocks do not influence education through other channels. In this section, I discuss potential sources of bias.

First, shocks should not influence the perceived returns to education which enter the education decisions. This kind of general equilibrium effect may happen when the shocks persist over time. Since this analysis focuses on shocks that persist only for relatively short periods of time, it is very unlikely that the shock of interest, current and school-age shocks, change the returns to education.

Another concern that has been discussed in section 3.4 is selective migration. If households who move towards prosperous locations have on average higher skills, children in districts which experienced positive productivity shocks are positively selected. In other words, when estimating the effect of positive productivity shock, the migration selection may induce an upward bias, which misleads the interpretations. To investigate whether it is an empirical issue, I regress the probability that a child aged 7 to 16 migrates in another district on lagged productivity shocks.<sup>42</sup> Results presented in Table A3.22 suggest that this selection bias is negligible and that productivity conditions do not significantly drive children's migration.

As Shah and Steinberg (2017) emphasize, mortality in early childhood also represents a potential source of bias. Indeed, exposure to negative productivity shocks such as drought may increase mortality in early life and change the composition of sampled children. Surviving children, who are more resistant, are likely to be positively selected and to better perform at school. Consequently, the results confound the direct effect of shocks on education and the effect from selection mortality. To address this concern, I use the LSMS data that provide the number of individuals, including infants, who died over the past two years. Then, I test whether mortality in early childhood (between 0 and 3 years

---

<sup>42</sup>Internal migration concerns 6 % of children from this age group.

old) depends on productivity shocks over the two last years. Table A3.22 reports no significant effect suggesting that mortality does not bias the results.

Last, but not least, results are biased if productivity shocks affect the quality of education. Heavy rains for instance, may make roads impassable, damage access to school and increase teachers' and children's absenteeism. By changing the attractiveness of agricultural activities, productivity shocks are expected to select teachers, but is unclear whether positive productivity shocks especially attract skilled or unskilled teachers. Although climate and price levels have no significant effect on the percentage of absent teachers, an increase of the price index by one standard deviation decreases the percentage of qualified teachers by 1 % (see Table A3.23). However, this effect is not robust when I consider the non-linearity in climate and prices.

Finally, to check whether the construction of price and climate variables is efficient at capturing productivity shocks, I test whether future shocks affect education decisions. Although future continuous price and climate variables have no impact on education and work decisions (see Table A3.19), three out of twelve coefficients become significant when I examine the non-linearity in prices (see Table A3.20). For instance, negative and positive price shocks significantly affect the enrollment decision. This anachronic result has no theoretical ground, but can be explained by the nature of price shocks. As Figure A3.9 and A3.8b depict, price variations last longer. Thus, adding prices of different periods may lead to collinearity issues<sup>43</sup>

## 3.6 Conclusion

In this paper, I study the effect of labor productivity on the demand for education and on schooling performance in Tanzania. To capture exogenous variations in labor productivity, I use a combination of geo-coded data to identify variations in climate conditions and in cash-crop prices over time and across location.

---

<sup>43</sup>The correlation between  $P_{j,y-1}$  and  $P_{jy+1}$  is 0,92 while the correlation between  $R_{jy-1}$  and  $R_{jy+1}$  is 0,05.



The core of this analysis is to investigate the effect of productivity shocks on children's education by considering two particular aspects, the age at which shocks occur, and the length of shocks.

The first findings support the idea that early life productivity shocks (from birth to 4 years old) are favorable to the development of future cognitive skills. Based on the theoretical model, this relationship is explained by two channels. Higher labor productivity in early-life improves children's nutrition and allows parents to save money to finance education later in life. Since labor productivity has negligible effect on education enrollment, I argue that this relationship between early-life shocks and schooling performance is mainly explained by the nutrition-learning nexus assumption<sup>44</sup> (Almond and Currie, 2011; Currie and Vogl, 2013; De Vreyer *et al.*, 2014).

In contrast, when children are of school age, the relationship between positive productivity shocks and education becomes counter-cyclical. This result, close to Shah and Steinberg (2017) findings, suggests that the substitution effect outweighs the income effect. When children are of school age and can work, positive productivity shocks increase the available income for education, but also increase the labor productivity which encourage households to call on child labor.

Interestingly, households adopt different coping mechanisms depending on the length of shocks. In response to current shocks, households increase child labor and put their children out of school. Schooling performance which results from a cumulative process, does not decrease yet. When shocks become recurrent, the conclusion is slightly different. Households still increase child labor, but do not change their education decisions. Despite this, schooling performance significantly drops. In other words, productivity shocks are detrimental to schooling performance even when children stay enrolled in school. This emphasizes that limiting the analysis to education enrollment is not satisfactory. To test whether the results are due to erratic attendance, detailed data on children's schedule are needed.

---

<sup>44</sup>Children who were exposed to positive productivity shocks in early life benefit from a better nutrition, which is necessary for the cognitive development.

In terms of public policies, these results imply that it is necessary to alleviate tuition fees, but also to account for the opportunity costs of children's time. In this regard, it would be interesting to test whether access to labor market allows households to cope with positive shocks. This research question, which requires rich data on the labor market, is left for future research.

## Appendix 3

### 3.7 Construction of the Standard Precipitation Evapotranspiration Index (SPEI)

To account for several climatic parameters, the SPEI provides a simple drought measure defined by the difference  $D$  between the original SPI and the Potential Evapotranspiration (PET). The PET corresponds to the evapotranspiration that would occur if the surface was sufficiently watered to be green and to have an active growth. Naturally, this PET varies between locations and depends on climate conditions and on the nature of the soil. This index is not observed and has to be modelled. The most wide-known computation used in the SPEI data is the Penman-Monthei equation.<sup>45</sup> Therefore,  $D$  represents the monthly water surplus or water deficit.

Similarly to the SPI, the SPEI accounts for different time scales that determine the nature of droughts. Short time scales represent soil water content and discharge in headwaters, while medium time scales refer to storage of water sources and long-time scales illustrate variations in groundwater. The various time scales are computed difference  $D$  by aggregating various time periods. For instance, the 6-month SPEI index is measured by adding the  $D$  values of the last 5 months before the current month.<sup>46</sup>

Then, to obtain comparable SPEI values in time and in space, the SPEI index is standardized using the Log-Logistic distribution. By construction, the historical mean is 0 for each geographical cell and the SPEI index is expressed in units of standard deviation

---

<sup>45</sup>According to the FAO manual on crop evapotranspiration (?), the FAO-56 Penman-Monteith equation estimates a reference evapotranspiration  $Et_0$  and is the most efficient method to approximate the PET:

$$Et_0 = \frac{0.408(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)}$$

Where  $R_n$  is the net radiation of the crop surface,  $G$ , the soil heat flux density,  $T$  the mean daily air temperature at 2 m height,  $u_2$  is the wind speed at 2 m height,  $e_s$  is the saturation vapour pressure,  $e_a$  is the actual vapour pressure,  $\Delta$  is the slope vapour pressure curve and  $\gamma$  is the psychrometric constant.

<sup>46</sup>To give a decreasing weight of the data from the past, Vicente-Serrano *et al.* (2010) adopt a Gaussian kernel function.

from the historical average.

### 3.8 Descriptive statistics

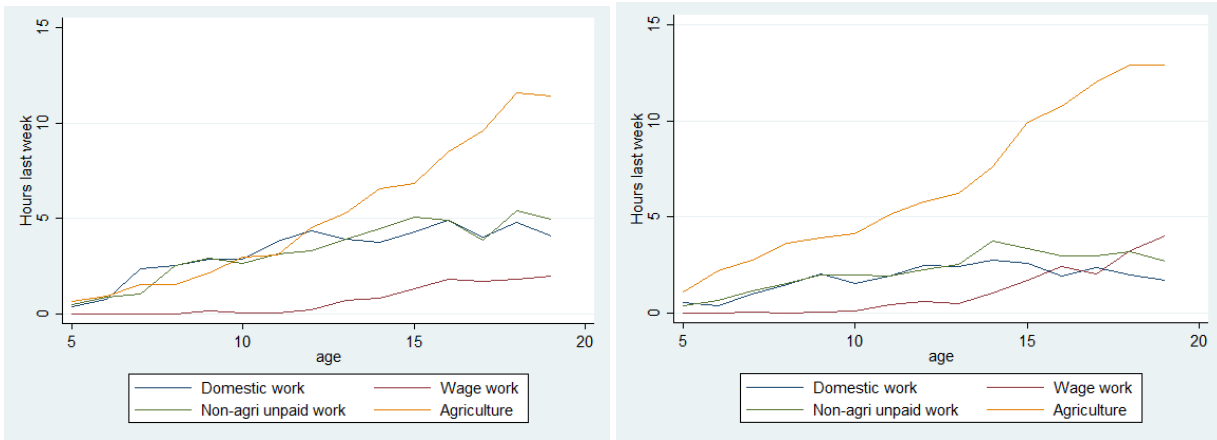
TABLE A3.1: DESCRIPTIVE STATISTICS FROM THE LSMS-ISA DATA

Year	2008	2010	2012
Composition of samples			
Number of districts	126	128	131
Number of wards	87	103	108
Number of HH	3265	3921	5004
Number of children	4512	5239	6236
Household characteristics			
Number of adults	3.208	3.425	3.511
Number of children	3.315	3.321	3.305
Household production (TSH)	246957.4	372482.4	708005.2
cultivated area (acres)	4.822	4.548	5.759
Children characteristics			
Child's age	11.407	11.370	11.335
Child is female	0.507	0.503	0.507
Child is enrolled in school	0.815	0.883	0.858
Number of completed grade	4.295	4.388	4.324
Child dropout school this year	0.043	0.050	0.061
Child has repeated a grade this year	0.123	0.117	0.129
Child reads and writes	.	0.733	0.720
Child works last week	0.114	0.422	0.262
Child works last year	0.116	0.444	0.382

TABLE A3.2: DESCRIPTIVE STATISTICS FROM THE TANZANIAN UWEZO SURVEY.

Year	2010	2011	2012	2013	2014
Composition of samples					
Number of districts	42	131	124	129	45
Number of villages	1077	3825	3752	3844	1313
Number of HH	18098	57945	56106	52808	16013
Number of children	35540	110435	105352	104162	32694
Household and children characteristics					
Size of households	7.156	7.284	7.016	6.672	7.040
Household is poor	0.815	0.816	0.781	0.775	0.593
Household is ultra poor	0.320	0.335	0.306	0.301	0.162
Number of children	5.474	2.797	2.761	3.197	3.523
Child's age	11.354	11.187	11.167	11.162	11.062
Child is female	0.507	0.502	0.497	0.496	0.495
Child is enrolled in school	0.897	0.886	0.884	0.880	0.796
Number of completed grade	4.184	4.018	3.950	4.101	3.091
Child drops out school this year	0.053	0.056	0.061	0.066	0.037
Child never enrolled	0.050	0.059	0.056	0.054	0.167
Child attends government school	0.804	0.974	0.969	0.970	0.715
Children test scores					
Child reads words	0.698	0.642	0.643	0.759	0.747
Child does basic maths	0.893	0.834	0.858	0.767	0.766
Child reads words and does basic maths	0.682	0.628	0.634	0.704	0.694
Child passes math test	0.364	0.487	0.538	0.389	0.361
Child passes language test	0.490	0.433	0.419	0.510	0.493
Child passes math and language test	0.196	0.202	0.217	0.229	0.194
Child has an imputed score	0.028	0.039	0.027	0.167	0.146

FIGURE A3.1: CHILDREN ACTIVITIES BY AGE IN RURAL AREAS.



