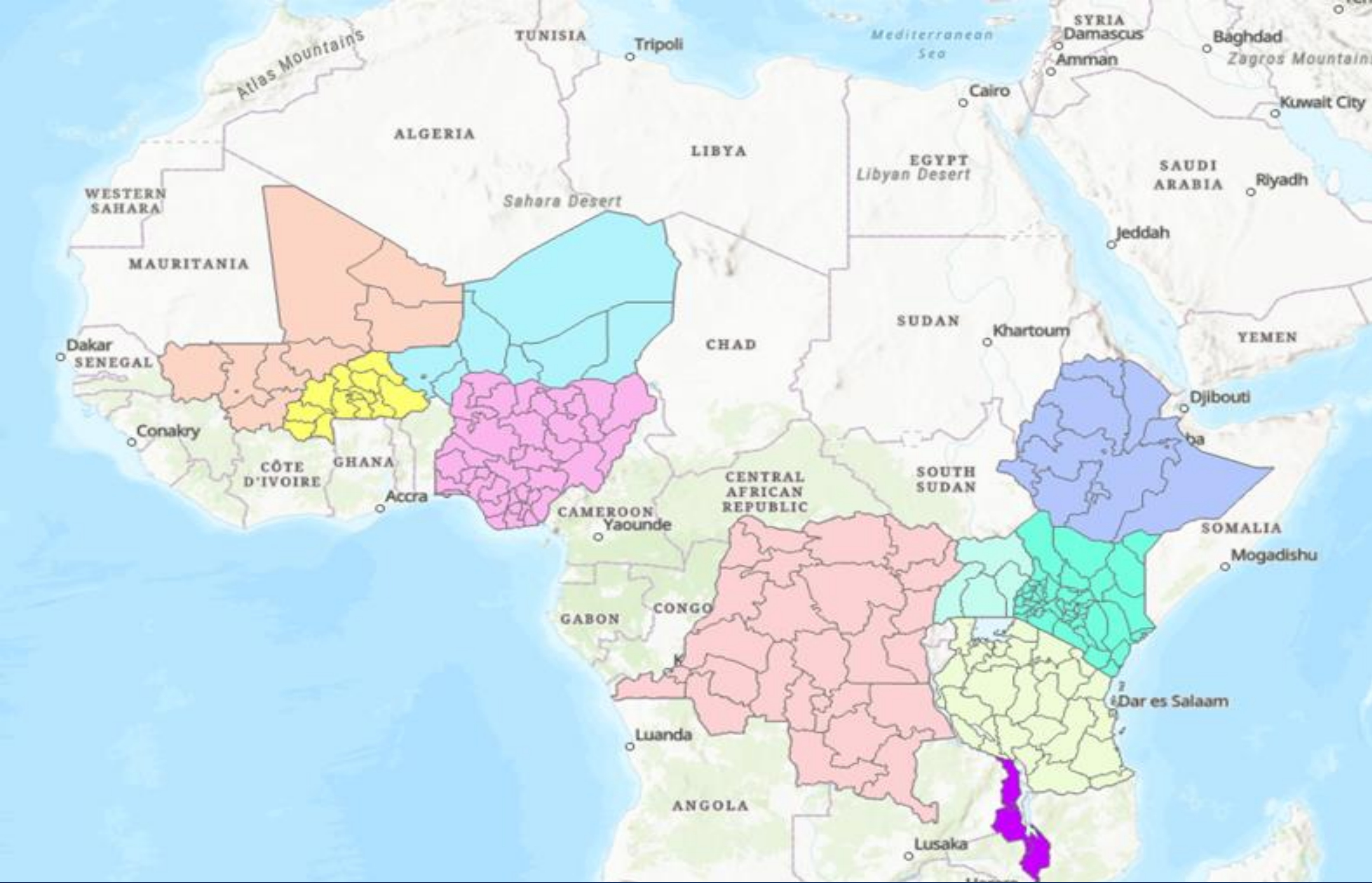


Report 2022/05 | For the University of Oslo and Vista Analyse AS

VISTA ANALYSE

UNIVERSITY OF OSLO

The Impact of Drought on Educational Attainment

An Empirical Cross-Country Study of Sub-Saharan Africa

Master of Philosophy in Economics

Mina Skille Mariussen

Document details

Title	The Impact of Drought on Educational Attainment
Report Number	2022/05
Author	Mina Skille Mariussen Mina Skille Mariussen
Date of completion	11.11.2021
Source front page photo	Arc GIS Pro
Keywords	Educational attainment, drought, climate change, empirical analysis, master thesis

About us

Vista Analysis is a social science consultancy with its main emphasis on economic research, policy analysis and advice, and evaluations. We carry out projects to the highest professional standards, with independence and integrity. Our key thematic areas include climate change, energy, transport, urban planning and welfare issues.

Our employees have high academic credentials and broad experience within consulting. When needed we utilise an extensive network of companies and resource persons nationally and internationally. The company is fully employee-owned.

Foreword

Climate change will probably intensify drought in areas that already are dry. It is important for adaptation and mitigation to assess and understand the impact of drought. Using econometric methods this thesis estimates the impact of drought on educational attainment in Sub-Saharan Africa. The thesis is co-financed by Vista Analyse as part of our research program in environment and development.

March 10, 2022

Haakon Vennemo
Managing partner
Vista Analyse

The Impact of Drought on Educational Attainment

An Empirical Cross-Country Study of Sub-Saharan Africa

Mina Skille Mariussen

Master of Philosophy in Economics

Department of Economics

Submitted: November 2021

2021

The Impact of Drought on Educational Attainment: An Empirical Cross-Country Study of Sub-Saharan Africa

Mina Skille Mariussen

http://www.duo.uio.no/

Trykk: Reprosentralen, Universitetet i Oslo

Abstract

This thesis investigates the effect of drought on educational attainment in ten countries south of Sahara, by using geocoded data on climate and households across a time span of seventy years. Most previous studies have considered smaller geographical areas and periods in isolation and remain inconclusive when it comes to the effect of weather shocks on educational outcomes (Randell & Gray, 2016; Shah, 2017).

Combining climate data sourced from the Peace Research Institute of Oslo with micro-level household data sourced from the Demographic and Health Survey Program was made possible by using the software ArcGIS Pro.

Results show that in the presence of drought, individuals attain around a quarter of a year less education, and the probability of having any education is reduced by roughly three percent. This effect is mainly driven by an adverse effect for females. Females seem to be kept from education in times of drought, while males seem to increase their attendance. This heterogeneous effect on genders seems to be reversed when comparing subsamples of those born in the 1950s and 1990s, indicating a shift in gender norms with regards to education across time.

The results also show that having any education increases the probability of being literate by sixty percent, highlighting the importance of ensuring universal access to education. Having experienced a drought reduces the probability of literacy by around two percent. There seem to be no apparent differences in educational adaptation patterns between rural and urban residents, in the presence of a drought. An event study of the famine in Ethiopia in the 1980s shows that persistent drought incidences can increase the length of education, possibly through altering long-term investment patterns in education.

Preface

I would like to thank my supervisor Jo Thori Lind for valuable guidance throughout the work with this thesis. He has provided invaluable feedback and encouragement thorough the whole process. I also would like to thank Vista Analyse for the opportunity to write the thesis in an inclusive and strongly competent environment. Particularly I thank co-supervisor Haakon Vennemo for useful discussions and feedback on the thesis.

I also thank my previous colleagues at the Development Co-Operation Directorate of the OECD for invaluable discussions on relevant trends and topics within the field of development co-operation, which ultimately led to the topic of this thesis.

Any errors in this thesis are solely my own responsibility.

Table of Contents

List of figures

List of tables

List of tables and figures in appendix

1 Introduction

Climate change is often referred to as the most severe challenge of our time. Rising temperatures increase the frequency of extreme weather events such as floods, droughts, wildfires, and heatwaves across the globe (Masson-Delmotte, et al., 2021).

Agricultural dependent populations in sub-Saharan Africa are especially vulnerable when faced with extreme climatic events, and their adaptability is limited by lack of outside opportunities and relatively low levels of education (Agamile, Dimova, & Golan, 2021; The World Bank, 2021).

Education is one of the most important tools in combating poverty and achieving the Sustainable Development Goals ahead of 2030 (United Nations, 2021). As extreme weather events become more frequent it is therefore increasingly important to better understand the linkages between extreme weather events and educational outcomes. Several previous studies have investigated the effect of extreme weather events on educational attainment, but there exists no consensus on the magnitude and direction of effect (Randell & Gray, 2016; Shah, 2017).

This thesis combines local climatic data on drought occurrences from the Standardized Precipitation-Evapotranspiration Index, with household data across ten countries, sourced from the Demographic and Health Surveys. This is done with the goal of contributing to a better understanding of the effect of extreme weather events on educational attainment. The data is combined using the software ArcGIS Pro, connecting the geocoded data on individual school attendance with data on drought occurrences from 1946 through 2014. The main study includes 244 000 individuals, distributed over ten countries located south of Sahara. This thesis has a particular focus on heterogeneous effects on gender in the presence of drought.

Linkages between drought and educational attainment are investigated using ordinary least squares estimation, and a linear probability model, both with fixed effects and clustered standard errors. The methods employ fixed effects both on small geographical areas, and across time, with the aim of investigating whether a causal relationship between drought and educational attainment exists, and how this has developed across time. Clustered standard errors are employed to obtain heteroscedasticity robust standard errors.

First, the effect of drought on years of education attained is investigated, both on the full sample and on two separate regions based on climatic conditions. Results show that having experienced moderate drought while in school yields a reduction in years of attended education by around a quarter of a year. This effect is mainly driven by an adverse effect for the females in the sample and can indicate that the parents´ valuation of expected future payoff from education is differentiated between the genders. The effect is however reversed when studying in isolation the individuals born in the 1990s, indicating a shift in gender norms with regards to education.

Secondly, the effect of drought on the level of attained education is investigated, finding that individuals who have experienced drought are three percent less likely to ever have attended education. Thirdly, the effects of both education and drought on literacy, are studied separately, showing that education is important to increase literacy, and that individuals are two percent less likely to be literate having experienced a drought.

Lastly, the effect of persistent drought is briefly considered, through an event study of the Ethiopian famine in the 1980s. This is done using synthetic control methodology to construct an artificial control group. The latter analysis indicates that the long-term effects on educational attainment from persistent and severe shocks to household income can increase investments in education.

The statistical software Stata is used in generation of all estimations and calculations that follow.

The thesis is structured in the following way: section 2 presents a background on the topics of the thesis and section 3 presents a brief overview of the current literature on the topic. Section 4 describes the data used to investigate the topic, section 5 describes the methods that are applied, and section 6 presents the findings. Lastly, section 7 discusses possible weaknesses and extensions, before section 8 presents a brief conclusion. The references are found in section 9 and an appendix is included in section 10.

2 Background

This chapter will provide further insight into why the topic of drought occurrences and educational attainment is an interesting and relevant field of study. The chapter starts with background on climate change, agricultural dependence, and drought prevalence, before it briefly elaborates on the importance of education and determinants of educational attainment in developing countries. The chapter ends by introducing the gender aspect of educational attainment and weather shocks.