(A) GIRLS

(B) BOYS

Sources: LSMS pooled data (2008, 2010, 2012).

FIGURE A3.2: PERCENTAGE OF ENROLLED CHILDREN BY AGE COHORT.

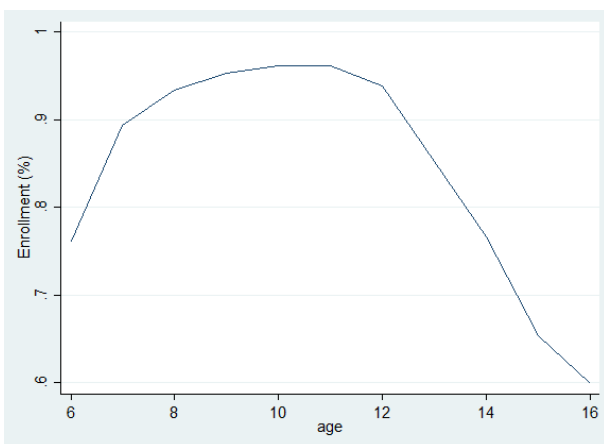
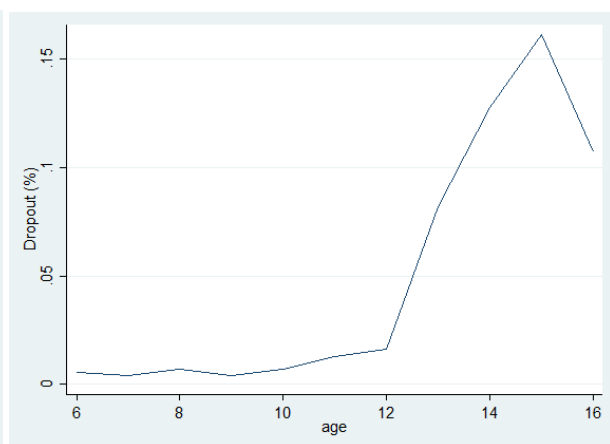
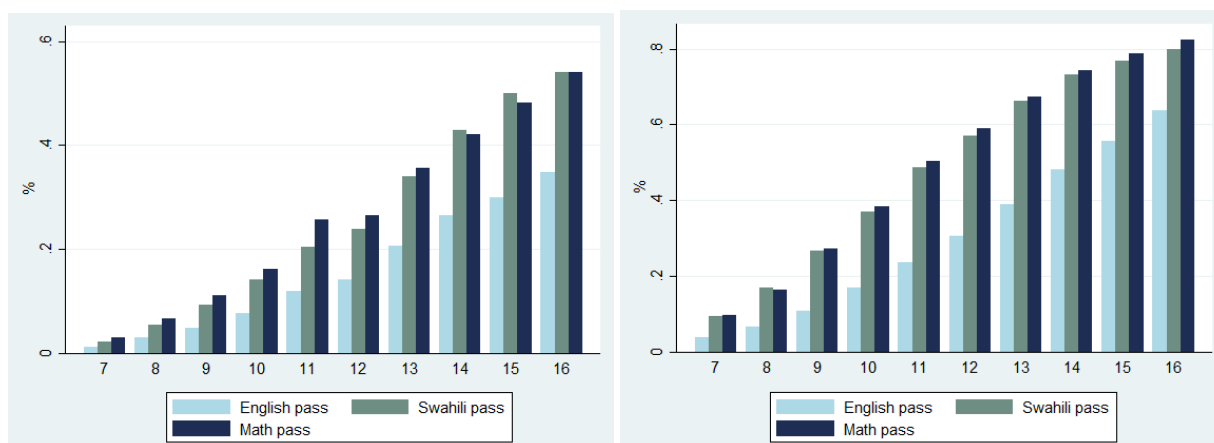


FIGURE A3.3: PERCENTAGE OF DROPOUT CHILDREN BY AGE COHORT.



Sources: LSMS pooled data (2008, 2010, 2012).

FIGURE A3.4: PERCENTAGE OF CHILDREN WHO PASSED THE EXAM BY AGE COHORT.

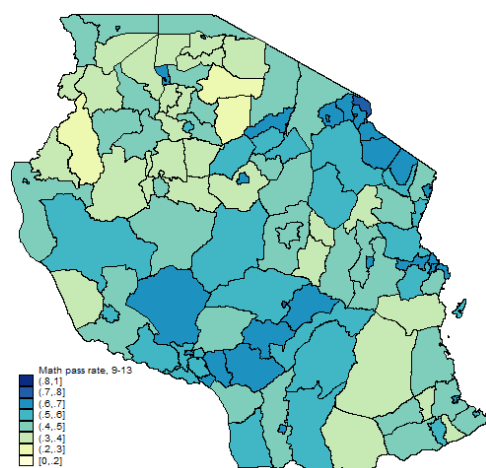


(A) UNENROLLED CHILDREN

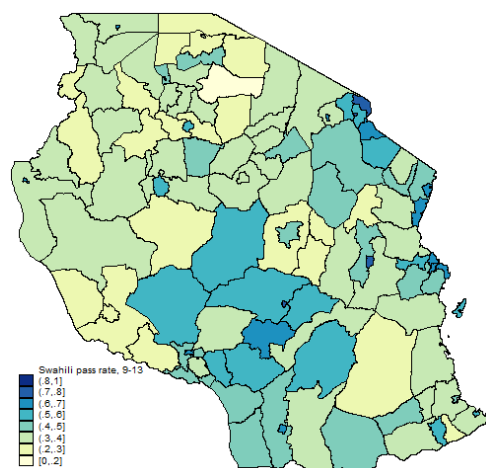
(B) ENROLLED CHILDREN

Sources: Uwezo data (20011, 2012, 2013, 2014).

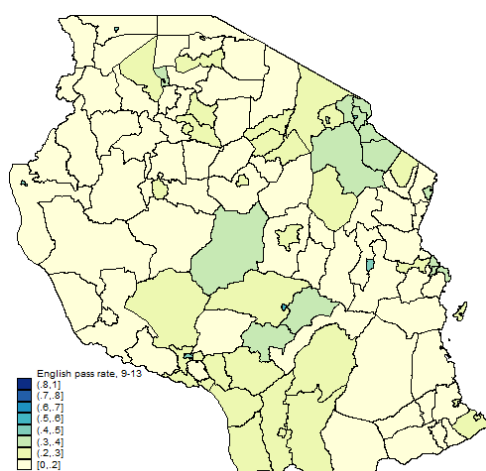
FIGURE A3.5: DISTRIBUTION OF CHILDREN WHO PASSED THE TESTS



(A) MATHS TEST, 9-13 YEARS OF AGE.



(B) SWAHILI TEST, 9-13 YEARS OF AGE.



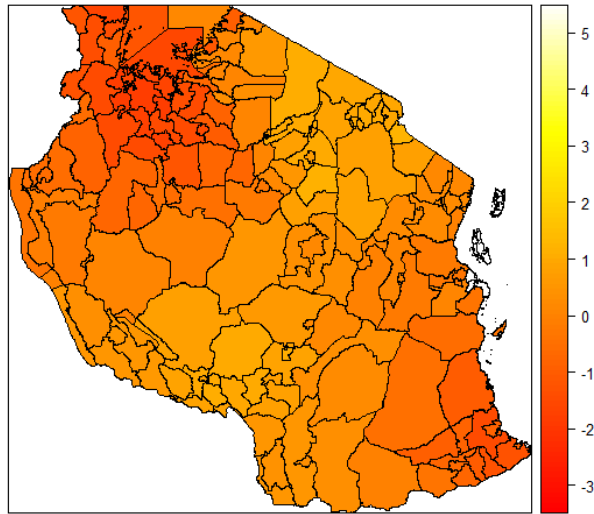
(C) ENGLISH TEST, 9-13 YEARS OF AGE.

Sources: Uwezo 2011 data.

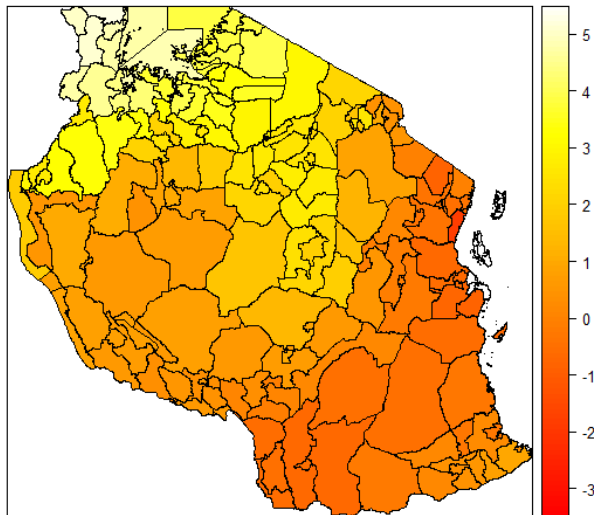


FIGURE A3.6: DISTRIBUTION OF SPEI BY DISTRICT

(A) IN 2008



(B) IN 2010



(C) IN 2012

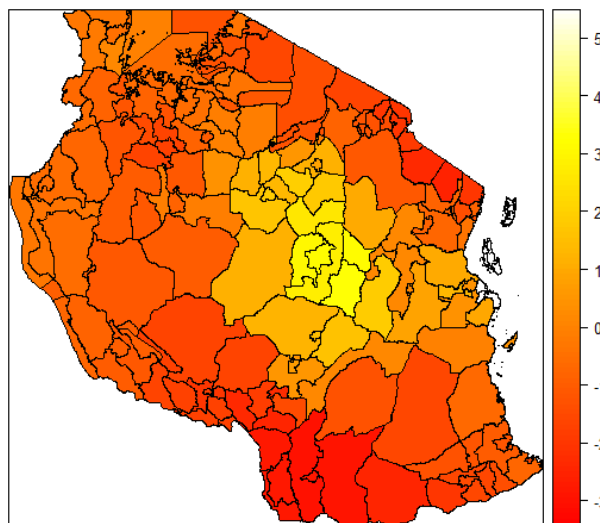
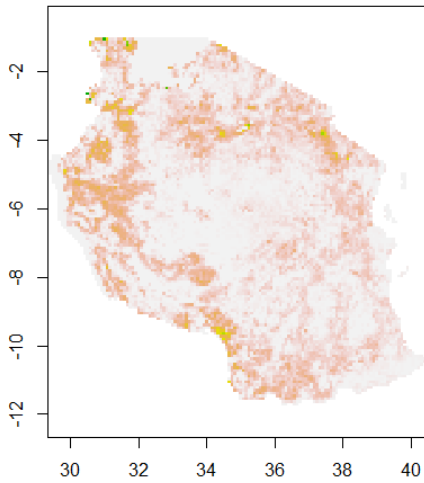
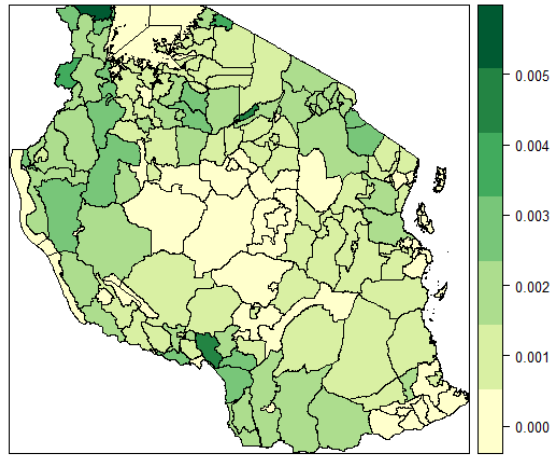


FIGURE A3.7: PERCENTAGE OF LAND ALLOCATED TO COFFEE PLANTATION IN TANZANIA.