2.1 Climate change, agricultural dependence, and drought

Rising global surface temperatures are causing extreme weather events, such as droughts, to become more frequent (Masson-Delmotte, et al., 2021). In this thesis the focus will lie on the possible impacts on educational attainment through weather induced shifts in household resources, possibly through changes in agricultural production and food prices. Droughts are one of the extreme weather types which have the most unambiguous implication for agricultural production, and for which there exist available measures across space and time (Beguería, Latorre, Reig, & Vicente-Serrano, 2021). This thesis will therefore focus on the effect of droughts on educational outcomes.

Climate change is likely to yield both more frequent and severe droughts towards the end of the 21st century (Haile, et al., 2020).[1] The study by Haile et al. find that droughts are likely to increase in frequency in Somalia, South Sudan, and Tanzania, while a decrease in drought frequency is expected to occur in Uganda, Kenya and in the highlands of Ethiopia. The different expected patterns of drought prevalence across regions in coming decades makes understanding its possible implications critical to increase adaptability and achieve the SDGs ahead of 2030 (United Nations, 2021).

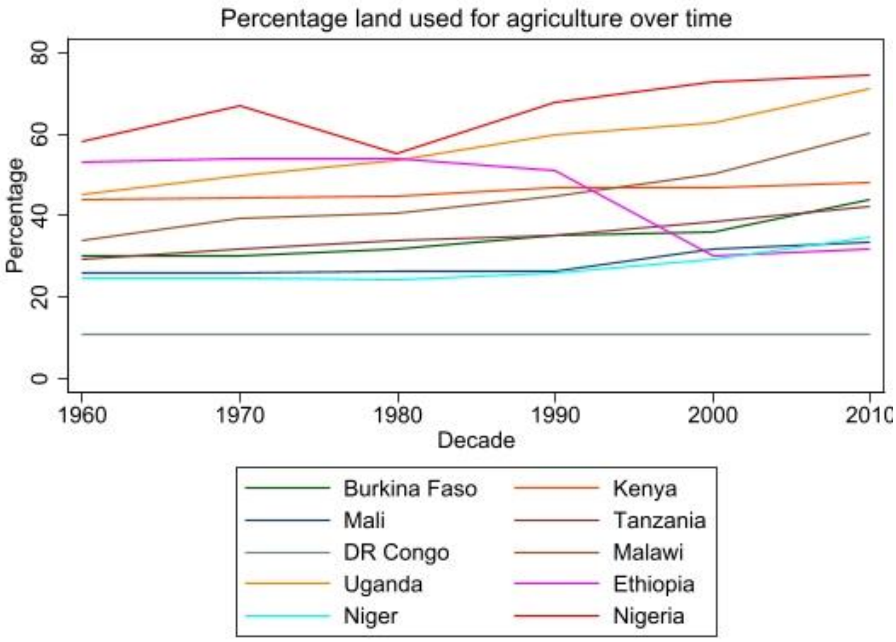
One of the ways droughts can affect educational attainment is through changes in agricultural production, spurring adaptations in the choice of education. In sub-Saharan Africa the percent employed in agriculture was measured to 53 percent in 2019, making it the largest industry in the region (International Labour Organization, ILOSTAT database, 2021). Given the high

[1] Climate change refers to long-term changes in weather and temperatures. Climate change induced by humans commenced as the industrial revolution yielded an increasing need for energy, which largely has been acquired by burning fossil fuels, generating greenhouse gas emissions. At an aggregate level these greenhouse gas emissions trap heat from the sun, causing a long-term shift towards higher temperatures (United Nations, 2021).

proportion of the working population employed in this industry the region is vulnerable to climatic changes affecting the industry.

Figure 2.1 shows the development in land used for agriculture in the ten selected countries from 1960 until 2010. The total average of land used for agriculture in these countries has risen with about ten percentage points during these decades, indicating high dependence on agricultural production.

Figure 2.1: Land used for agriculture over time in selected countries



Note: Development in land used for agriculture in the ten selected countries from 1960 until 2010. The figure is developed using Stata. Figure source: (UNESCO Institute for Statistics, 2021).

Studies have observed that variations in rainfall and temperatures have historically impacted agricultural production more adversely in sub-Saharan Africa since the 1950s, relative to other regions of the world (Barrios, Bazoumana, & Strobl, 2008). The study by Barrios, Bazoumana and Eric find that even small climatic changes is impacting agricultural production negatively in this region due to a lack of endowments and technology to adequately adapt. This thesis will therefore focus primarily on the effect of smaller drought incidences on educational attainment.

2.2 Educational attainment

Sustainable development goal number 4 states "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (United Nations, 2021). Within this lies the goal of universal access to education, ensuring opportunities for all individuals. Reaching this goal would yield value for individuals, but also on the aggregate level a country can benefit by reaching higher long-term economic growth rates if higher educational levels are attained (Crespo Cuaresma & Lutz, 2007). Researchers have found that education is necessary for long-term economic growth, by increasing human capital development, and that the highest level of growth is obtained by ensuring universal access to primary and secondary education (Crespo Cuaresma & Lutz, 2007).

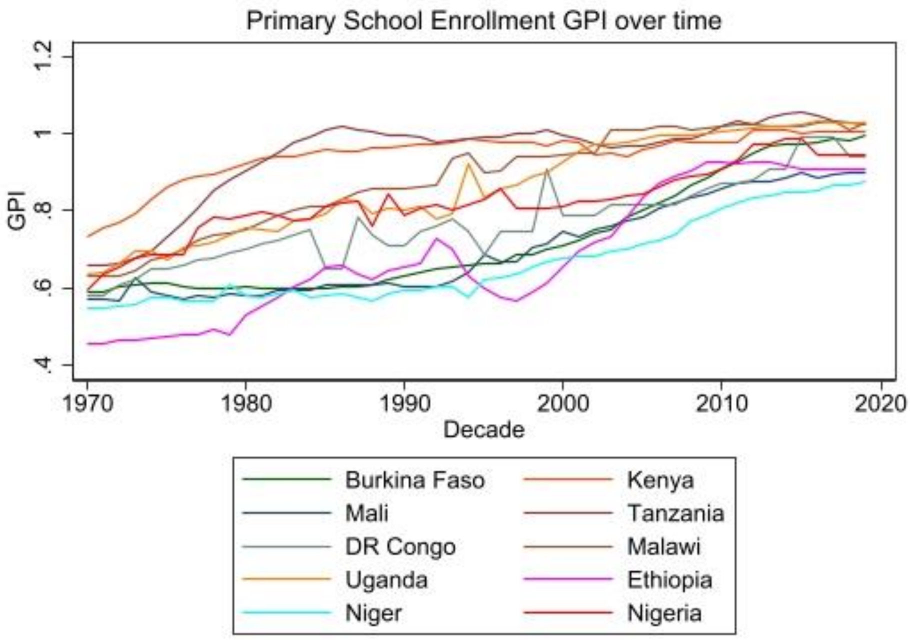
There are several factors influencing educational attainment in developing countries. One important factor is the dependency and employment in agricultural production. In economies where agricultural production is the main source of income, the true cost of education is non-zero despite school enrolment being free of charge (Glory & Nsikak-Abasi, 2013). This is due to the high alternative value children may produce partaking in agricultural production. Glory and Nsikak-Abasi find poverty and family size to be the most important factors influencing child labor rates in agriculture.

Another factor influencing educational enrollment is the distance to the nearest school, which is largely predicted by rural and urban residency (Alesina, Hohmann, Michalopoulos, & Papaioannou, 2021). Further, cultural norms towards education plays an important role. The latter is especially important when evaluating gender aspects of educational attainment. These factors will all be considered as determinants of educational attainment throughout this thesis.

2.3 The gender aspect of weather shocks and educational attainment

In year 2000 world leaders gathered around the Millennium Development Goals (MDS). The audacious goals set out to achieve universal primary education, promote gender equality and empower women, among other goals (United Nations, 2021). With the increased focus on equal access to opportunities the world has seen a rise in education for females. For sub-Saharan Africa the percentage of female pupils in primary education has risen from 45 percent in year 2000 to 48 percent in 2019 (UNESCO Institute for Statistics, 2021). This is in line with the overall trend for the globe.

Figure 2.2: Gender parity index for primary school attendance over time



Note: The gender parity index (GPI) of primary school attendance in the ten selected countries from 1970 to 2020. The figure is developed using Stata. Figure source: (UNESCO Institute for Statistics, 2021)

From figure 2.2 we observe that in 1970 the GPI indicated that for every tenth boy enrolled in school there were only 4.5 girls in Ethiopia and 5.5 girls in Niger. The ratio has however tended towards one for most of the studied countries during the last decades (UNESCO Institute for Statistics, 2021).

Access to education and paid employment outside the household are on average better in urban areas than rural ones. This implies a relaxation of social values and norms that have historically been imposed on females, enabling them to easier achieve greater independence (Tacoli & Satterthwaite, 2013). Therefore, advocating for females´ school attendance has the possibility of inducing great long-term effects. Education helps limit poverty and lower fertility rates, it increases labor force participation and empowerment, which in turn increases the likelihood that future children are sent to school (Murray, Paola, Demeranville, & Hurst, 2010; Tacoli & Satterthwaite, 2013). Given the importance of universal access to education and the changing patterns of drought occurrences it seems increasingly important to understand the mechanisms it might induce with regards to gender disparity in education.

3 Literature review

This chapter will provide an overview of some of the available literature on relevant economic mechanisms induced by droughts and the linkages to educational attainment.

3.1 Drought and mechanisms

Extreme climatic events such as floods and droughts usually affect crops and thereby livelihoods across large regions simultaneously. Consequences of such events are diverse, ranging from increased poverty and forced migration, to changing employment and educational patterns (Agamile, et al., 2021; Barrios, et al., 2008; Bjorkman-Nyqvist, 2013).

A recently published report by several multilateral organizations showed that 60 percent of Africa's population currently suffer from moderate to severe food insecurity, and that this share has been rising since 2014 (FAO, IFAD, UNICEF, WFP and WHO, 2021). Climatic changes, such as the increase in drought frequency, affecting agricultural productivity, and thereby increasing food prices is identified as one of the largest drivers of this development in recent years.

The adverse effects of climate change in the less developed regions are often studied through the lens of crop failures, leading to loss of livelihoods and higher food prices, reducing real income (Hertel & Rosch, 2010). This effect is particularly strong for the very poor, who already live on marginalized plots of land, with limited possibilities of adapting to new conditions (Randell & Gray, 2016). However, higher food prices can lead to an increase in income for producers in regions that are not as adversely affected, generating heterogeneous effects between those that are directly affected and those that are not (Hertel & Rosch, 2010). The effects of weather induced shifts in agricultural productivity can also translate into shifts in labor markets (Agamile et al., 2021).