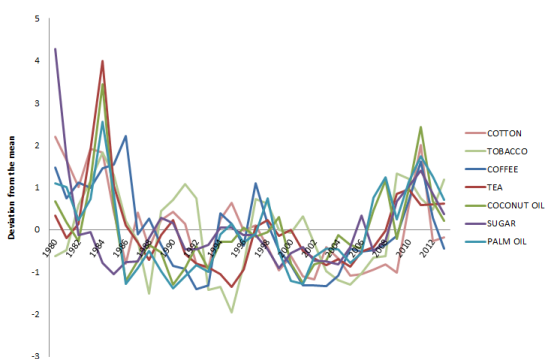
(A) CELLS OF 10KM\*10KM.



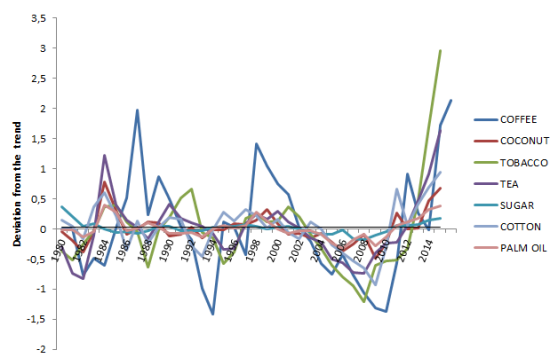
(B) AVERAGE BY DISTRICT.

Sources: Earth Stat data (2000).

FIGURE A3.8: STANDARDIZED PRICE DEVIATIONS FOR THE MAIN CASH-CROP COMMODITIES IN TANZANIA.



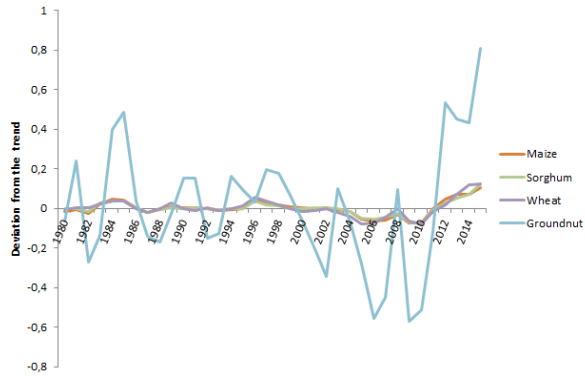
(A) DEVIATIONS FROM THE MEAN.



(B) DEVIATIONS FROM THE HP TREND.

Sources: World Bank Commodities Price Data.

FIGURE A3.9: STANDARDIZED PRICE DEVIATIONS FOR THE MAIN SUBSISTENCE COMMODITIES IN TANZANIA (DEVIATIONS FROM HP TREND).



Sources: World Bank Commodities Price Data.

## 3.9 Estimations of the mechanisms

### 3.9.1 Effect of shocks on production

TABLE A3.3: EFFECT OF CLIMATIC SHOCKS ON HOUSEHOLD PRODUCTION.

Climatic variable	log(Production)	log(Production)
Rainfall January-December	0.055 (0.087)	0.084 (0.092)
Temperature	-0.047*** (0.007)	0.003 (0.022)
Within R-squared	0.42	0.18
Rainfall March-May	0.171* (0.097)	0.219** (0.110)
Temperature	-0.047*** (0.007)	0.002 (0.022)
Within R-square	0.42	0.18
SPEI-6 months March-May	0.409*** (0.090)	0.326*** (0.099)
Within R-squared	0.404	0.177
SPEI-12 months March-May	0.383*** (0.084)	0.260*** (0.094)
Within R-squared	0.404	0.176
Localities and Times F.E	×	×
Households F.E		×
Observations	12,181	12,181

Sources: LSMS-ISA from 2008, 2010 and 2012. Note: Production and Consumption are computed in Tanzanian shillings. Standard errors, clustered by geographical units (0.5\*0.5° of precision), are reported in parentheses. Controls are survey month dummies, cultivated lands, the number of days of labor in the field and the age of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

### 3.9.2 Effect of shocks on labor allocation decisions

TABLE A3.4: EFFECT OF AGGREGATE SUBSISTENCE PRICE VARIABLE ON DAYS OF LABOR IN THE FIELD (BETA COEFFICIENTS).

	Hired labor	Adult labor	Child labor
$P_{jy-1}$ <i>subsistence crops</i>	-51.72 (48.40)	121.9 (216.2)	-136.3 (86.82)
Within R-squared	0.061	0.083	0.026
Observations	6,540	6,282	4,387
Localities and Time F.E	×	×	×
Household F.E	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: Standard errors, clustered at the household level, are reported in parentheses. Controls are survey month dummies, cultivated lands, the number of adults and the number of children in the household and the age of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.5: EFFECTS OF PRODUCTIVITY SHOCKS ON DAYS OF LABOR IN THE FIELDS BY AGE GROUPS (BETA COEFFICIENTS).

	(1)	(2)	(3)	(4)
	7-13	14-16	7-13	14-16
SPEI-6 months March-May	25.09*** (7.948)	27.38*** (7.848)	24.84*** (7.891)	27.39*** (7.807)
$P_{jy-1}$	5.575* (2.979)	8.839* (4.729)		
$P_{jy-1} * cash - crop producer$	31.34 (19.11)	20.78** (9.961)		
$P_{jy-1} S.R.$		5.572*	8.860* (3.016)	(4.784)
$P_{jy-1} S.R. * S.R. cash - crop producer$			30.19 (19.29)	19.59** (8.982)
Within R-squared	0.104	0.125	0.104	0.126
Observations	4,678	2,758	4,678	2,758
Localities and Times F.E	×	×	×	×
Households F.E	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: Standard errors, clustered by geographical units (0.5\*0.5° of precision), are reported in parentheses. Controls are survey month dummies, cultivated lands, the number of days of labor in the field and the age of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.6: EFFECTS OF PRODUCTIVITY SHOCKS ON THE PROBABILITY OF WORKING THE WEEK PRIOR THE SURVEY.

	(1)	(2)	(3)	(4)	(5)
	Work	Paid	Unpaid	Agriculture	Domestic
SPEI-6 months March-May	0.071*** (0.018)	0.010*** (0.004)	-0.022** (0.010)	0.022* (0.013)	0.009 (0.012)
$P_{jy-1}$	0.006 (0.010)	0.004* (0.002)	0.004 (0.005)	0.001 (0.012)	0.012** (0.006)
Within R-squared	0.113	0.036	0.557	0.080	0.086
Observations	12,674	12,788	12,788	12,788	12,788
Times F.E	×	×	×	×	×
Households F.E	×	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: standard errors, clustered by geographical units (0.5\*0.5° of precision), are reported in parentheses. Controls are survey month dummies, cultivated lands, the number of days of labor in the field and the age of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.7: EFFECTS OF PRODUCTIVITY SHOCKS ON THE LABOR INTENSITY BY ACTIVITY (HOURS OVER LAST WEEK).

	(1)	(2)	(3)	(4)
	Paid	Unpaid	Agriculture	Domestic
SPEI-6 months March-May	-0.970	-0.274	1.535*	0.278
	(0.750)	(0.473)	(0.920)	(0.455)
$P_{jy-1}$	0.513	0.101	0.515**	0.106
	(0.358)	(0.151)	(0.207)	(0.143)
Within R-squared	0.064	0.475	0.159	0.081
Observations	3,094	3,094	3,094	3,094
Times F.E	×	×	×	×
Households F.E	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: standard errors, clustered by geographical units ( $0.5^{\circ} \times 0.5^{\circ}$  of precision), are reported in parentheses. Controls are survey month dummies, cultivated lands, the number of days of labor in the field and the age of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.8: EFFECT OF CLIMATE AND PRICE VARIATIONS ON Z-SCORE OF HEIGHT FOR AGE (BETA COEFFICIENTS).

Age	0	1	2	3	4	5
$P_{jy}$	0.004	0.046*	-0.094	0.009	-0.039	0.039
	(0.040)	(0.026)	(0.093)	(0.033)	(0.032)	(0.034)
SPEI-6 months March-May <sub>y</sub>	0.387***	0.178**	-0.074	-0.077	0.053	0.098
	(0.146)	(0.082)	(0.074)	(0.070)	(0.090)	(0.083)
Within R-squared R-squared	0.086	0.032	0.062	0.030	0.089	0.048
Observations	1,294	1,304	1,406	1,304	1,360	1,278
Localities and Times F.E	×	×	×	×	×	×
Households F.E	×	×	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: standard errors, clustered by geographical units ( $0.5^{\circ} \times 0.5^{\circ}$  of precision), are reported in parentheses. Controls are survey month dummies, age dummies and years of the survey. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

## 3.10 Heterogeneous effects

### 3.10.1 Current shocks

TABLE A3.9: EFFECT OF POSITIVE AND NEGATIVE SHOCKS ON CHILDREN'S ACTIVITIES ACCORDING TO HOUSEHOLD CONSUMPTION

	Work	Enrolled	Dropout	Grade
Below the median consumption				
Positive Price Shock <sub>y-1</sub>	0.059* (0.033)	-0.056*** (0.021)	0.017 (0.013)	-0.065 (0.083)
Positive Rainfall Shock <sub>y-1</sub>	0.088** (0.035)	0.003 (0.014)	0.012 (0.009)	-0.107** (0.054)
Negative Price Shock <sub>y-1</sub>	-0.008 (0.027)	-0.020 (0.016)	0.003 (0.010)	0.014 (0.076)
Drought <sub>y-1</sub>	-0.009 (0.027)	0.026 (0.018)	-0.014 (0.008)	0.015 (0.058)
Above the median consumption				
Positive Price Shock <sub>y-1</sub>	0.059 (0.039)	-0.022 (0.017)	-0.005 (0.014)	-0.085 (0.085)
Positive Rainfall Shock <sub>y-1</sub>	0.073* (0.041)	-0.004 (0.018)	0.019* (0.011)	-0.143* (0.073)
Negative Price Shock <sub>y-1</sub>	-0.027 (0.033)	0.029* (0.017)	-0.024* (0.015)	-0.038 (0.084)
Drought <sub>y-1</sub>	0.030 (0.037)	-0.012 (0.017)	0.008 (0.011)	-0.073 (0.066)
R-squared	0.17	0.16	0.0856	0.70
Observations	12,677	11,625	11,230	10,588
Localities F.E	×	×	×	×
Year F.E	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: standard errors, clustered by geographical units (0.5°×0.5° of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the number of boys among siblings, the age and the education of the household head. Coefficients are computed with the delta method. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.