A study from Uganda find that women are allocated more productive land following droughts (Agamile et al., 2021). Authors argue that the pattern is likely driven by a difference in outside options for the genders. Males have on average higher education and access to more resources than females and are therefore more likely to be more mobile and able to transition to non-agricultural employment, leaving women with more productive plots of land after the drought has subsided.

Another possible effect of extreme weather events is that there is a short-term increase in the need for unskilled labor, such as child labor, for instance in order to replant fields and salvage existing crops (Hertel & Rosch, 2010). Extreme weather events can also reduce the need for unskilled labor, through the loss of crops to cultivate. Therefore, the short-term effects on demand for unskilled labor depend largely on the nature of the weather events and available adaptation mechanisms. A study from Zambia in 2012 show however that the impact of climatic variability on crops alone can reduce the growth rate of GDP by 0.4 percentage points yearly (Thurlow, Zhu, & Xinshen, 2012). Thurlow et al. demonstrate that the impacts on the growth rate of GDP are even larger in years with extreme drought or other extreme weather events.

Long-term effects of such events on labor markets are also ambiguous, but some observations point in the direction of less use of child labor, and more investment in education, as work outside the agricultural sector becomes more attractive when the vulnerability in the sector becomes more apparent (Opiyo, Wasonga, Nyangito, & Schilling, 2015). In conclusion, the nature of the mechanisms induced by extreme weather events depend on the length and severity of the episodes and can therefore have diverse effects on the economy.

3.2 Linkages between droughts and educational attainment

Extreme weather events have direct negative effects on development, leaving countries to rebuild rather than improve on existing structures. If climate change also impairs educational attainment the adverse effect on development might be amplified over time. Despite this research on linkages between weather shocks and educational attainment in developing countries is limited and to a large extent diverging in conclusions (Randell & Gray, 2016).

When considering linkages between drought frequency and educational attainment the notion of human capital investments and alternative value of time becomes important (Shah, 2017). A study of nearly two million children in rural India find evidence that children are more likely to attend school during years of drought (Shah, 2017). The study asserts that a viable alternative to education is in many cases agricultural work or work in the home.

Given that extreme weather events yield less crops to cultivate and less need for child labor, the value of time spent in agricultural production is reduced. This can induce more education for children. This finding is corroborated by a study performed using data from rural

Zimbabwe (Nordstrom & Cotton, 2020). In addition to exposing the above mechanism, the study from Zimbabwe also find that the learning output of children is reduced in the presence of drought, regardless of more time spent in school. This suggests that the adverse effect of excess stress put on families during droughts is offsetting the effect of more time spent in school on learning outcomes.

Contrary to findings in Zimbabwe and India a study of educational choices in rural Ethiopia shows that rainfall above mean long-term levels results in an increased likelihood of more education (Randell & Gray, 2016). The study find that higher levels of rainfall is associated with higher crop productivity, and higher household food consumption, which the authors argue reduces the need for children participating in income generating activities. It seems that less droughts, and thereby stress on family resources increases the likelihood of attaining more education for rural households in Ethiopia.

The literature is also concerned with understanding how effects of weather shocks yield heterogeneous effects on the genders. A study from Uganda finds that drier seasons lower the school enrolment of older females (Bjorkman-Nyqvist, 2013). This study finds no significant effect on school enrollment of younger females, nor on males of any age. The study implies that when stress is asserted on the family, the value of home production for older females increases, while males and young females are kept in school due to the lower need for their labor in agricultural production.

A study of Ethiopia concludes that more humid seasons reduce schooling for all pupils, and that the effect is more negative for males than it is for females (Mani, Hoddinott, & Strauss, 2013). The two studies from Uganda and Ethiopia thus conclude similarly in terms of the relative change in schooling between the genders in response to a drier environment. The studies concludes that more rainfall increases crops, and this is when the value of male production increases relative to the value of traditional female production. Based on the two studies of Mani et al. and Bjorkman-Nyqvist more humid seasons reduces male schooling, while drier seasons reduces female schooling, and the explanation of such effects lie in a change in relative value of their labor in home or agricultural production compared to the value of remaining in school.

The difference in results regarding the overall effect of drought on educational outcomes could be due to the different approaches adopted by researchers but is also likely to be a product of different samples. Zimbabwe, Ethiopia, Uganda and India are different countries, with different adaptation mechanism and differing cultural aspects. Therefore, studying each country in isolation, as is primarily done in the existing literature, is likely to yield a difference of results.

Furthermore, it is difficult in such studies to know if all possible underlying effects influencing educational attainment are controlled for. The problem of omitted variable bias, endogeneity and outliers often limit the validity of results. Therefore, studying the effects across several countries, with many individuals across time, and with climate data for small areas, as done in this thesis, is an interesting approach to contribute to- and expand on the existing literature within this research field.

4 Data

This chapter will provide a description of the data sources used in this thesis, some of the processing of the data, and descriptive statistics of the dataset that is constructed. The main data sources used are the Demographic and Health Surveys (DHS) from ten countries in sub-Saharan Africa, and climatic data sourced from SPEI, made available through the PRIO-GRID structure. Data on first administrative borders is sourced from Humanitarian Data Exchange.

4.1 Demographic and Health Surveys and data management

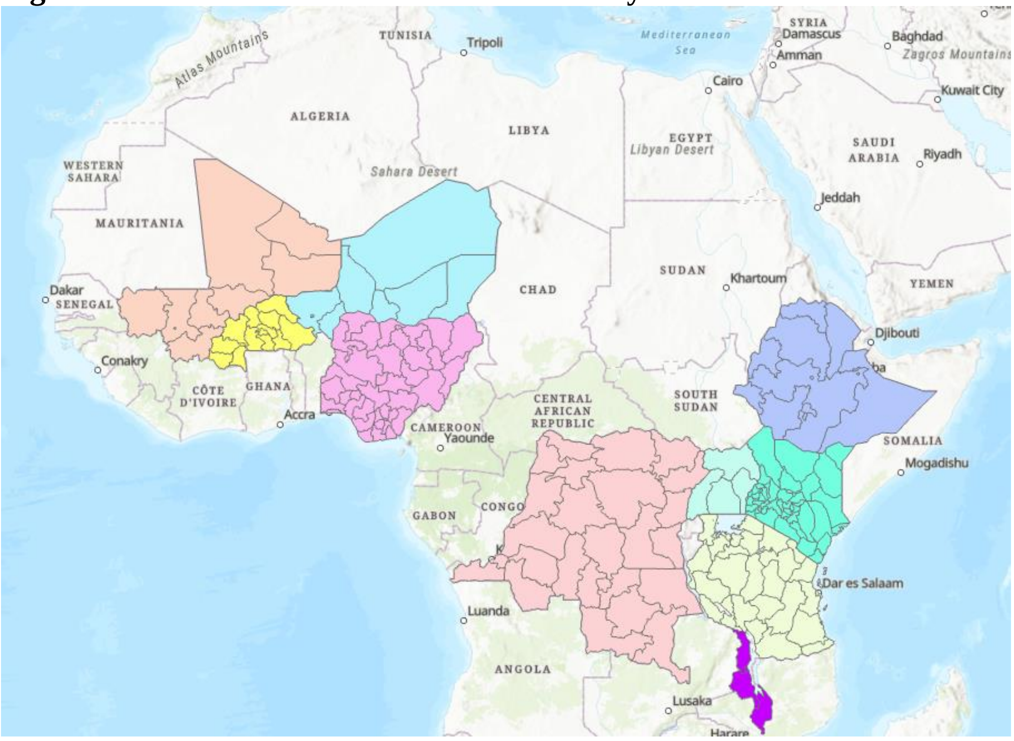
4.1.1 The DHS data

Data on individual educational attainment and other individual characteristics was sourced from the Demographic and Health Survey Program. The DHS conduct surveys by country usually at five-year intervals. The sample size of the surveys range between five and thirty thousand individuals during each survey round. The DHS records households into clusters based on their location. A small village will usually be recorded as one cluster, while in a larger city several clusters will be recorded. The DHS clusters are given GPS coordinates, yielding an accuracy of about fifteen meters of clusters (Burgert, Zachary, & Colston, 2013).

The primary purpose of the DHS is to record health outcomes, but individuals are also asked about living conditions, education, employment, and status within the household, among other characteristics (The DHS Program, 2021).

For use in this thesis access to surveys from ten countries was applied for based on locations of countries and the availability of surveys with geocoded clusters. The countries selected for analysis are Burkina Faso, the Democratic Republic of Congo, Ethiopia, Kenya, Malawi, Mali, Niger, Nigeria, Tanzania, and Uganda. Surveys in the selected countries are conducted in the timespan from 1990 until 2018.

Figure 4.1: The ten countries selected for analysis



Note: The ten selected countries are represented in different colors, and with first administrative levels outlined within each country. The figure is developed using the software ArcGIS Pro.

By merging the cross-sectional datasets from the DHS across both countries and time a large dataset is constructed containing around 750 000 individual observations. Individuals in the dataset are interviewed once, they are born in the year span from 1932 to 2003 and are eligible to attend school between 1937 and 2013. This span in time means we can study droughts and educational attainment for around seventy years.

4.1.2 Data restrictions

Due to the structure of the data several restrictions had to be made. Firstly, as this thesis considers linkages between educational outcomes and experiences of drought during childhood, the sample had to be limited to individuals who reported having always lived in the same place. This is done as the respondents are recorded in the geocoded clusters at the time of the survey, and do not give location of their childhood residency. The possible restrictions this impose on validity of results are investigated in section 6.5.

Secondly, individuals who had missing values or non-logical values for central variables, such as educational attainment, were excluded. Values which are not likely to be correct, such as an individual having attended school for forty years, while being thirty-five years of age are considered as non-logical values.

Thirdly, as the PRIO-GRID only is operationalized until year 2014 individuals who were eligible to end schooling past this year were excluded from the sample. These restrictions limit the sample size to around 244 000 individuals across the ten countries.

After imposing restrictions, two of the selected countries only have one survey to draw from. As individuals of different ages are asked to partake in each survey round, it is possible to study developments across time despite only having one survey for these two countries. The remaining eight countries have from two to four available surveys from which data is sourced.

4.2 PRIO-GRID and SPEI

The climatic data from SPEI and percentage of land used for agriculture, used in this thesis is sourced through the PRIO-GRID map structure.

4.2.1 PRIO-GRID

Data on drought frequency was accessed using a spatial grid structure developed by the Peace Research Institute Oslo (PRIO).[2] The grid structure developed by PRIO consists of a 0.5x0.5-degree grid resolution map covering the globe (Tollefsen, Strand, & Buhaug, 2012).[3]

PRIO-GRID draws on multiple data sources to generate various variables describing drought and other extreme weather events within each grid cell. As the PRIO-GRID draws on several data sources, the availability of variables varies. Some variables are given as snapshots in given years, such as land within the grid cell that is used for agriculture, which is given every tenth year, while other variables, such as drought measures are given every year, for every grid cell (Tollefsen, Bahgat, Nordkvelle, & Buhaug, 2015). PRIO draws on three different sources of drought measurements; the SPI index, the SPEI index and the SPEI Global Drought Monitor, allowing users to identify the measures that is best suited for their analysis.

[2] The purpose of PRIO-GRID is to facilitate easier use of spatial and comparable data worldwide, by drawing on multiple different spatial data sources. The structure compiles data from a range of other spatial data sources and enables the usage of those within the given grid structure (Tollefsen, Strand, & Buhaug, PRIO-GRID: A unified spatial data structure, 2012).

[3] This implies that the cell area covered by grids around equator has an area of roughly 55 x 55 kilometers.

The PRIO-GRID cell structure and data pertaining to the grids are freely available and downloadable from their online portal (Peace Research Institute Oslo, 2021). For use in this thesis variables describing droughts and land used for agriculture within grid cells was downloaded from the online portal and combined with micro-level household data from DHS using ArcGIS Pro.

4.2.2 SPEI global dataset

The Standardized Precipitation-Evapotranspiration Index (SPEI) is one of several indexes measuring droughts across the globe (Tollefsen, et al., 2012).[4] The SPEI contains monthly values for drought measurements across the globe on a 0.5x0.5-degree resolution. The values of the SPEI are given as deviations from normal, long-term conditions in each locality, which makes it is well suited to compare adaptations to droughts across regions (Begueria, Vicente-Serrano, & Angulo-Martinez, 2010).

This implies that the same value of the index would present the same deviation from the mean in two different localities, reflecting relative conditions rather than absolute ones (Begueria, et al., 2010). For use in this thesis, it is relevant that the SPEI reports deviation from mean conditions, as what is studied are mechanisms of adaptation to shocks. Therefore, other indexes, reporting absolute measures are inferior to the SPEI for the given purpose.[5]

The values of the SPEI typically range between 2,5 and -2,5, although the index in theory could reach infinity in either direction. The drought variable captures the severity of drought within the rain season of the given grid cell.[6]

[4] The SPEI is made available to the public through a global dataset named the SPEIbase (Begueria, Vicente-Serrano, & Angulo-Martinez, 2010). The SPEI uses monthly climatic water balance as measurement, effectively measuring the difference between monthly rainfall and monthly potential evapotranspiration.

[5] Note that the similar SPI differ from the SPEI by not including temperatures in its measures. This makes the SPEI better suited to capture drought incidences as consequence of climate changes, considering temperatures are changing in response to climate change (Begueria, Vicente-Serrano, & Angulo-Martinez, 2010) (Masson-Delmotte, et al., 2021) .

[6] The rain season is defined as the three consecutive months where it on average rained the most during a year within a grid cell. Data on rain seasons within each cell used in the calculation of the SPEIbase is obtained from Schneider et.al (Schneider, et al., 2015).

Figure 4.2: SPEI drought values and their possible implications

Dryness	Possible Impacts	Standardized Precipitation-Evapotranspiration Index (SPEI)	Moderate drought dummy	Extreme drought dummy
Moderate Drought	Some damage to crops	-0.8 to -1.2	1	0
Severe Drought	Crop or pasture losses likely Water shortages common	-1.3 to -1.5	1	0
Extreme Drought	Major crop losses Widespread water shortages or restrictions	-1.6 to -1.9	1	1
Exceptional Drought	Exceptional and widespread crop losses Shortages of water creating water emergencies	-2.0 or less	1	1

Note: The values of the SPEI that captures dry conditions and their possible impacts on crops and livelihoods, along with a representation of the two drought dummy variables. Figure source: (National Drought Mitigation Center, 2021)

In this thesis two threshold values of drought prevalence are used to study educational adaptations to droughts. One set of dummy variables capturing the existence of at least a moderate drought (yielding 1 for SPEI values below -0,8) within each grid cell in each year is generated, while a second set of dummy variables capture the existence of at least an extreme drought (yielding 1 for SPEI values below -1,6) is also generated.[7]

This construction implies that if the dummy variable for moderate drought is one, while the extreme drought dummy is zero, the drought experienced ranges from moderate to severe, while if both dummy variables are one the drought ranges from moderate to exceptional drought. Consequently, the extreme drought dummy variable cannot be one unless the moderate drought dummy variable is also one.

A dataset with the two drought dummy variables is merged with the DHS clusters, where the common variable is the grid cell of each observation.[8] The process of combining the two geocoded datasets is performed using ArcGIS and is described in more detail below.

[7] The two threshold values are generated using the variable from the SPEIbase named "*droughtend_speibase*" in the PRIO-GRID dataset (Tollefsen, Bahgat, Nordkvelle, & Buhaug, 2015) (Begueria, Vicente-Serrano, & Angulo-Martinez, 2010). The SPEIbase and "*droughtend_speibase*" was selected to generate the threshold values based on several criteria. The variable is recorded for the longest period of time within the PRIO-GRID structure, 1949 to 2014, and it includes temperatures and measurements are deviations from long-term mean values.

[8] To be able to combine data from the SPEI with DHS data the data is reshaped separately for each dummy threshold such that dummies for each year 1946 to 2014 are kept on the x-axis while keeping the grid numbering on the y-axis of the dataset. The two resulting datasets are then merged yielding one dataset with dummies showing whether there was existence of moderate and extreme drought for each year within each grid cell in the PRIO-GRID structure.

4.3 Humanitarian Data Exchange

In addition to the use of data from the DHS and the SPEI, data on first administrative levels (ADM1) is sourced from the Humanitarian Data Exchange (Humanitarian Data Exchange , 2021). The inclusion of data on administrative levels is done to facilitate clustering of the standard errors at this level to obtain heteroscedasticity robust standard errors. This method is further explained in chapter 5. Shapefiles with administrative borders for each country are downloaded from the Humanitarian Data Exchange website and merged with the other data sources using ArcGIS Pro. For each of the ten countries the selected shapefile is the most recent version available.

This is done despite the DHS also recording corresponding ADM1 levels for each interviewed cluster. This is done to ensure grid cells pertain to one ADM1 level, rather than evolving with name changes of ADM1 levels between DHS rounds.

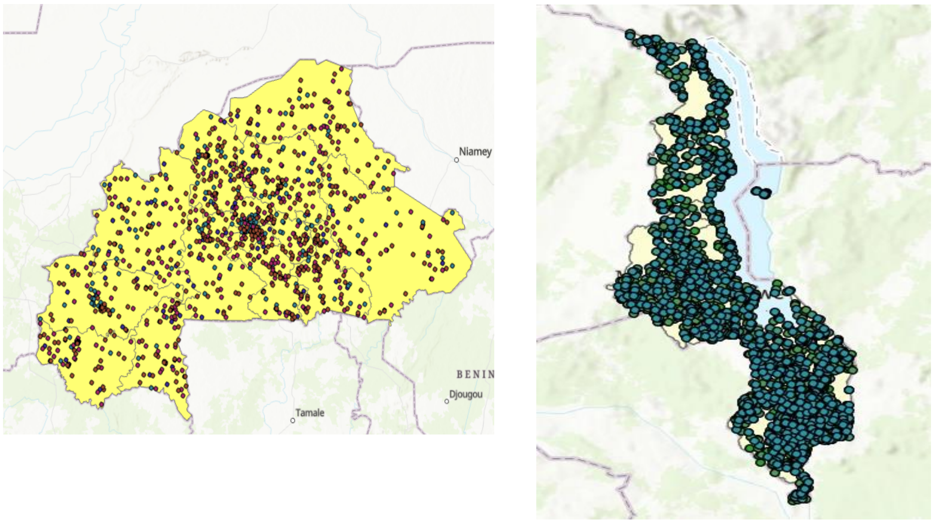
4.4 Combining the datasets using ArcGIS Pro

The software ArcGIS Pro is used in combining the DHS clusters with PRIO-GRID and ADM1 levels. All three data sources are available in shapefiles, which are representable using the software. The DHS clusters with longitude and latitude coordinates from each survey were uploaded and merged into one shapefile, before uploading the PRIO-GRID structure onto the software. The polygons of PRIO-GRID were joined with the points of the DHS clusters using a spatial join tool, capturing the points lying within each grid polygon, and generating a new layer of data with the combined information (ArcGIS Pro, 2021). The clusters that lie on polygon borders are excluded in this procedure, as including them in an arbitrary grid cells could potentially lead to biased estimates.

The new data layer containing numbers for identification, clusters and grid cells, among other variables, is then exportable to Excel format, which in turn is imported into the statistical software Stata. This is then merged with the combined DHS dataset, using cluster number, year of interview and country of residence as the common identifiers. The merge is performed using a one-to-many merge in Stata, as one grid cell contains several individual records in the DHS dataset.

ArcGIS Pro is also used in the same way to combine DHS clusters with first administrative levels (ADM1). Figure 4.3 illustrates how this looks like for Burkina Faso and Malawi.

Figure 4.3: Combining ADM1 with DHS clusters using ArcGIS Pro



Note: The DHS clusters from different survey rounds are represented by different colored points, while the within country borders are those of the ADM1 levels. Burkina Faso is represented to the left, while Malawi is represented to the right. The figures are developed using ArcGIS Pro.

4.5 Construction of two variables for drought experiences

In this thesis individuals' experiences of drought occurrences are recorded in the period in which individuals are eligible to attend school. The eligibility to attend school is defined as the years in which individuals are five to fifteen years of age. The selection of this span of years is based on the mean and standard deviation of the main dependent variable "Years of education". The variable has a mean of 4,5 years, and a standard deviation of 4,6 years across the whole sample. Therefore, the selected year span of eligibility to attend school is deemed reasonable.

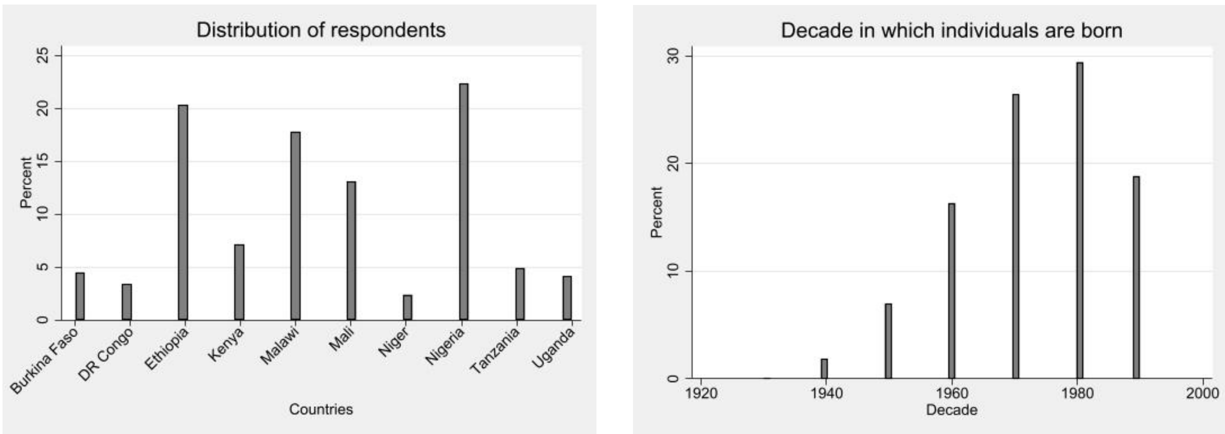
For each individual two variables describing the years they were five and fifteen years of age are constructed. This is then merged with the two sets of threshold values for drought, for which two dummy variables for the years from 1946 to 2014 exists, describing whether there was moderate and extreme drought in each of the years. Based on these yearly dummy variables for drought occurrence and the range of years where individuals are eligible to attend school, individuals are either affected by a drought during these years, or they are not. The result are two dummy variables named “Moderate drought while in school” and “Extreme drought while in school” for each individual observation in the dataset.