TABLE A3.10: EFFECT OF POSITIVE AND NEGATIVE SHOCKS ON CHILDREN'S ACTIVITIES BY GENDER.

	Work	Enrolled	Dropout	Grade
Girls				
Positive Price Shock <sub>y-1</sub>	0.058* (0.035)	-0.044** (0.019)	0.005 (0.012)	-0.061 (0.084)
Positive Rainfall Shock <sub>y-1</sub>	0.071** (0.035)	-0.003 (0.016)	0.013 (0.011)	-0.113** (0.054)
Negative Price Shock <sub>y-1</sub>	0.005 (0.028)	-0.022 (0.016)	0.001 (0.011)	-0.059 (0.087)
Negative Rainfall Shock <sub>y-1</sub>	0.004 (0.032)	0.010 (0.017)	-0.008 (0.009)	-0.032 (0.047)
Boys				
Positive Price Shock <sub>y-1</sub>	0.059* (0.034)	-0.025 (0.019)	0.004 (0.013)	-0.062 (0.095)
Positive Rainfall Shock <sub>y-1</sub>	0.097*** (0.033)	0.005 (0.017)	0.016 (0.010)	-0.131** (0.062)
Negative Price Shock <sub>y-1</sub>	-0.033 (0.025)	0.015 (0.016)	-0.013 (0.011)	0.074 (0.083)
Negative Rainfall Shock <sub>y-1</sub>	0.009 (0.027)	0.009 (0.020)	-0.001 (0.009)	-0.028 (0.059)
R-squared	0.17	0.09	0.315	0.67
Observations	12,677	11,625	11,230	10,588
Localities F.E	×	×	×	×
Year F.E	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: standard errors, clustered by geographical units (0.5\*0.5° of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the number of boys among siblings, the age and the education of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.11: EFFECT OF POSITIVE AND NEGATIVE SHOCKS ON CHILDREN'S ACTIVITIES ACROSS AGE-GROUP.

	Work	Enrolled	Dropout	Grade
7-13 age group				
Positive Price Shock <sub>t-1</sub>	0.021 (0.031)	0.003 (0.017)	0.002 (0.012)	-0.089 (0.086)
Positive Rainfall Shock <sub>t-1</sub>	0.073** (0.033)	0.042*** (0.014)	0.004 (0.008)	-0.086* (0.045)
Negative Price Shock <sub>t-1</sub>	-0.011 (0.027)	-0.002 (0.016)	-0.008 (0.011)	0.055 (0.075)
Drought <sub>t-1</sub>	0.005 (0.027)	0.008 (0.015)	-0.003 (0.008)	-0.002 (0.048)
14-16 age group				
Positive Price Shock <sub>t-1</sub>	0.154*** (0.043)	-0.127*** (0.034)	0.011 (0.024)	0.017 (0.119)
Positive Rainfall Shock <sub>t-1</sub>	0.113*** (0.037)	-0.096*** (0.027)	0.039** (0.019)	-0.214* (0.110)
Negative Price Shock <sub>t-1</sub>	-0.020 (0.030)	-0.010 (0.026)	-0.001 (0.021)	-0.113 (0.119)
Drought <sub>t-1</sub>	0.007 (0.035)	0.010 (0.030)	-0.007 (0.021)	-0.111 (0.091)
R-squared	0.171	0.162	0.0852	0.694
Observations	12,677	11,625	11,230	10,588
Localities F.E	×	×	×	×
Year F.E	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: Read and Write variable is only available for 2010 and 2012. Standard errors, clustered by geographical units ( $0.5^\circ \times 0.5^\circ$  of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the birth order of the child, the number of boys among siblings, the age and the education of the household head. Coefficients are computed with the delta method. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

### 3.10.2 Cumulative shocks

TABLE A3.12: EFFECT OF THE NUMBER OF SHOCKS ACCORDING TO HOUSEHOLD CONSUMPTION.

	Work	Overage	Grade	Read and write
Below the median consumption				
Number Positive Price Shocks	0.013 (0.019)	0.007 (0.012)	-0.046 (0.053)	-0.029* (0.016)
Number Positive Rainfall Shocks	0.052*** (0.013)	0.041*** (0.013)	-0.134*** (0.047)	0.019 (0.014)
Above the median consumption				
Number Positive Price Shocks	0.019 (0.020)	0.031** (0.014)	-0.056 (0.050)	-0.022* (0.012)
Number Positive Rainfall Shockst	0.030** (0.014)	-0.020 (0.013)	0.074 (0.047)	0.002 (0.014)
R-squared	0.167	0.259	0.700	0.236
Observations	12,677	10,588	10,588	9,100
Localities F.E	×	×	×	×
Year F.E	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: Read and Write variable is only available for 2010 and 2012. Standard errors, clustered by geographical units ( $0.5^\circ \times 0.5^\circ$  of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the birth order of the child, the number of boys among siblings, the age and the education of the household head. Coefficients are computed with the delta method. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.13: EFFECT OF THE NUMBER OF SHOCKS ACROSS AGE-GROUPS.

	Work	Overage	Grade	Read and write
7-13 years old				
Number Positive Price Shocks	0.000 (0.018)	-0.004 (0.005)	-0.035 (0.042)	-0.030** (0.014)
Number Positive Rainfall Shocks	0.051*** (0.014)	-0.005 (0.004)	-0.008 (0.035)	-0.001 (0.013)
14-16 years old				
Number Positive Price Shocks	0.041** (0.019)	-0.006 (0.011)	-0.045 (0.061)	-0.015 (0.016)
Number Positive Rainfall Shocks	0.042*** (0.013)	0.008 (0.009)	-0.102* (0.058)	0.022 (0.014)
R-squared	0.167	0.0845	0.694	0.230
Observations	12,677	11,230	10,588	9,100
Localities F.E	×	×	×	×
Year F.E	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: Read and Write variable is only available for 2010 and 2012. Standard errors, clustered by geographical units ( $0.5^\circ \times 0.5^\circ$  of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the birth order of the child, the number of boys among siblings, the age and the education of the household head. Coefficients are computed with the delta method. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.14: EFFECT OF THE NUMBER OF SHOCKS ACROSS GENDER.

	Work	Overage	Grade	Read and write
Girls				
Number Positive Price Shocks	0.042*** (0.013)	0.010 (0.012)	-0.007 (0.043)	-0.011 (0.015)
Number Positive Rainfall Shocks	0.054*** (0.014)	0.016 (0.015)	-0.045 (0.056)	0.014 (0.015)
Boys				
Number Positive Price Shocks	0.023 (0.022)	-0.017 (0.046)	-0.003 (0.016)	-0.017 (0.015)
Number Positive Rainfall Shocks	0.054*** (0.013)	0.022* (0.012)	-0.101** (0.044)	0.010 (0.013)
R-squared	0.191	0.249	0.743	0.258
Observations	12,677	10,588	10,588	9,100
Localities F.E	×	×	×	×
Year F.E	×	×	×	×

Note: Sources: LSMS-ISA from 2008, 2010 and 2012. Note: Standard errors, clustered by geographical units ( $0.5^\circ \times 0.5^\circ$  of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the birth order of the child, the number of boys in the sibling, the age and the education of the household head. Coefficients are computed with the delta method. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.15: EFFECT OF SHOCKS DURING SCHOOL AGE ON TEST SCORES BY AGE GROUPS

	Swahili	Maths	Swahili	Maths
Primary age children				
Number Positive Price Shocks	-0.022*** (0.008)	-0.035*** (0.010)	-0.017** (0.007)	-0.028*** (0.009)
Number Positive Rainfall Shocks	-0.014 (0.011)	-0.021 (0.014)	-0.020** (0.010)	-0.023* (0.012)
Secondary age children				
Number Positive Price Shocks	-0.026*** (0.007)	-0.0314*** (0.00717)	-0.029*** (0.007)	-0.0315*** (0.00715)
Number Positive Rainfall Shocks	-0.032*** (0.012)	-0.019 (0.013)	-0.0330** (0.0129)	-0.018 (0.013)
R-squared	0.323	0.295	0.323	0.289
Observations	328,948	328,948	294,521	294,521
District F.E	×	×	×	×
Year F.E	×	×	×	×
Attend School			×	×

Sources: Uwezo data from 2011 to 2014. Note: Standard errors are clustered at the district level and are reported in parentheses. Controls are years dummies, the percentage of lands allocated to crop production by district, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the age and the education of the household head. Coefficients are computed with the delta method. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.16: EFFECT OF SHOCKS DURING SCHOOL AGE ON TEST SCORES BY GENDER

	Swahili	Maths	Swahili	Maths
Girls				
Number Positive Price Shocks	-0.028*** (0.007)	-0.0318*** (0.00787)	-0.027*** (0.007)	-0.0283*** (0.00726)
Number Positive Rainfall Shocks	-0.028*** (0.010)	-0.025* (0.013)	-0.0341*** (0.0110)	-0.027** (0.013)
Boys				
Number Positive Price Shocks	-0.025*** (0.007)	-0.033*** (0.007)	-0.027*** (0.007)	-0.033*** (0.007)
Number Positive Rainfall Shocks	-0.026** (0.011)	-0.013 (0.012)	-0.027** (0.012)	-0.011 (0.012)
R-squared	0.325	0.295	0.324	0.289
Observations	328,948	328,948	294,521	294,521
District F.E	×	×	×	×
Year F.E	×	×	×	×
Attend School			×	×

Sources: Uwezo data from 2011 to 2014. Note: Standard errors are clustered at the district level and are reported in parentheses. Controls are years dummies, the percentage of lands allocated to crop production by district, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the age and the education of the household head. Coefficients are computed with the delta method. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

### 3.11 Robustness checks

TABLE A3.17: EFFECT OF CONTINUOUS CLIMATE AND PRICE VARIABLES ON CHILDREN'S ACTIVITIES (BETA COEFFICIENTS).

	Work	Enrolled	Dropout	Grade
$P_{jy-1}$	-0.009 (0.017)	0.001 (0.018)	0.013 (0.011)	0.011 (0.031)
SPEI-6 months March-May $_{y-1}$	0.040 (0.025)	0.018 (0.014)	0.003 (0.011)	-0.003 (0.046)
R-squared	0.138	0.181	0.083	0.667
Observations	3,653	3,010	3,245	3,052
Localities F.E	×	×	×	×
Year F.E	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Note: Standard errors, clustered by geographical units ( $0.5^\circ \times 0.5^\circ$  of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the number of boys in the sibling, the age and the education of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.18: EFFECT OF CONTINUOUS CLIMATE AND PRICE VARIABLES ON TEST SCORES (BETA COEFFICIENTS).

	Swahili	Maths	Swahili	Maths
$P_{jy-1}$	-0.003 (0.008)	0.000 (0.006)	-0.002 (0.006)	0.001 (0.005)
SPEI-6 months March-May $_{y-1}$	-0.032** (0.014)	-0.018 (0.015)	-0.033** (0.014)	-0.023 (0.014)
R-squared	0.320	0.292	0.321	0.287
Observations	321,900	321,900	287,856	287,856
District F.E	×	×	×	×
Year F.E	×	×	×	×
Attend school			×	×

Sources: Uwezo data from 2011 to 2014. Note: Standard errors are clustered at the district level and are reported in parentheses. Controls are years dummies, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child,, the age and the education of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.19: EFFECT OF FUTURE PRICE AND RAINFALLS ON EDUCATION DECISIONS.(BETA COEFFICIENTS).