Using a year span of eligibility of school attendance, rather than a record of actual school attendance has the benefit of capturing individuals who might have left school short-term due to drought experiences. Using the year span of when individuals attended school would not have captured this short-term effect.

4.6 Descriptive statistics of constructed dataset

The resulting dataset on which analysis is based consists of around 244 000 individual respondents, interviewed in years between 1990 and 2018 and attending school between 1937 and 2003.

Figure 4.4: Distribution of respondents by country residence and decade born



Note: The percent of individual observations by country of residence and decade in which respondents were born. The figures are developed using Stata.

From figure 4.4 we observe that Ethiopia, Malawi and Nigeria are the three largest countries in the sample, while Niger, the Democratic Republic of Congo and Uganda are the three countries with fewest observations. The majority of respondents, almost thirty percent, were born in the 1980s. Only 236 individual respondents were born in the 1930s and due to the imposing restrictions of completion of school within 2014, which is the last recording by PRIO-GRID, no one is born in the 21st century. The sample size of those born in the earliest decades is much smaller than the samples of those born in more recent decades. This fact, together with the acknowledgement that some of the ones born in the 1930s and 1940s can have difficulties remembering the exact years they attended school, mean the accuracy of estimates based on these samples will be lower than the estimates for the more recent samples.

The DHS primarily ask women about health and family planning, and thus females are overrepresented in the surveys. In the constructed dataset, sixty-five percent of respondents are female, and thirty-five percent are male. Therefore, it is expected that the results are more accurate for females than they are for males due to the larger sample size. Also, given that the effect on males and females of experiencing drought is heterogenous, the overall effect will be driven mostly by the effect experienced by females.

Table 4.1: Gender frequency

Gender	Frequency	Percent
Male	84 064	34.47
Female	159 796	65.53
Total	243 860	100.00

4.6.1 Drought experiences

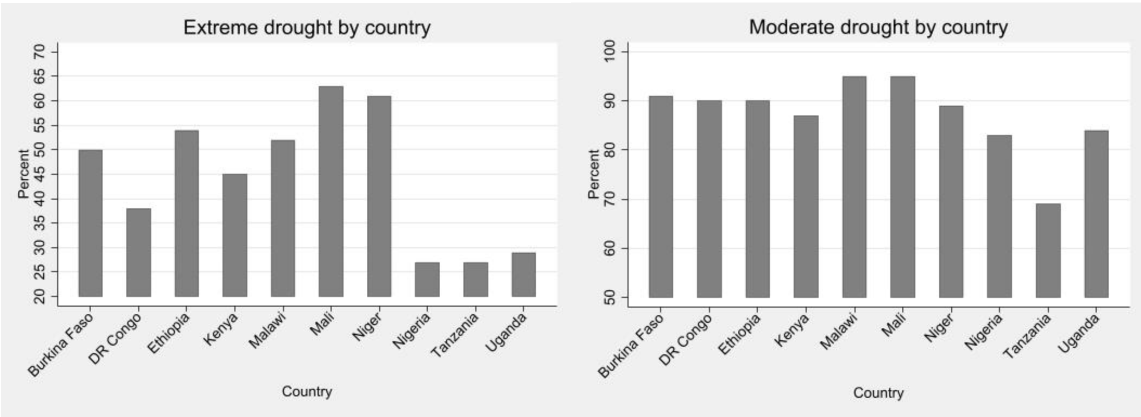
Due to the way the drought dummy variables are constructed, more individuals have experienced moderate drought than extreme drought. Table 4.2 shows these percentages in the constructed dataset.

Table 4.2: Frequency of extreme and moderate drought during eligibility to attend school

	Extreme drought		Moderate drought	
	Frequency	**Percent**	**Frequency**	**Percent**
No	133 734	54.84	27 096	11.11
Yes	110 126	45.16	216 764	88.89
Total	243 860	100.00	243 860	100.00

From table 4.2 we observe that around fifty percent of individuals have experienced extreme drought during their years when eligible to attend school. In contrast, almost ninety percent of individuals have experienced moderate drought in the years when eligible to attend school. Naturally, there are differences in these percentages across the ten countries in the sample, as. represented in figure 4.5.

Figure 4.5: Frequency of drought during eligibility to attend school by country



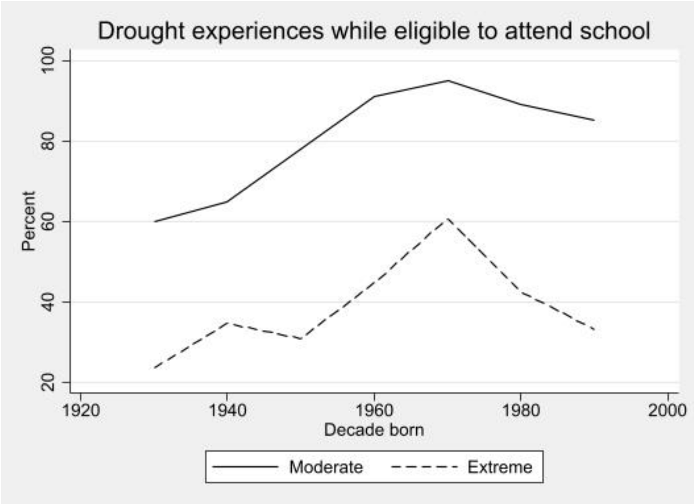
Note: The percent of individuals who have experienced extreme and moderate drought during eligibility to attend school, within each of the selected countries. The figures are developed using Stata.

From figure 4.5 we observe that the highest percentage has experienced extreme drought in Mali and Niger, while in Nigeria, Tanzania and Uganda fewest individuals have experienced drought while in school.

The percentage of those that have experienced moderate drought is on average higher and more even across the ten countries than that of extreme drought. We also observe that Tanzania and Uganda are two of the countries where there have been relatively few droughts while individuals were eligible to attend school.

Across time there are differences in the percentage of the sample who experienced drought while eligible to attend school. Figure 4.6 show how this percentage has developed over time.

Figure 4.6: Percent of individuals who experienced drought while eligible to attend school



Note: The percent of individuals who have experienced moderate and extreme drought while eligible to attend school, over time in the sample. Calculated by the mean level for those born in the same decade. The figure is developed using Stata.

From figure 4.6 we observe that more individuals have experienced moderate drought, than extreme drought while eligible to attend school. For those born around 1970, almost sixty percent experienced extreme drought while eligible to attend school. The estimated percentage for the earliest decades is less reliable than the more recent decades, due to the larger sample size.

4.6.2 Rural and urban areas

Acknowledging the existence of a rural-urban divide in educational attainment, it is interesting to consider how this is represented in the constructed dataset (Sumida & Kawata, 2021). From table 4.3 we observe the distribution of the sample that reside in urban and rural areas.

Table 4.3: Distribution of urban and rural residence among respondents

Respondents by residence

Residence	**Frequency**	**Percent**
Urban	57 993	23.78
Rural	185 867	76.22
Total	243 860	100.00

From table 4.3 we observe that seventy-six percent reside in rural residences, while twenty-four percent live in urban areas. Considering the divide in educational attainment by residence, table 4.4 represents the distribution of educational attainment according to type of residence.

Table 4.4: Percentage distribution of educational attainment

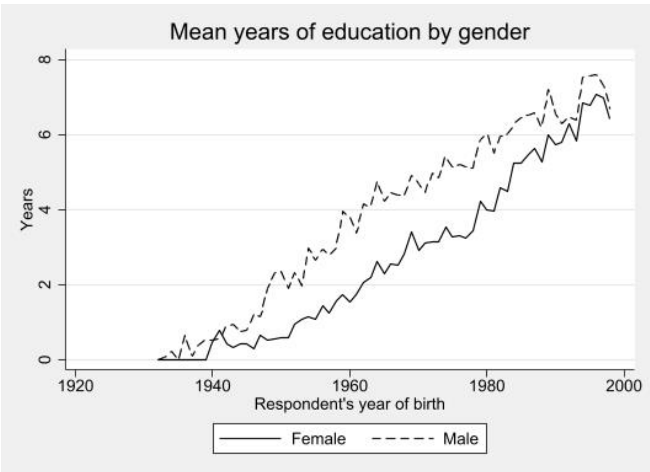
Education	Urban percent	Rural percent
Any education	79.01	54.65
Completed primary	53.57	24.67
Secondary or higher	41.87	14.53

From table 4.4 we observe that the educational attainment is much lower in rural residences, compared to urban ones. In urban areas seventy-nine percent report having any education, compared to only fifty-five percent in rural areas. The difference is even larger when considering secondary or higher education, with urban and rural residents reporting respectively forty-two and fifteen percent.

4.6.3 Educational attainment

The DHS records several variables related to educational attainment. Individuals report their highest year of education, they report how many years they have attended education and their highest level of completed education. As this thesis investigates the possibility that individuals alter their educational patterns in the short term, the main variable of interest is the variable that records how many years individuals have attended education. Figure 4.7 shows the development in this variable over time.

Figure 4.7: Mean years of education reported across time

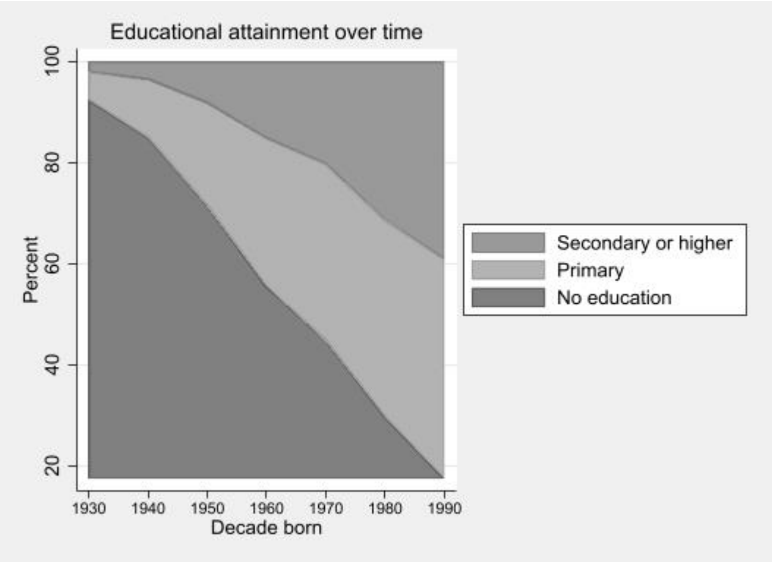


Note: Mean years of education over time, calculated by the mean reported by those born in the same year for each gender separately. The figure is developed using Stata.