	Work	Enrolled	Dropout	Grade
$P_{jy-1}$	0.003 (0.015)	-0.006 (0.009)	0.010* (0.006)	0.005 (0.020)
$P_{jy+1}$	0.000 (0.000)	0.000 (0.000)	-0.000 (0.000)	-0.000 (0.000)
SPEI-6 months March-May $_{y-1}$	0.061** (0.028)	0.001 (0.012)	0.010* (0.006)	0.021 (0.014)
SPEI-6 months March-May $_{y+1}$	0.029 (0.024)	-0.007 (0.011)	-0.000 (0.005)	0.012 (0.013)
R-squared	0.167	0.154	0.085	0.230
Observations	12,677	11,625	11,230	9,100
Localities F.E	×	×	×	×
Year F.E	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: Standard errors, clustered by geographical units ( $0.5^\circ \times 0.5^\circ$  of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the number of boys among siblings, the age and the education of the household head. Coefficients are computed with the delta method. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.20: EFFECT OF FUTURE SHOCKS ON EDUCATION DECISIONS.

	Work	Enrolled	Dropout	Grade
Positive Price Shock <sub>y-1</sub>	0.017 (0.054)	-0.079*** (0.025)	0.026 (0.018)	0.021 (0.108)
Positive Price Shock <sub>y+1</sub>	0.043 (0.048)	0.052** (0.024)	-0.023 (0.018)	-0.078 (0.092)
Positive Rainfall Shock <sub>y-1</sub>	0.082** (0.036)	-0.004 (0.017)	0.018* (0.010)	-0.084 (0.052)
Positive Rainfall Shock <sub>y+1</sub>	0.011 (0.025)	-0.026 (0.025)	0.010 (0.010)	0.082 (0.058)
Negative Price Shock <sub>y-1</sub>	-0.002 (0.025)	-0.008 (0.013)	-0.008 (0.009)	-0.006 (0.074)
Negative Price Shock <sub>y+1</sub>	-0.094** (0.039)	0.055** (0.023)	0.003 (0.012)	0.084 (0.088)
Drought <sub>y-1</sub>	0.003 (0.026)	0.014 (0.015)	-0.010 (0.009)	-0.048 (0.058)
Drought <sub>y+1</sub>	-0.033 (0.027)	0.008 (0.014)	-0.005 (0.008)	0.058 (0.053)
R-squared	0.170	0.156	0.085	0.694
Observations	12,677	11,625	11,230	10,588
Localities F.E	×	×	×	×
Year F.E	×	×	×	×

Sources: LSMS-ISA from 2008, 2010 and 2012. Note: Standard errors, clustered by geographical units ( $0.5^\circ \times 0.5^\circ$  of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the number of boys among siblings, the age and the education of the household head. Coefficients are computed with the delta method. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.



TABLE A3.21: EFFECT OF SHOCKS DURING SCHOOL AGE ON TEST SCORES STANDARDIZED BY AGE.

	Swahili	Maths	Swahili	Maths
Primary age children				
Number Positive Price Shocks	-0.017* (0.010)	-0.044*** (0.014)	-0.013 (0.009)	-0.033** (0.013)
Number Positive Rainfall Shocks	-0.020* (0.012)	-0.027* (0.014)	-0.032*** (0.011)	-0.034*** (0.013)
R-squared	0.306	0.276	0.295	0.257
Observations	249,186	249,186	226,020	226,020
Secondary age children				
Number Positive Price Shocks	-0.036 (0.024)	-0.064** (0.028)	-0.030 (0.021)	-0.058** (0.024)
Number Positive Rainfall Shocks	-0.038 (0.027)	-0.030 (0.029)	-0.041 (0.030)	-0.037 (0.029)
R-squared	0.060	0.072	0.051	0.077
Observations	48,072	48,072	40,279	40,279
District F.E	×	×	×	×
Year F.E	×	×	×	×
Attend School			×	×

Sources: Uwezo data from 2011 to 2014. Note: Standard errors are clustered at the district level and are reported in parentheses. Controls are years dummies, the percentage of lands allocated to crop production by district, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the age and the education of the household head. Coefficients are computed with the delta method. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.22: EFFECT OF CONTINUOUS CLIMATE AND PRICE VARIABLES ON SAMPLE SELECTION (BETA COEFFICIENTS).

	Mortality	Migration
$P_{jy-1}$	0.013 (0.015)	-0.018 (0.019)
$P_{jy-2}$	-0.009 (0.014)	0.021 (0.019)
SPEI-6 months March-May $_{y-1}$	0.002 (0.010)	-0.007 (0.009)
SPEI-6 months March-May $_{y-2}$	-0.005 (0.012)	-0.007 (0.010)
R-squared	0.066	0.103
Observations	4,784	2,618
Localities F.E	×	×
Households F.E	×	×
Sample	Agricultural HH	Children aged 7-16

Sources: LSMS-ISA from 2008, 2010 and 2012. Notes: Standard errors, clustered by geographical units ( $0.5^\circ \times 0.5^\circ$  of precision), are reported in parentheses. Controls are survey month dummies, the number of adults and the number of children in the household, age dummies, the gender and the birth order of the child, the number of boys among siblings, the age and the education of the household head. \*\*\*, \*\*, \* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

TABLE A3.23: EFFECT OF CLIMATE AND PRICES ON QUALITY OF EDUCATION

	(1)	(2)
	Attend teachers	Qualified teachers
Continuous variables		
SPEI-6months March-May <sub><i>jy</i></sub>	0.006 (0.007)	0.009 (0.013)
$P_{jy}$	0.001 (0.005)	-0.013*** (0.004)
Within R-squared	0.029	0.013
Shocks		
Positive Rainfall Shock <sub><i>jy</i></sub>	-0.000 (0.010)	0.014 (0.036)
Positive Price Shock <sub><i>jy</i></sub>	0.015 (0.012)	-0.009 (0.017)
Negative Rainfall Shock <sub><i>jy</i></sub>	0.007 (0.009)	0.017 (0.021)
Negative Price Shock <sub><i>jy</i></sub>	-0.009 (0.013)	0.096 (0.069)
Within R-squared	0.03	0.016
Observations	9,356	9,356
Localities F.E	×	×
Month and Year F.E	×	×

Sources: Uwezo data from 2011 to 2014. Notes: Standard errors are clustered at the district level and are reported in parentheses. In columns (2) and (3), I control by the number of recorded actual teachers. \*\*\*,\*\*,\* mean respectively that the coefficients are significantly different from 0 at the level of 1%, 5% and 10%.

# References

- ADDISON, T., GHOSHRAJ, A. and STAMATOIANNIS, M.P., 2016, 'Agricultural commodity price shocks and their effect on growth in Sub-Saharan Africa', *Journal of Agricultural Economics*, 67(1), 47–61.
- AKABAYASHI, H. and PSACHAROPOULOS, G., 1999, 'The trade-off between child labour and human capital formation: A Tanzanian case study', *The Journal of Development Studies*, 35(5), 120–140.
- ALMOND, D. and CURRIE, J., 2011, 'Killing me softly: The fetal origins hypothesis', *The Journal of Economic Perspectives*, 25(3), 153–172.
- ANDERSON, K. and BRÜCKNER, M., 2012, 'Distortions to agriculture and economic growth in sub-Saharan Africa', Mimeo. Available at [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2151768](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2151768).
- ANDRE, P., DELESALLE, E. and DUMAS, C., 2017, 'Returns to Farm Child Labor in Tanzania', Mimeo. Available at [https://economics.handels.gu.se/digitalAssets/1643/1643674\\_57.-delesalle\\_esther-returns-to-farm.pdf](https://economics.handels.gu.se/digitalAssets/1643/1643674_57.-delesalle_esther-returns-to-farm.pdf).
- ANKER, R., BARGE, S., ASHRAF, S. and LEVISON, D., 1998, 'Economics of Child Labour in India's Carpet Industry', In: R. Anker, S. Barge, S. Rajagopal and M. Joseph (eds.), *Economics of Child Labour in Hazardous Industries of India*, Center for Operations Research and Training.
- APPLETON, S., BALIHUTA, A. *et al.*, 1996, *Education and agricultural productivity: Evidence from Uganda*, University of Oxford, Centre for the Study of African Economies.

- ATTANASIO, O., FITZSIMONS, E., GOMEZ, A., GUTIERREZ, M.I., MEGHIR, C. and MESNARD, A., 2010, 'Children's schooling and work in the presence of a conditional cash transfer program in rural Colombia', *Economic development and cultural change*, 58(2), 181–210.
- BARHAM, B. and BOUCHER, S., 1998, 'Migration, remittances, and inequality: estimating the net effects of migration on income distribution', *Journal of Development Economics*, 55(2), 307–331.
- BARRERA-OSORIO, F., BERTRAND, M., LINDEN, L.L. and PEREZ-CALLE, F., 2011, 'Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia', *American Economic Journal: Applied Economics*, 3(2), 167–95.
- BASALIRWA, C., ODIYO, J., MNGODO, R. and MPETA, E., 1999, 'The climatological regions of Tanzania based on the rainfall characteristics', *International Journal of Climatology*, 19(1), 69–80.
- BASU, K. and VAN, P., 1998, 'The economics of child labor', *American Economic Review*, 88(3), 412–427.
- BAUMEISTER, C. and KILIAN, L., 2014, 'Do oil price increases cause higher food prices?', *Economic Policy*, 29(80), 691–747.
- BEEGLE, K., DEHEJIA, R.H. and GATTI, R., 2006a, 'Child labor and agricultural shocks', *Journal of Development Economics*, 81(1), 80–96.
- BEEGLE, K., DEHEJIA, R.H. and GATTI, R., 2006b, 'Child Labor and Agricultural Shocks', *Journal of Development Economics*, 81(1), 80–96.
- BEHRMAN, J.R., PARKER, S.W. and TODD, P.E., 2011, 'Do Conditional Cash Transfers for Schooling Generate Lasting Benefits?: A Five-Year Followup of PROGRESA/Oportunidades', *Journal of Human Resources*, 46(1), 93–122.
- BENJAMIN, D., 1992, 'Household composition, labor markets, and labor demand: testing for separation in agricultural household models', *Econometrica: Journal of the*

*Econometric Society*, 287–322.

BERTRAND, M., DUFLO, E. and MULLAINATHAN, S., 2004, ‘How much should we trust differences-in-differences estimates?’, *The Quarterly Journal of Economics*, 119(1), 249–275.

BHALOTRA, S. and HEADY, C., 2003, ‘Child Farm Labour: the Wealth Paradox.’, *World Bank Economic Review*, 17(2), 197–227.

BONINI, N., 2003, ‘Un siècle d’éducation scolaire en Tanzanie’, *Cahiers d’études africaines*, (1), 40–62.

BOOZER, M. and SURI, T., 2001, ‘Child labor and schooling decisions in Ghana’, Mimeo, Yale University.

BOS, J.W. and KOETTER, M., 2011, ‘Handling losses in translog profit models’, *Applied Economics*, 43(3), 307–312.

BOURGUIGNON, F., FERREIRA, F.H.G. and LEITE, P.G., 2003, ‘Conditional Cash Transfers, Schooling, and Child Labor: Micro-Simulating Brazil’s Bolsa Escola Program’, *World Bank Economic Review*, 17(2), 229–254.