From figure 4.7 we observe that the mean years of education has been trending upwards since 1930. The mean years of education has risen remarkably from 1930s until the 1990s, from an overall mean of 0.3 to almost 6.5 years of education. We also observe that there is a gender gap in the mean years of reported education but that this gap is narrowing.

From the dataset constructed we note changes across time also with respect to the highest level of education attained. Figure 4.8 displays this overall trend, with the highest level of educational attainment reported divided into three categories.

Figure 4.8: Trends in the highest level of education attained over time



Note: Developments in educational attainment over time, calculated by the percentages of those born in the same decade. The figure is developed using Stata.

From figure 4.8 we observe that the share reporting having no education declines over the decades, while those reporting having any primary education and secondary or higher education increase, indicating that most individuals now start education.

4.6.4 Literacy

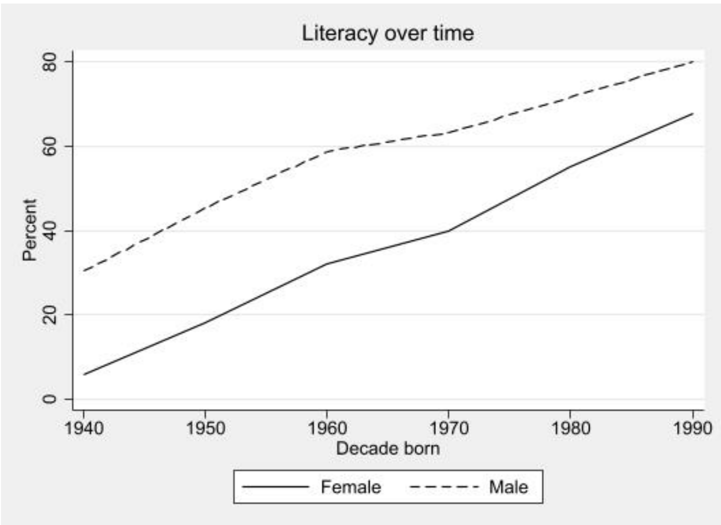
Another variable that will be considered in the analysis is the level of literacy, as an indicator of the quality of education. Table 4.5 shows the level of literacy in the whole sample, by gender. A more comprehensive explanation on the construction of this variable will be given in section 6.3.

Table 4.5: Percentage of literacy by gender

Literacy	Female	Male
Not literal	54.66	35.79
Literal	45.34	64.21
Total	100.00	100.00

From table 4.5 we observe that more males than women in the total sample are literal. However, there has been changes to this percentage across time. Figure 4.9 shows this evolution.

Figure 4.9: Evolution of literacy by gender



Note: Percent of literal individuals over time, for the genders separately. Calculated by the percentage of literacy for those born in the same decade. The figure is developed using Stata.

From figure 4.9 we observe that that the percentage of females who are literate lie consistently below the one for males, but that the gap is narrowing.

As this thesis considers the effect on drought on educational attainment, it is interesting to note how educational attainment correlates with literacy in the sample. In table 4.6 four variables, representing the highest level of education attained by individuals, and the correlation to literacy is presented.

Table 4.6: Correlation of educational levels and literacy

Level of education	Literal
No education	-0.291
Primary education	0.195
Secondary education	0.354
Higher education	0.391

From table 4.6 we observe that all levels of education have a positive correlation with literacy, while having no education correlates inversely, indicating that education is an important tool to increase literacy.

5 Empirical strategies

The following section will describe the empirical strategies that are used to study linkages between drought prevalence and educational attainment. In a first step ordinary least squares estimation is used to evaluate the effect of drought on years of education obtained. Secondly, a linear probability model is used to evaluate the effect on the level of education attained and to evaluate the effect of drought on literacy. Both strategies involve fixed effects and clustering to obtain heteroskedasticity robust standard errors. Lastly, the method of synthetic controls is applied to consider the effect of persistent drought on educational attainment.

5.1 Ordinary Least Squares estimation with fixed effects

The method of ordinary least squares (OLS) is applied to study the effect of drought on years of attained education. The dataset that is created consists of individuals who either have experienced drought during years eligible to attend school, or they have not. Having this data structure, it is theoretically possible to estimate the causal effect of drought experiences on educational attainment, conditional on several criteria.

For OLS estimation to yield unbiased and consistent estimators, the Gauss-Markov assumptions must be fulfilled. If the assumptions are fulfilled the resulting least squares estimators, are those that have the lowest possible sampling variance (Stock & Watson, 2015). The following section mentions assumptions that are central to OLS estimation and how they relate to the dataset and study design used, before elaborating on the use of fixed effects in the estimation.

A central underlying assumption for OLS to yield unbiased estimates is that the conditional mean of the error term should be zero, meaning that there is no relationship between the explanatory variables and the error term. If this assumption holds, there should exist no omitted variable bias influencing the estimated coefficients of the model.

If there is no presence of omitted variable bias, the results of the estimations are close to the true population parameters, and not influenced by other factors that are not controlled for. When studying the effect of drought on educational attainment, it is important to note that there exist several factors which affects educational attainment, that could potentially bias the results. In the analysis, some control variables are included, but the majority of omitted variable bias is excluded due to the use of entity fixed effects on small geographical areas, as well as time fixed effects.

There are several factors that can influence educational attainment and be correlated with drought prevalence. Cultural aspects, norms, conflict, and attitudes towards female education are only some of the factors that can play an important role in determining the level of educational attainment reached by the individual. These are factors which are difficult to control for in the analysis and may result in omitted variable bias affecting the validity of the estimate if not controlled for using alternative methods.

As observations had to be restricted to certain individuals who have not moved residence since childhood, the constructed dataset violates the OLS assumption concerning the randomness of the sample. However, an estimation of the magnitude of this potential bias is performed in section 6.5, indicating that this does not introduce major bias.

Another assumption states that the model should be linear in its parameters (Stock & Watson, 2015). In the specification used no coefficients enter multiplicatively or quadratically, but explanatory variables are interacted. For the constructed dataset, the presence of perfect multi-collinearity is tested for by considering the correlation of explanatory variables. This test indicates that the explanatory variables do not perfectly predict one another, and the presence of perfect multi-collinearity is excluded.

The assumption of no heteroscedasticity and no serial autocorrelation in the sample is important for OLS to yield precise estimates. Heteroscedasticity is observed when the variance of the error terms vary systematically with the predictor and serial correlation means that the covariance of the error terms is non-zero, meaning that the error terms can be predicted by one another (Stock & Watson, 2015). To test for heteroscedasticity in the dataset the White test is applied to a regression without fixed effects or clustered standard errors. This test indicates that the data is heteroscedastic with high significance.

The presence of spatial correlation between treated and untreated groups in the constructed dataset and heteroscedastic error terms presents a possible issue for the estimation, if not addressed. The adverse effects from this are limited using clustered standard errors in estimation, producing heteroscedasticity robust standard errors. This method is described more in depth in section 5.4.

As omitted variable bias is one of the most common issues when trying to expose causal relationships, several methods have been developed to limit its adverse effects (Stock & Watson, 2015). One way of addressing issues caused by omitted variable bias is to use fixed effects on entity and/or time. An alternative to fixed effects estimation is the usage of random effects estimation. To rule out the usage of random effects in the estimation, the Hausman test is performed using Stata. Prior to running the Hausman test two estimations are performed, one with fixed effects and one with random effects. The Hausman test then compares the estimated coefficients from both models and rejects the usage of random effects if the estimated coefficients are sufficiently different from one another. The Hausman test confirmed that fixed effects perform better than random effect for estimation with the constructed dataset.

In this thesis effects are fixed at relatively small geographical areas, as well as across time, implicitly studying variation within each selected area for individuals that are born in the same decade. Employing both entity- and time fixed effects omits much of the potential omitted variable bias from variables that vary across areas and across time (Stock & Watson, 2015). Using fixed effects on small geographical areas, within each country, mean cross-country differences are also accounted for.

Employing both entity and time fixed effects is appropriate when there are omitted variables that are constant over time but vary across areas, and omitted variables that are constant across areas but vary over time. In this analysis it can be argued that both types of omitted variables exist. Two examples of such variables are that cultural aspects vary between areas, and attitudes towards universal education has varied across time.

When employing entity fixed effects one is subtracting the mean values for each area from individual observations in the given area, and therefore studying deviations from mean values. Fixed effects estimation on area assumes that everyone in the same area has been subject to the same incidences and trends, allowing us to detect the effect of someone being affected by drought.

Similarly, to the intuition of the entity fixed model, employing time fixed effects allows each time period, or decade, to have its own intercept. This allows to study effects that vary over time, but not across areas. The intercept of different decades means we subtract the mean of each decade, leaving us to study the effect for those born in the same decade.

The functional form with both entity- and time fixed effects is illustrated in equation 5.1 where α_i is the intercept for each area studied, and λ_t is the intercept for each time period studied.

$$\text{Years of education}_{it} = \beta_0 + \beta_1 \text{Drought}_{1,it} + \cdots + \beta_k X_{k,it} + \alpha_i + \lambda_t + \mu_{it} \tag{5.1}$$

In equation 5.1 β_1 is the main coefficient of interest in this thesis, while β_k represents the coefficients of the control variables. This is the equation that is used to study the effect of drought on years of attained education.

The statistical software Stata is used to run the model with fixed effects using the constructed dataset. Employing this method means that there needs to exist variation in experiences of drought within the fixed areas for those born in the same decade, for meaningful results to emerge. If everyone in an area, born in the same decade, experience drought, the model will not provide meaningful results.

Omitted variable bias can still occur due to area specific characteristics that vary over time, for which the fixed effect method does not control. For this to be an issue, there must be such omitted variables that are correlated to both drought occurrence and educational attainment.

5.2 Linear probability model on the level of education attained

In addition to recording years of attained education, the DHS also records the highest level of education attained by the individual. To study the effect of drought on different levels of attained education a linear probability model with OLS, fixed effects and clustering is applied.

The DHS records the highest level of educational attainment in five categories: no education, incomplete primary education, complete primary, incomplete secondary, complete secondary and higher education.

We have reason to believe that a drought occurrence affects children of different ages differently (Bjorkman-Nyqvist, 2013; Mani, et al., 2013). This indicates that there is a non-linearity in the effect of drought on different levels of education. This implies that regression analysis with a categorical dependent variable can yield biased results. To resolve this issue a linear probability model, with a binary dependent variable, is applied on three different levels of education separately. The linear probability model is chosen over the ordered logistic and ordered probit models due to its clear interpretation and the simple inclusion of fixed effects. A brief demonstration of the performance of the ordered logistic and ordered probit models is presented in appendix 10.1 table A10.

The linear probability model entails that the dependent variable takes the value of either one or zero, and the coefficients of the explanatory variables are interpreted as the change in probability that the dependent variable takes the value one if the explanatory variable increase by one unit. The three dependent variables used to study the effect of drought on educational level attained are: having any education, having completed primary education, and having completed secondary or higher education. The functional form of this model is presented by equation 5.2.

$$\mathrm{P}(\mathrm{Educational\ level}_{it} = 1) = \beta_0 + \beta_1 \mathrm{Drought}_{it} + \beta_2 X_{1,it} + \cdots + \beta_k X_{k,it} + \alpha_i + \lambda_t + \mu_{it} \quad (5.2)$$

5.3 Control variables included in the analysis

Control variables are added to the estimation, to reduce the omitted variable bias that is not eliminated by fixed effects, as much as possible. This is done to resolve issues related to endogeneity and identification issues of the model. However, including more control variables can increase the variance of the main variable of interest, leaving us with a less accurate estimate (Stock & Watson, 2015). Therefore, the control variables that are added to the estimation are thoroughly assessed before inclusion, with respect to the possibility that they cause omitted variable bias.

An example of a variable that can be correlated to both drought occurrence and educational obtainment, while varying over time and within locations, is the wealth quintile to which an individual belongs. This also varies for individuals born in the same decade.

Wealth of a family is a strong predictor of educational attainment of its children (Knight & Shi, 1996). At the same time wealth could be strongly correlated with drought occurrence in agricultural dependent areas. The wealth of families can vary over time within the areas and is therefore included as a control variable in the specification. However, it is important to note that controlling for wealth can pose a potential issue, as some of the educational adaptation patterns to drought are expected to happen due to a shift in family income, which potentially also affects wealth. This means income could have been the dependent variable studied, and therefore should not be used as an explanatory variable.

However, the DHS does not record income directly but calculates a wealth index that is relative for each survey based on the household’s ownership of assets such as refrigerator, TV and access to water sources, that are indicative of their living conditions (Rutstein & Johnson, 2004). As living conditions can be an important predictor of educational attainment, the wealth quintile is included as a control variable despite being related to income and the potential issues this might cause (Filmer & Pritchett, 2001).

Other control variables are added as they pose great predictive power on educational attainment. One of these variables is the type of residence of individuals, whether they live in urban or rural areas. As demonstrated in section 4.6.2 educational attainment vary greatly between rural and urban areas. There exist several explanations of this difference in educational attainment between rural and urban areas. In rural areas the distance to school is on average longer, infrastructure is generally less developed, and cultural norms towards especially female education tends to be more conservative for rural residents (Knight & Shi, 1996; Linard, et al., 2012).

Furthermore, child labor and agricultural production is highly correlated in many countries. In 2010 it was estimated that 60 percent of all child laborers work in agriculture, and the majority of these are unpaid family members (International Labour Organization, 2010). With this being an issue mainly in rural areas, where agricultural production is higher, it adds to the divide in educational attainment between rural and urban areas.

Another control variable that is included in the specification is the gender of individuals. Due to the gender disparity in education gender proves an important control variable. This becomes especially important when studying heterogeneous effects of drought on gender.

Religion is included as another control variable, as this can be indicative of cultural norms, which can vary both across and within areas. The composition of religion in an area could also vary from decade to decade, although families tend to keep to the same religion over time.

The land used for agriculture within the area is another control variable, sourced from PRIO-GRID and the European Space Agency Global Cover Portal (ESA Globcover Project, 2021). It is possible to infer that the higher this variable is, the more dependent on agriculture individuals within this area are. This control variable is recorded in the year the individuals were eligible to attend school, given that this was during or after the 1960s. For individuals attending school prior to 1960 the value recorded in 1960 is inferred. This is done as the first recording of the variable is in 1960.

The age of respondents is also included as a control variable. This is important as wealth and religion are variables which are recorded at the time of the survey. Individuals who are in their 30s might differ from individuals who are in their 40s, with respect to for instance accumulated wealth. In addition, the decade in which individuals were born is included as a time fixed effect, as we know there has been great development with regards to educational attainment over time.

Information is given in the year individuals partook in the survey. To limit bias from this, years in which individuals partook is divided into ranges of five years and included as a time fixed effect. However, as information of variables such as religion and wealth are given at the time of the survey in which individuals partook, assumptions must be made regarding the relevance of those variables in predicting conditions during childhood. It is possible to argue that low social mobility makes relative wealth quintile today predictive of conditions in the past (Alesina, et al., 2021). The choice has therefore been made to include it in the analysis regardless of its shortcomings. Furthermore, individuals seldomly move between religions. Therefore, religion today is used to control for the outcome in the past.

Controlling for age, survey period and decade in which individuals are born mean we have a strong correlation between these three variables, and the signs and magnitudes of these estimated coefficients should therefore not be interpreted. The decision is made to include these three variables as controls as it is believed that they together control for much of potential omitted variable bias.

The main coefficients of interest are the coefficients of the two drought threshold variables. Therefore, the coefficients of several of the control variables are not displayed in the regression tables. The full regression tables and dataset can be obtained at request. A summary of the variables and their functions in the regressions is found in appendix 10.2 table A9.

5.4 Clustered standard errors to resolve spatial correlation in the error terms

As climatic shocks are determined by large weather systems, they are likely to affect large areas simultaneously. This means that the grid cells experiencing droughts are not at random within an area, implying existence of spatial correlation in the error terms.

This issue can be solved using clustered standard errors. By employing clustering on the standard errors, we allow the error terms to be arbitrarily correlated withing the clusters. If the appropriate clustering level is employed the method yield heteroscedasticity robust standard errors. This can matter greatly for the statistical significance of the independent variables, as the size of standard errors usually increase in size, compared with estimations without clustering.

Clustering on first administrative level (ADM1) is likely to be the best option to resolve the spatial correlation in the error terms, as these are larger areas than the grid cells and might resolve the issue of drought prevalence over larger areas. Clustering could also be done on country level, but for the larger countries in the sample drought is less likely to affect the entire country at the same time.

To be able to cluster the grid cells within the ADM1 levels sourced from Humanitarian Data Exchange some grid cells need to be split. PRIO-GRID cells do not remain within country, nor within administrative borders, generating an overlap between grid cells and other borders. The overlap means grid cells are not nested within administrative levels. To solve this overlap a new set of grid cells are generated, by combining a unique country code, the code for ADM1 levels and the PRIO-GRID cell code. The resulting, new grid cells pertain to one country and one ADM1 level only. The procedure of generating new grid cells is explained in more detail in appendix 10.2. Depending on analysis, the resulting clusters based on ADM1 levels range between 160 and 190 clusters.

5.5 Synthetic control method to evaluate impact of persistent drought

The main focus of this thesis is on evaluating the effects of short-term drought experiences on educational attainment. However, as droughts can span across several years, the possible effects this can cause is also investigated briefly using a famine in Ethiopia as an event study. To evaluate whether mean years of education developed differently in Ethiopia compared to other countries, who did not experience this famine, the method of synthetic controls is applied (Abadie & Gardeazabal, 2003).

The method extends on the methodology of difference-in-difference estimation, where the trends of two groups, one affected and one unaffected, are compared before and after a shock or an intervention to determine its effects. A central assumption of the difference-in-difference methodology is that the treated and control groups have the same trend in the outcome variable prior to the shock (Stock & Watson, 2015). Otherwise, the observed difference-in-difference estimator can be biased by other underlying trends, and not be a product of the shock alone.

As it is often challenging to find two groups where the trend is equal prior to the shock, the method of synthetic controls poses an advantage. The method derived by Abadie and Gardeazabal entails creating an artificial control group by combining selected untreated groups, in a weighted average, so that the artificial control group resemble the trend of the treated group prior to the shock as much as possible (Abadie & Gardeazabal, 2003).

For the dataset at hand, the synthetic control method is used to study the effect of the infamous famine in Ethiopia during the 1980s. The other nine countries in the constructed dataset are used to generate the synthetic control group, in a weighted average, to lay as close to the pre-famine Ethiopian trend. This entails that some of the nine countries are used more than others in the construction of the weighted average, but some information is sourced from all nine countries.

6 Results

This section will give an overview of the results of the analysis. Several linkages between drought and educational outcomes are investigated, including the effect on years of education, the level of education, the effect on literacy and heterogeneous effects of drought. An analysis is also conducted to assess the effects of droughts across time.
Lastly the section provides a brief event study of the Ethiopian famine, and a robustness analysis of the models that are used to investigate the relations between education and drought.

In all the sections of this chapter both threshold values of drought are included. Due to the way these variables are constructed, anyone who experiences extreme drought has also experienced moderate drought. Therefore, in all instances where the extreme drought dummy variable is one, the moderate drought dummy variable is also one.

When interpreting the coefficients of the drought dummies it is important to consider the change in control group between the two dummy variables. The coefficient of the moderate drought dummy is interpreted as the effect of going from no drought, to experiencing a moderate drought or worse, while the coefficient of the extreme drought dummy is the effect of going from a situation of moderate drought to even worse drought. Therefore, the effect of going from no drought to a situation with extreme drought is interpreted as the sum of the two estimated drought coefficients.

Using fixed effects, the interpretation of the coefficients is the estimated effect within each grid cell area for all specifications, and within each decade when decade born is included as a fixed effect.

6.1 Years of education and drought

A correlation matrix of the two threshold values of drought with years of education is generated and can be found in appendix 10.1 table A8. The matrix shows negative and significant correlation of both types of drought and education. However, this effect might be driven by other factors, and do not indicate a causal relationship between the variables. To investigate whether a significant relationship exists the methods described in chapter 5 are applied.

The following four subsections all investigate the relationship between drought and years of attained education.

6.1.1 Effect of drought on years of education attained for the full sample

In table 6.1 five specifications are used to study the links between drought and years of attained education. All specifications in this, and the following regression tables, include fixed effects on the constructed grid cell level, and clustering on first administrative level to obtain heteroscedasticity robust standard errors. Significant results at minimum 95 percent confidence intervals for key variables of interest are outlined in bold in all tables to follow.

Table 6.1: The effect of drought on years of education attained

Years of education attained	(1)	(2)	(3)	(4)	(5)
Moderate Drought	-0.124	-0.0993	**-0.252*****	**-0.249*****	**-0.232*****
	(0.130)	(0.125)	(0.0590)	(0.0593)	(0.0571)
Extreme Drought	-0.0265	-0.00866	0.0327	0.0371	0.0481
	(0.177)	(0.164)	(0.0591)	(0.0626)	(0.0582)
Female		-1.443***	-1.440***	-1.451***	-1.420***
		(0.0884)	(0.0880)	(0.0888)	(0.0933)
Urban		2.848***	2.677***	2.687***	1.604***
		(0.211)	(0.206)	(0.191)	(0.163)
Constant	4.600***	4.839***	1.941***	1.199**	-0.214
	(0.0909)	(0.111)	(0.201)	(0.374)	(0.444)
N	243718	243718	243718	233690	228872
R^2	0.000	0.091	0.162	0.180	**0.238**
F (1, 188)	1.16	0.68	9.77	8.65	7.28
P-value	0.2834	0.4101	**0.0021****	**0.0037*****	**0.0077****
Religion	No	No	No	Yes	Yes
Wealth	No	No	No	No	Yes
Agriculture	No	No	No	No	Yes
Decade born	No	No	Yes	Yes	Yes
Age	No	No	No	Yes	Yes
Year interviewed	No	No	No	No	Yes

Note: Standard errors are clustered at the ADM1 level and are heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level.
The F-test is testing the significance of the sum of the two drought dummy variables. Degrees of freedom are given in parenthesis, and the p-value in the following row. For specification 5 the F-statistic has degrees of freedom F (1, 157).
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

From regression table 6.1 we observe that experiencing moderate drought has a negative and significant effect on years of education attained in specification 3-5. Depending on selected specification, individuals who have experienced moderate drought spend from a fifth of a year, to a quarter of a year, less in school than those who have not experienced drought.