BURBIDGE, J.B., MAGEE, L. and ROBB, A.L., 1988, ‘Alternative Transformations to Handle Extreme Values of the Dependent Variable’, *Journal of the American Statistical Association*, 83(401), 123–127.

CAMERON, A.C., GELBACH, J.B. and MILLER, D.L., 2008, ‘Bootstrap-based improvements for inference with clustered errors’, *The Review of Economics and Statistics*, 90(3), 414–427.

CARD, D., 2001, ‘Estimating the return to schooling: Progress on some persistent econometric problems’, *Econometrica*, 69(5), 1127–1160.

CARPIO, X.V.D., LOAYZA, N.V. and WADA, T., 2016, ‘The impact of conditional cash transfers on the amount and type of child labor’, *World Development*, 80, 33 – 47.

CHAVAS, J.P., PETRIE, R. and ROTH, M., 2005, ‘Farm household production efficiency: Evidence from the Gambia’, *American Journal of Agricultural Economics*, 87(1), 160–

- CHENNAREDDY, V., 1967, 'Production Efficiency in South Indian Agriculture', *Journal of Farm Economics*, 49(4), 816–820.
- COGNEAU, D. and JEDWAB, R., 2012, 'Commodity Price Shocks and Child Outcomes: The 1990 Cocoa Crisis in Côte d'Ivoire', *Economic Development and Cultural Change*, 60(3), 507–534.
- COURT, D. and KINYANJUI, K., 1980, 'Development policy and educational opportunity: the experience of Kenya and Tanzania', Occasional Paper 33, Nairobi: Institute for Development Studies, University of Nairobi.
- CURRIE, J. and VOGL, T., 2013, 'Early-life health and adult circumstance in developing countries', *Annual Review of Economics*, 5(1), 1–36.
- DAMMERT, A.C., 2009, 'Heterogeneous impacts of conditional cash transfers: Evidence from Nicaragua', *Economic Development and Cultural Change*, 58(1), 53–83.
- DE CHAISEMARTIN, C. and D HAULTFOEUILLE, X., 2015, 'Fuzzy differences-in-differences', Cemmap working paper No. CWP69/15, Centre for Microdata Methods and Practice.
- DE JANVRY, A., FAFCHAMPS, M. and SADOULET, E., 1991, 'Peasant household behaviour with missing markets: Some paradoxes explained', *The Economic Journal*, 101(409), 1400–1417.
- DE LEON, J. and PARKER, S., 2000, 'The impact of anti-poverty programs on children's time use: The case of Progresa in Mexico', Working paper, Progresa.
- DE VREYER, P., GUILBERT, N. and MESPLE-SOMPS, S., 2014, 'Impact of natural disasters on education outcomes: Evidence from the 1987–89 locust plague in Mali', *Journal of African Economies*, 24(1), 57–100.
- DEATON, A. and ZAIDI, S., 2002, *Guidelines for constructing consumption aggregates for welfare analysis*, volume 135, World Bank Publications.
- DEHEJIA, R.H. and GATTI, R., 2005, 'Child labor: the role of financial development

- and income variability across countries', *Economic Development and Cultural Change*, 53(4), 913–931.
- DEVEREUX, S., MAXWELL, S. *et al.*, 2001, *Food security in sub-Saharan Africa.*, ITDG Publishing.
- DIALLO, Y., ETIENNE, A. and MEHRAN, F., 2013, 'Global Child Labour Trends 2008 to 2012', Technical report, International Labour Office, International Programme on the Elimination of Child Labour (IPEC), Geneva.
- DREZE, J., SEN, A. *et al.*, 1999, *India: Economic development and social opportunity*, Oxford University Press.
- DUBE, O. and VARGAS, J.F., 2013, 'Commodity price shocks and civil conflict: Evidence from Colombia', *The Review of Economic Studies*, 80(4), 1384–1421.
- DUFLO, E., 2001, 'Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment', *American economic review*, 91(4), 795–813.
- DUMAS, C., 2007, 'Why do parents make their children work? A Test of the Poverty hypothesis in rural areas of Burkina Faso.', *Oxford Economic Papers*, 59(2), 301–329.
- DUMAS, C., 2012, 'Does work impede child learning? The case of Senegal', *Economic Development and Cultural Change*, 60(4), 773–793.
- DUMAS, C., 2015, *Shocks and child labor: the role of markets*, Faculté des sciences économiques et sociales.
- EDMONDS, E.V. and SCHADY, N., 2012, 'Poverty alleviation and child labor', *American Economic Journal: Economic Policy*, 4(4), 100–124.
- EDMONDS, E.V. and SHRESTHA, M., 2014, 'You get what you pay for: Schooling incentives and child labor', *Journal of Development Economics*, 111(C), 196–211.
- ELBERS, C., LANJOUW, J.O. and LANJOUW, P., 2003, 'Micro-level estimation of poverty and inequality', *Econometrica*, 71(1), 355–364.
- FERREIRA, F.H. and SCHADY, N., 2009, 'Aggregate economic shocks, child schooling,



- and child health', *The World Bank Research Observer*, 24(2), 147–181.
- FILMER, D. and PRITCHETT, L., 2001, 'Estimating Wealth Effects without Expenditure Data - or Tears: an Application of Educational Enrollment in States of India.', *Demography*, 38(1), 115–132.
- FOSTER, A.D. and ROSENZWEIG, M.R., 1996, 'Technical change and human-capital returns and investments: evidence from the green revolution', *The American economic review*, 931–953.
- GALIANI, S. and MCEWAN, P.J., 2013, 'The heterogeneous impact of conditional cash transfers', *Journal of Public Economics*, 103, 85 – 96.
- GERTLER, P. and GLEWWE, P., 1992, 'The willingness to pay for education for daughters in contrast to sons: Evidence from rural Peru', *The World Bank Economic Review*, 6(1), 171–188.
- GLEWWE, P., 2002, 'Schools and skills in developing countries: Education policies and socioeconomic outcomes', *Journal of Economic Literature*, 40(2), p. 436.
- GRILICHES, Z., 1964, 'Research expenditures, education, and the aggregate agricultural production function', *The American Economic Review*, 54(6), 961–974.
- GRILICHES, Z., 1977, 'Estimating the returns to schooling: Some econometric problems', *Econometrica*, 1–22.
- GROOTAERT, C. and KANBUR, R., 1995, 'Child labour: an economic perspective', *International Labour Review*, 134, p. 187.
- GUARCELLO, L., LYON, S. and VALDIVIA, C., 2016, 'Evolution of the Relationship between Child Labour and Education since 2000: Evidence from 19 Developing Countries', Background paper prepared for the Education for all global monitoring report 2015, Education for All 2000-2015: Achievements and challenges.
- GUARCELLO, L., MEALLI, F. and ROSATI, F.C., 2010, 'Household vulnerability and child labor: the effect of shocks, credit rationing, and insurance', *Journal of Population Economics*, 23(1), 169–198.

- GUBERT, F., LASSOURD, T. and MESPLÉ-SOMPS, S., 2010, 'Transferts de fonds des migrants, pauvreté et inégalités au Mali', *Revue économique*, 61(6), 1023–1050.
- GUBERT, F. and ROBILLIARD, A.S., 2007, 'Risk and schooling decisions in rural Madagascar: A panel data-analysis', *Journal of African Economies*, 17(2), 207–238.
- HARARI, M. and LA FERRARA, E., 2014, 'Conflict, climate and cells: A disaggregated analysis', Mimeo, Bocconi University.
- HENNINGSEN, A. and HENNINGSEN, G., 2012, 'On estimation of the CES production function - Revisited', *Economics Letters*, 115(1), 67–69.
- IMBERT, C., SEROR, M., ZHANG, Y. and ZYLBERBERG, Y., 2016, 'Internal Migration and Firm Growth: Evidence from China.', Mimeo.
- IPEC, 2007, 'Child labour wages and productivity: Results from demand-side surveys', Technical report, International Labour Office.
- JACOBY, H. and SKOUFIAS, E., 1997a, 'Risk, financial markets and human capital in a developing country', *Review of Economic Studies*, 64(3), 311–335.
- JACOBY, H.G., 1993, 'Shadow wages and peasant family labour supply: An econometric application to the Peruvian sierra', *The Review of Economic Studies*, 60(4), 903–921.
- JACOBY, H.G. and SKOUFIAS, E., 1997b, 'Risk, financial markets, and human capital in a developing country', *The Review of Economic Studies*, 64(3), 311–335.
- JENSEN, R., 2000, 'Agricultural volatility and investments in children', *The American Economic Review*, 90(2), 399–404.
- JENSEN, S., MKAMA, J. and YA UCHUMI NA MIPANGO YA MAENDELEO, T.W., 1968, *District data, Tanzania, 1967*, Ministry of Economic Affairs and Development Planning.
- JOLLIFFE, D., 1998, 'Skills, schooling, and household income in Ghana', *The World Bank Economic Review*, 12(1), 81–104.
- JOLLIFFE, D., 2004, 'The impact of education in rural Ghana: Examining household labor allocation and returns on and off the farm', *Journal of Development Economics*, 73(1), 287–314.

- KAKWANI, N., SOARES, F.V. and SON, H.H., 2005, 'Conditional cash transfers in African countries', International Policy Centre for Inclusive Growth Working Papers No. 9.
- KING, K., 1984, *The end of educational self-reliance in Tanzania?*, volume 1, Centre of African Studies, Edinburgh University.
- KINUNDA, M.J., 1975, *Experience in Tanzania in identifying and satisfying local needs in Education: A Contribution to the IIEP Seminar on "The Planning of learning arrangements of all kinds for local Communities", 9-17 December 1974*, volume 14, International Institute for Educational Planning.
- KMENTA, J., 1967, 'On estimation of the CES production function', *International Economic Review*, 8, 180–189.
- KRUGER, D.I., 2007, 'Coffee production effects on child labor and schooling in rural Brazil', *Journal of Development Economics*, 82(2), 448–463.
- KUBIK, Z. and MAUREL, M., 2016, 'Climate Variability and Migration: Evidence from Tanzania', *The Journal of Development Studies*.
- LAMBERT, S. and MAGNAC, T., 1997, 'Implicit prices and recursivity of agricultural households' decisions', Crest working paper no. 9731.
- LE, K.T., 2009, 'Shadow wages and shadow income in farmers' labor supply functions', *American Journal of Agricultural Economics*, 91(3), 685–696.
- LEROY, J., 2011, zSCORE06: Stata module to calculate anthropometric z-scores using the 2006 WHO child growth standards. Retrieved 1 February 2017 from <http://www.ifpri.org/staffprofile/jef-leroy>.
- LEVISON, D., ANKER, R., BARGE, S. and ASHRAF, S., 1998, 'Is Child Labour Really Necessary in India's Carpet Industry?', In: R. Anker, S. Barge, S. Rajagopal and M. Joseph (eds.), *Economics of Child Labour in Hazardous Industries of India*, Center for Operations Research and Training.
- LITTLE, A.W., 2008, *Size Matters for EFA*, research Monograph No. 26. Available

- at [http://www.create-rpc.org/pdf\\_documents/PTA26.pdf](http://www.create-rpc.org/pdf_documents/PTA26.pdf) (accessed on September 1,2017).
- LOCKHEED, M.E., JAMISON, T. and LAU, L.J., 1980, 'Farmer education and farm efficiency: A survey', *Economic Development and Cultural Change*, 29(1), 37–76.
- MACKINNON, J.G. and MAGEE, L., 1990, 'Transforming the dependent variable in regression models', *International Economic Review*, 31(2), 315–39.
- MALUCCIO, J., 1998, 'Endogeneity of schooling in the wage function: Evidence from the rural Philippines', *Food Consumption and Nutrition Division Discussion Paper*, 54.
- MARO, P.S. and MLAY, W.F., 1979, 'Decentralization and the organization of space in Tanzania', *Africa*, 49(03), 291–301.
- MARTIN, D., 1988, *Tanzanie: l'invention d'une culture politique*, KARTHALA Editions.
- MCKENZIE, D.J., 2005, 'Measuring inequality with asset indicators', *Journal of Population Economics*, 18(2), 229–260.
- MEYER, B.D. and SULLIVAN, J.X., 2003, 'Measuring the well-being of the poor using income and consumption', National Bureau of Economic Research Working Paper No. 9760.
- MORDUCH, J., 2005, *Consumption smoothing across space: Testing theories of risk-sharing in the ICRISAT study region of South India*, Oxford: Oxford University Press.
- MSAMBICHAKA, L.A., NDULU, B.J. and AMANI, H., 1983, *Agricultural Development in Tanzania: Policy Evolution, Performance and Evaluation: the First Two Decades of Independence*, Friedrich-Ebert-Stiftung.
- MSAMBICHAKA L, M.N. and O, M., 2006, 'Globalization and Development Challenges in Tanzania', *Dar-es-Salaam University Press (DUP), Dar-es-Salaam*.
- NIKOLOSKI, Z., CHRISTIAENSEN, L. and HILL, R., 2018, 'Household shocks and coping mechanism: evidence from Sub-Saharan Africa', In: *Agriculture in Africa : Telling Myths from Facts. Directions in Development - Agriculture and Rural Development*, 123–134, Washington D.C : World Bank 2018.

- NYERERE, J.K., 1967, 'Education for self-reliance', *The Ecumenical Review*, 19(4), 382–403.
- NYERERE, J.K., 1987, 'Ujamma: the basis of African socialism', *The Journal of Pan-African Studies*, 1, 4–11.
- PENCE, K., 2006, 'The Role of Wealth Transformations: An Application to Estimating the Effect of Tax Incentives on Saving', *The B.E. Journal of Economic Analysis & Policy*, 5(1), 1–26.
- RAVALLION, M. and WODON, Q., 2000, 'Does child labor displace child schooling? Evidence on behavioral responses to an enrollment subsidy', *Economic Journal*, 110, C158–C176.
- ROHRBACH, D. and KIRIWAGGULU, J., 2007, 'Commercialization prospects for sorghum and pearl millet in Tanzania', *Journal of SAT Agricultural Research*, 3(1), 1–25.
- ROSENZWEIG, M.R. and UDRY, C., 2014, 'Rainfall forecasts, weather, and wages over the agricultural production cycle', *The American Economic Review*, 104(5), 278–283.
- SABATES, R., WESTBROOK, J. and HERNANDEZ-FERNANDEZ, J., 2011, *The Health and Education Benefits of Universal Primary Education for the Next Generation: Evidence from Tanzania. Research Monograph No. 62 Available at <https://files.eric.ed.gov/fulltext/ED522066.pdf>*, ERIC.
- SAFIR, A., 2009, 'Who leaves, who moves in? The impact of positive and negative income shocks on migration in Senegal', PSE Working Papers No. 2008-76.
- SAHN, D. and STIFEL, D., 2003, 'Exploring alternative measures of welfare in the absence of expenditure data.', *Review of Income and Wealth*, 49(4), 463–489.
- SANDRI, D., VALENZUELA, E., ANDERSON, K. *et al.*, 2007, 'Economic and trade indicators for Asia, 1960 to 2004', Technical report, World Bank.
- SCHADY, N. and ARAUJO, M.C., 2006, 'Cash transfers, conditions, school enrollment, and child work : evidence from a randomized experiment in Ecuador', .
- SHAH, M. and STEINBERG, B.M., 2017, 'Drought of opportunities: Contemporaneous

- and long-term impacts of rainfall shocks on human capital', *Journal of Political Economy*, 125(2), 527–561.
- SHAO, I.F., 1982, 'A neo-colony and its problems during the process of attempting to bring about socialist rural transformation: the case of Tanzania', In: *Taamuli: a Political Science Forum*, volume 12, 29–46.
- SINGH, I., SQUIRE, L. and STRAUSS, J., 1986, *Agricultural Household Models*, chapter 10, 79–83, Washington, D.C.: World Bank.
- SKOUFIAS, E., 1994, 'Using shadow wages to estimate labor supply of agricultural households', *American Journal of Agricultural Economics*, 76(2), 215–227.
- SKOUFIAS, E., PARKER, S.W., BEHRMAN, J.R. and PESSINO, C., 2001, 'Conditional cash Transfers and their impact on child work and schooling: Evidence from the PROGRESA Program in Mexico [with Comments]', *Economía*, 2(1), 45–96.
- TAROZZI, A. and DEATON, A., 2009, 'Using census and survey data to estimate poverty and inequality for small areas', *The Review of Economics and Statistics*, 91(4), 773–792.
- TOWSE, P., KENT, D., OSAKI, F. and KIRUA, N., 2002, 'Non-graduate teacher recruitment and retention: Some factors affecting teacher effectiveness in Tanzania', *Teaching and Teacher Education*, 18(6), 637–652.
- VICENTE-SERRANO, S.M., BEGUERÍA, S. and LÓPEZ-MORENO, J.I., 2010, 'A multiscale drought index sensitive to global warming: the standardized precipitation evapotranspiration index', *Journal of Climate*, 23(7), 1696–1718.
- VICENTE-SERRANO, S.M., BEGUERÍA, S., LORENZO-LACRUZ, J., CAMARERO, J.J., LÓPEZ-MORENO, J.I., AZORIN-MOLINA, C., REVUELTO, J., MORÁN-TEJEDA, E. and SANCHEZ-LORENZO, A., 2012, 'Performance of drought indices for ecological, agricultural, and hydrological applications', *Earth Interactions*, 16(10), 1–27.
- VYAS, S. and KUMARANAYAKE, L., 2006, 'Constructing socio-economic status indices: how to use principal components analysis', *Health policy and planning*, 21(6), 459–468.
- WHEELER, D., 2011, 'Quantifying vulnerability to climate change: implications

for adaptation assistance’, Working Paper 240, Center for Global Development, Washington DC, pages 1-53.

WOOLDRIDGE, J., 2015, ‘Control Function Methods in Applied Econometrics’, *Journal of Human Resources*, 50, 420–445.

WOOLDRIDGE, J.M., 2010, *Econometric Analysis of Cross Section and Panel Data*, second edition, MIT press.

WOOLDRIDGE, J.M., 2014, ‘Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables’, *Journal of Econometrics*, 182(1), 226–234.

# List of Tables

1.1	Agricultural child labor by age . . . . .	25
1.2	Average semi-elasticities of labor: Perfect substitute specification (simulations from Table A1.4 ). . . . .	38
1.3	Average semi-elasticities of labor: Imperfect substitutes specification (simulations from Table A1.5 ). . . . .	39
1.4	Average semi-elasticities of labor: Translog specification (simulations from Table A1.7 ). . . . .	40
1.5	Effect of the number of children $N_c$ on child labor (first stage) . . . . .	41
1.6	Effect of household labor and of household demographic characteristics on the use of inputs . . . . .	44
1.7	Average semi-elasticities of labor, Simulation from OLS and IV estimations.	46
1.8	Average value of semi-elasticities in Tanzanian Shillings . . . . .	48
1.9	Average semi-elasticities with gender heterogeneity (Perfect substitutes) .	51
1.10	Mean of semi-elasticities from perfect specification with age heterogeneity (equation 1.19) . . . . .	52
A1.1	Descriptive statistics: characteristics of agricultural households . . . . .	57
A1.2	Number of days of farm labor per household, by category . . . . .	58
A1.3	Percentage of children without parent living in the household. . . . .	58
A1.4	OLS estimations: perfect substitute specification . . . . .	59
A1.5	OLS estimations: imperfect substitute specification . . . . .	60
A1.6	OLS estimations: Translog specification . . . . .	61
A1.7	OLS estimations: Translog specification with interaction terms between labor and other inputs . . . . .	62



A1.8	IV estimations of the production function . . . . .	66
A1.9	Estimation of the substitution parameter $\rho$ . . . . .	67
A1.10	OLS estimations of the production function for the unrestricted sample (with households observed only at one wave). . . . .	68
A1.11	Effect of the number of children on child labor by gender (first stage) . . .	69
A1.12	Effect of the number of children by age groups on child labor (first stage) .	70
A1.13	Mean of semi-elasticities: perfect specification with age heterogeneity (equation (1.19)). . . . .	71
A1.14	Mean of the semi-elasticity of children aged 11 to 16 and adults aged 17 to 65 . . . . .	72
A1.15	Mean of the semi-elasticity of labor. The instruments are the number of children by age groups $N_{10} \dots N_{15}$ . . . . .	72
A1.16	Comparison of households with and without livestock . . . . .	73
2.1	Villages in Tanzania . . . . .	81
2.2	Age Cohorts . . . . .	87
2.3	Effect of the program on education: $\gamma$ coefficients of equation (2.3) . . . .	93
2.4	Estimations of education on <i>consumption</i> . . . . .	98
2.5	IV estimates of the returns to education by sectors . . . . .	101
2.6	Average marginal effect of education on the probability of working in each sector of activity (mult. logit) . . . . .	104
2.7	The cumulative effect of education . . . . .	106
A2.1	Descriptive statistics from the 2002 census . . . . .	109
A2.2	Descriptive statistics from the LSMS-ISA data . . . . .	110
A2.3	Effect of dwelling characteristics on consumption . . . . .	111
A2.4	Effect of the program on the education level: $\gamma$ coefficients of 2.4 . . . . .	113
A2.5	Effect of the program on the education level from the household survey ( $\gamma$ coefficients of 2.3). . . . .	114
A2.6	Evolution of education attainment by period and region groups . . . . .	115

A2.7 OLS estimates of the returns to education . . . . .	118
A2.8 Effect of education on consumption : IV estimations with education level and primary education level. . . . .	119
A2.9 Effect of education on consumption: IV estimations with partially treated.	119
A2.10 IV estimations at different geographical scales . . . . .	120
A2.11 Effect of education on the wealth index (Income index constructed from the 2002 census data) . . . . .	121
A2.12 IV estimates of the returns to education by sector . . . . .	122
3.1 Effect of positive shocks on children’s activities . . . . .	144
3.2 Effect of Contemporaneous Shocks on Test Scores . . . . .	146
3.3 Effect of shocks during schooling on education decisions. . . . .	147
3.4 Effect of Shocks during schooling on Test Scores . . . . .	149
3.5 Effect of Early Life Shocks on children’s activities(beta coefficients) . . . .	152
3.6 Effect of Early Life Shocks on Schooling Outcomes (beta coefficients) . . .	154
3.7 Effects of productivity shocks on the log of Household Production (beta coefficients). . . . .	155
3.8 Effect of climate and aggregate price variables on days of labor in the field (beta coefficients). . . . .	156
A3.1 Descriptive statistics from the LSMS-ISA data . . . . .	163
A3.2 Descriptive statistics from the Tanzanian Uwezo survey. . . . .	164
A3.3 Effect of Climatic Shocks on Household Production. . . . .	171
A3.4 Effect of aggregate subsistence price variable on days of labor in the field (beta coefficients). . . . .	172
A3.5 Effects of productivity shocks on days of labor in the fields by age groups (beta coefficients). . . . .	173
A3.6 Effects of productivity shocks on the probability of working the week prior the survey. . . . .	173

A3.7	Effects of productivity shocks on the labor intensity by activity (hours over last week).	174
A3.8	Effect of climate and price variations on Z-score of height for age (beta coefficients).	174
A3.9	Effect of positive and negative shocks on children’s activities according to household consumption	175
A3.10	Effect of positive and negative shocks on children’s activities by gender.	176
A3.11	Effect of positive and negative shocks on children’s activities across age-group.	177
A3.12	Effect of the number of shocks according to household consumption.	178
A3.13	Effect of the number of shocks across age-groups.	179
A3.14	Effect of the number of shocks across gender.	179
A3.15	Effect of shocks during school age on Test Scores by age groups	180
A3.16	Effect of shocks during school age on test scores by gender	180
A3.17	Effect of continuous climate and price variables on children’s activities (beta coefficients).	181
A3.18	Effect of continuous climate and price variables on test scores (beta coefficients).	181
A3.19	Effect of future price and rainfalls on education decisions.(beta coefficients).	182
A3.20	Effect of future shocks on education decisions.	183
A3.21	Effect of shocks during school age on test scores standardized by age.	184
A3.22	Effect of continuous climate and price variables on sample selection (beta coefficients).	185
A3.23	Effect of climate and prices on quality of education	186

## List of Figures

1	Evolution of Enrollment Rates in sub-Saharan Africa	3
---	---	---

2	Percentage of the population living in rural areas . . . . .	6
3	Enrollment in school, by work status and age . . . . .	9
A1.1	Distribution of children within the pool of children. . . . .	58
A1.2	Activities in agricultural work . . . . .	63
A1.3	Time spent at working by activities over the last week . . . . .	64
A1.4	Time spent at working (in hours) . . . . .	64
A1.5	Time spent in domestic tasks (in hours) . . . . .	65
A1.6	Labor distribution among main crops . . . . .	65
2.1	Evolution of the education attainment by Age-Cohort. . . . .	87
2.2	$\gamma_t$ coefficients of the interaction between age-cohorts and education level by region in 1967. . . . .	95
A2.1	Distribution of education spendings. . . . .	112
A2.2	Distribution of food spendings. . . . .	112
A2.3	Relationship between the expected consumption $\widehat{lnC}$ and $lnC$ . . . . .	112
A2.4	Access to education in Tanzania . . . . .	116
A2.5	Consumption level of household heads . . . . .	117
A2.6	Evolution of education attainment by region from T0 to T1 according to the education level in T0. . . . .	120
A2.7	Evolution of education attainment by region from T0 to T1 according to the number of schools in 1967. . . . .	120
A2.8	Evolution of the education attainment by district according to the education level in 1967. . . . .	121
A2.9	Education level by district in 1967. . . . .	123
A2.10	Education level by district in 1978. . . . .	123
A2.11	Primary education level (in years) by age cohorts. . . . .	124
A3.1	Children activities by age in rural areas. . . . .	165
A3.2	Percentage of enrolled children by age cohort. . . . .	165

A3.3 Percentage of dropout children by age cohort. . . . .	165
A3.4 Percentage of children who passed the exam by age cohort. . . . .	166
A3.5 Distribution of children who passed the tests . . . . .	167
A3.6 Distribution of SPEI by district . . . . .	168
A3.7 Percentage of land allocated to coffee plantation in Tanzania. . . . .	169
A3.8 Standardized price deviations for the main cash-crop commodities in Tanzania. . . . .	169
A3.9 Standardized price deviations for the main subsistence commodities in Tanzania (deviations from HP trend). . . . .	170

# Table of Contents

<b>Remerciements</b>	<b>i</b>
<b>General Introduction</b>	<b>1</b>
0.1 A particular feature of sub-Saharan African countries: an economy highly dependent on agriculture	4
0.2 Shocks in agriculture	7
0.3 Which coping strategy to adopt?	8
0.4 Child labor, which consequences on education?	9
0.5 Motivations and chapters' summaries	10
0.5.1 Chapter 1 . . . . .	11
0.5.2 Chapter 2 . . . . .	12
0.5.3 Chapter 3 . . . . .	12
0.6 The methodology and the data	13
<b>Chapter 1: Returns to farm child labor in Tanzania</b>	<b>17</b>
1.1 Introduction	19
1.2 Data	22
1.2.1 LSMS-ISA data . . . . .	22
1.2.2 Production . . . . .	23
1.2.3 Child labor in the data . . . . .	24
1.2.4 Sample . . . . .	26
1.3 Production functions and identification	26
1.3.1 Production function . . . . .	27
1.3.2 Child labor productivity . . . . .	29

1.3.3	Inputs . . . . .	31
1.3.4	Identification and specifications . . . . .	32
1.4	Results	37
1.4.1	OLS estimations . . . . .	37
1.4.2	First-stage of IV specifications . . . . .	41
1.4.3	Validity tests for IV specifications . . . . .	42
1.4.4	IV estimations . . . . .	45
1.5	How much should children be compensated?	47
1.6	Heterogeneity in productivity	49
1.6.1	Specification . . . . .	49
1.6.2	Results . . . . .	50
1.7	Conclusion	53
1.7.1	Variables definition . . . . .	55
1.7.2	Additional tables and figures . . . . .	57
1.7.3	Semi-elasticities with function $\log^M$ . . . . .	74

**Chapter 2: The Impact of the Universal Primary Education Program on  
Labor Market Outcomes: Evidence from Tanzania** **75**

2.1	Introduction	77
2.2	The program	79
2.2.1	Historical background and the UPE program. . . . .	79
2.2.2	Data . . . . .	81
2.2.2.1	Data sets . . . . .	81
2.2.2.2	Measuring household wealth . . . . .	82
2.2.3	Measuring intensity of the UPE program . . . . .	85
2.2.4	Sample and Descriptive statistics . . . . .	86
2.3	Empirical strategy	88
2.3.1	Identification strategy . . . . .	89
2.3.2	The impact of the UPE program on education expansion . . . . .	92

2.4	Results	95
2.4.1	The returns to education . . . . .	96
2.4.1.1	Quality bias . . . . .	100
2.4.2	Returns to education by sector of activity . . . . .	101
2.4.2.1	Sample selection bias . . . . .	103
2.4.3	Effect of education on the labor market participation . . . . .	103
2.4.4	Decomposition of the education effects . . . . .	105
2.5	Conclusion	106
2.5.1	Measuring the effect of education by sector of activity with the Heckman selection model . . . . .	108
2.5.2	Sample and statistic descriptives . . . . .	109
2.5.3	Construction of the proxy for consumption . . . . .	111
2.5.4	First stages . . . . .	113
2.5.5	Robustness Checks : . . . . .	118

**Chapter 3: Good or bad timing? The pro-cyclical and counter-cyclical effect of shocks on education investment and on schooling performance. 125**

3.1	Introduction	127
3.2	Framework	129
3.2.1	The effect of shocks which occur during schooling . . . . .	133
3.2.2	The effect of shocks which occur in early life . . . . .	135
3.2.3	Heterogeneity of the effect of shocks . . . . .	136
3.2.3.1	By access to markets . . . . .	136
3.2.3.2	By households characteristics . . . . .	136
3.2.3.3	By shock characteristics . . . . .	137
3.3	Data	137
3.3.1	Education and child labor data . . . . .	138
3.3.2	Climate data . . . . .	140



3.3.3	Price data . . . . .	141
3.4	Estimation strategy and results	143
3.4.1	Effect of current productivity shocks . . . . .	143
3.4.2	Effect of the length and frequency of school-aged . . . . .	146
3.4.3	Effect of early shocks . . . . .	150
3.4.4	Mechanisms . . . . .	155
3.5	Discussion	158
3.6	Conclusion	159
3.7	Construction of the Standard Precipitation Evapotranspiration Index (SPEI)	162
3.8	Descriptive statistics	163
3.9	Estimations of the mechanisms	171
3.9.1	Effect of shocks on production . . . . .	171
3.9.2	Effect of shocks on labor allocation decisions . . . . .	172
3.10	Heterogeneous effects	175
3.10.1	Current shocks . . . . .	175
3.10.2	Cumulative shocks . . . . .	178
3.11	Robustness checks	181



## **Trois essais sur l'éducation en Tanzanie rurale**

Si de nombreux efforts ont été déployés pour améliorer l'accès à l'éducation en Afrique subsaharienne, la proportion d'enfants qui interrompent leur scolarité avant la fin du cycle primaire reste encore très élevée. Pour tenter de répondre à ce défi majeur, cette thèse se propose d'examiner les déterminants de la demande d'éducation dans un environnement rural exposé à de nombreux risques. Le premier chapitre s'intéresse à la place essentielle, et pourtant peu étudiée, que tiennent les coûts d'opportunité du temps des enfants dans les choix éducatifs. Il s'avère dans ce cas nécessaire d'estimer la productivité du travail des enfants afin d'identifier les coûts susceptibles de compromettre leur scolarisation. Afin de mieux appréhender les choix en matière d'éducation, le deuxième chapitre dresse quant à lui une estimation des bénéfices de l'éducation dans un contexte agricole qui se distingue à la fois par un accès limité aux nouvelles technologies et par la prédominance d'exploitations familiales de petites tailles. Enfin, le troisième chapitre interroge les effets des chocs de productivité sur les décisions de scolarisation et sur les performances scolaires des enfants. Deux critères retiennent notre attention, l'âge auquel les enfants sont confrontés à ces chocs et leur intensité. Ce sujet nous semble d'autant plus pertinent que la fréquence de ces chocs ne cesse aujourd'hui d'augmenter. À travers ces trois chapitres centrés sur la Tanzanie, cette thèse offre ainsi un aperçu du rôle des politiques publiques dans la protection et le développement de l'éducation.

## **Three essays on education in rural Tanzania**

Despite numerous investments that have been made to increase access to education in sub-Saharan Africa, a noteworthy share of children drop out of school prior to completing primary education. To address this issue, this thesis examines the factors that drive education decisions in a rural risky environment. The first chapter focuses on one of the core determinant of education investment that has been under-explored, the opportunity costs of education. To identify these costs that can significantly hinder education, we determine children's productivity on the farm and provide an estimate range of the value of one day of child labor. In order to better understand education decisions in rural sub-Saharan Africa, the second chapter assesses the different benefits of education in rural Tanzania, where family farm is the dominant structure in agriculture and where the technology level is low. Finally, the third chapter investigates whether productivity shocks are detrimental to educational achievement and children's cognitive skills by considering two particular aspects, the age at which shocks occur, and the length of shocks. This subject is all the more relevant today when the number of productivity shocks is growing. Throughout these three chapters which focus on rural Tanzania, this thesis provides some insight into the role of public policies in protecting and promoting education.