The added effect of going from a moderate drought to an extreme drought is small, and not statistically different from zero. The F-test is testing the significance of the sum of the two drought dummies. From this we observe that the effect of an extreme drought compared to a situation of no drought is significantly different from zero in specification 3-5. This indicates that some of the effect on education detected in the coefficient of the moderate drought variable is driven by the effect of an extreme drought.

The coefficients of the drought variables change as religion, wealth, agriculture, age and period interviewed are included as control variables. This indicate that the introduced control variables explain some of the variation in years of attained education. The included R^2 shows how much of the variation within each fixed area and time period in educational attainment is explained by the model. However, including more explanatory variables will always increase this R^2 value (Stock & Watson, 2015). Therefore, not much emphasis will be put on interpreting this value with respect to the goodness of fit of the model.

6.1.2 Effects of drought on subsets of countries

Several studies have found that in areas that frequently experience drought, people have adopted both short- and long-term adaption strategies, to mitigate the adverse effects of crop loss (Opiyo, et al., 2015). To investigate whether this implies that effects of droughts on educational outcomes are different depending on long-term climatic conditions, the dataset is divided into two subsets.

In table 6.2, panel A the relatively humid countries of Nigeria and the Democratic Republic of Congo are investigated and in panel B, the drier countries of Burkina Faso, Mali, Niger, Ethiopia, and Kenya are studied (Ellis & Galvin, 1994).

Table 6.2: The effect of drought in dry and humid countries

Years of education attained

Panel A: humid countries	(1)	(2)	(3)	(4)	(5)
Moderate Drought	-0.0735	-0.0636	-0.0589	-0.0599	-0.0408
	(0.187)	(0.187)	(0.0929)	(0.0908)	(0.0880)
Extreme Drought	**-0.957*** **	**-0.952*** **	**-0.202***	**-0.201***	**-0.200***
	(0.166)	(0.155)	(0.0990)	(0.0989)	(0.0903)
Female		-2.241***	-2.366***	-2.345***	-2.259***
		(0.141)	(0.140)	(0.145)	(0.155)
Urban		2.588***	2.386***	2.414***	0.820***
		(0.243)	(0.242)	(0.254)	(0.111)
Constant	6.208***	6.862***	2.345***	2.526***	0.289
	(0.162)	(0.191)	(0.339)	(0.671)	(0.473)
N	62714	62714	62714	62314	61611
R2	0.008	0.108	0.176	0.196	**0.291**
F (1, 62)	18.20	20.14	4.40	4.38	4.01
P-value	**0.0001*** **	**0.0000*** **	**0.0401***	**0.0404***	**0.0497***

Panel B: Dry countries	(1)	(2)	(3)	(4)	(5)
Moderate Drought	0.130	0.139	**-0.335*** **	**-0.311*** **	**-0.301*** **
	(0.147)	(0.138)	(0.0840)	(0.0843)	(0.0811)
Extreme Drought	-0.115	-0.0811	0.0307	0.0199	0.0255
	(0.0790)	(0.0727)	(0.0456)	(0.0433)	(0.0434)
Female		-1.262***	-1.184***	-1.142***	-1.154***
		(0.135)	(0.121)	(0.109)	(0.114)
Urban		3.360***	3.216***	3.044***	2.150***
		(0.425)	(0.411)	(0.338)	(0.295)
Constant	3.279***	3.178***	0.932***	0.0719	-0.622
	(0.137)	(0.144)	(0.232)	(0.537)	(0.476)
N	115625	115625	115625	115158	114682
R^2	0.000	0.117	0.175	0.191	**0.225**
F (1, 87)	0.01	0.14	15.13	13.04	12.25
P-value	0.9249	0.7057	**0.0002*** **	**0.0005*** **	**0.0007*** **
Religion	No	No	No	Yes	Yes
Wealth	No	No	No	No	Yes
Agriculture	No	No	No	No	Yes
Decade born	No	No	Yes	Yes	Yes
Age	No	No	No	Yes	Yes
Year interviewed	No	No	No	No	Yes
Gender interactions	No	No	No	No	No

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level.
The F-test is testing the significance of the sum of the two drought dummy variables. Degrees of freedom are given in parenthesis, and the p-value in the following row.
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In the relatively humid countries of DR Congo and Nigeria going from experiencing a moderate drought to experiencing an extreme drought has a negative and significant effect on years of education. Specification 3-5 estimate this effect to about a quarter of a year less time spent in school.

For the drier countries it is the moderate drought dummy that is negative and significant. Going from a situation with no drought, to a situation with drought implies individuals attain about a third of a year less schooling. The effect of going from a moderate drought to an extreme drought is not significant in explaining years of education.

The F-test of significance of the sum of the two drought coefficients is included in both panels, showing that the effect of an extreme drought is different from zero in the majority of specifications.

We observe that for the drier countries, the effect of a drought is pronounced enough to adapt educational patterns if the drought is moderate or worse, while for the more humid countries educational patterns are only adapted if a drought is categorized as extreme or worse. The observed difference between the dry and more humid countries can be explained by the construction of the SPEI as a deviation from mean long-term conditions in each locality, as described more in depth in section 4.2.2. Therefore, it is natural that the more humid countries only feel the impact of the drought enough to adapt educational patterns when the drought is categorized as extreme. For the drier countries, a drought of any kind is enough of a deviation from their already dry long-term mean to adapt.

To investigate whether a drought has fewer consequences in already drier areas, due to better adaptation mechanisms, an absolute measure of droughts should be applied. The results do however indicate that the variables used for analysis work appropriately for the purpose of this analysis, as they confirm that the construction of the SPEI values make sense.

Choosing selected countries for analysis is interesting but can risk threatening the validity of results by testing multiple hypothesis. The more hypothesizes are tested, the more likely it is to observe a significant result by random, as the chance that the result lie outside of the 95 percent confidence interval increases (Stock & Watson, 2015). Stated differently, if twenty hypotheses are tested, one is likely to yield significant results due to the acceptance of a significance level of 95 percent. This is worth noting as we assess the significance of repeated testing.

6.1.3 Heterogeneous effects on gender

The results from regression table 6.1 and 6.2 show that a drought does affect the years of education attained negatively from a third of a year to a quarter of a year, depending on sample and specification. From the descriptive statistics we noted that there are differences in educational attainment between the genders, and previous research has shown that there exists a gendered aspect of adaptation to weather shocks (Bjorkman-Nyqvist, 2013; Hertel & Rosch, 2010). Therefore, it is interesting investigate whether there are heterogeneous effects on years of education attained with respect to gender in our sample.

This is done by introducing an interaction term describing whether an individual is male and has experienced drought while eligible to attend school. In regression table 6.3 all specifications include control variables for age, wealth, agricultural land, type of residence, religion and year interviewed. All specifications have entity fixed effects on grid cell level and time fixed effects on decade in which individuals were born.

The tree panels in table 6.3 contain the same samples as presented in section 6.1 and 6.2, and the interaction terms are included for male. This allows us also to study the effect of drought for females, presented by the estimated coefficient of the drought variables. For males, the total effect of a drought on education is the sum of the estimated coefficient of the interaction variable and the corresponding estimated coefficient of the drought variable. A table with all four interaction terms is included in appendix 10.1 table A1.

Table 6.3: Heterogeneous effects on gender by sample

Years of education attained

Panel A: Full sample	(1) Moderate	(2) Extreme	(3) Moderate & Extreme
Extreme Drought		0.0315	0.0759
		(0.0800)	(0.0768)
Male x Extreme Drought		-0.0490	-0.0820
		(0.100)	(0.101)
Moderate Drought	**-0.278*****		**-0.311*****
	(0.0678)		(0.0677)
Male x Moderate Drought	0.168		**0.208***
	(0.103)		(0.104)
Female	-1.271***	-1.443***	-1.271***
	(0.126)	(0.0973)	(0.126)
Urban	1.604***	1.604***	1.604***
	(0.162)	(0.163)	(0.162)
Constant	-0.277	-0.341	-0.284
	(0.454)	(0.440)	(0.452)
N	228872	228872	228872
R^2	0.238	0.238	0.238
F (1, 157)			1.03
P-value			0.3115

Panel B: Humid countries	(1) Moderate	(2) Extreme	(3) Moderate & Extreme
Extreme Drought		**-0.454*****	**-0.427****
		(0.125)	(0.124)
Male x Extreme Drought		**0.764*****	**0.697*****
		(0.197)	(0.200)
Moderate Drought	**-0.264***		-0.143
	(0.115)		(0.115)
Male x Moderate Drought	**0.563****		0.330
	(0.172)		(0.174)
Female	-1.774***	-2.041***	-1.778***
	(0.215)	(0.174)	(0.215)
Urban	0.822***	0.823***	0.823***
	(0.111)	(0.110)	(0.110)
Constant	-0.0569	0.118	-0.0456
	(0.466)	(0.454)	(0.459)
N	61611	61611	61611
R^2	0.291	0.292	0.292
F (1, 62)			18.90
P-value			**0.0001*****

Panel C: Dry countries	(1) Moderate	(2) Extreme	(3) Moderate & Extreme
Extreme Drought		-0.0456	0.0167
		(0.0485)	(0.0480)
Male x Extreme Drought		0.0551	0.0210
		(0.0980)	(0.108)
Moderate Drought	**-0.403*****		**-0.408*****
	(0.0933)		(0.0896)
Male x Moderate Drought	0.234		0.221
	(0.126)		(0.142)

Text box 6.1.

The regression design in table 6.3 implies the following with respect to the effect of an extreme drought:

The effect on females:

$$\beta_{ModerateDrought} + \beta_{ExtremeDrought} \quad (1)$$

The effect on males:

$$\beta_{ModerateDrought} + \beta_{ExtremeDrought} + \beta_{Male\ x\ ModerateDrought} + \beta_{Male\ x\ ExtremeDrought} \quad (2)$$

By combining (1) and (2) we note that if heterogeneous effects of extreme drought exist:

$$\beta_{Male\ x\ ModerateDrought} + \beta_{Male\ x\ ExtremeDrought} \neq 0 \ (3)$$

The F-test in column 3 is testing equation (3)

Female	-0.940***	-1.126***	-0.940***
	(0.143)	(0.130)	(0.143)
Urban	2.150***	2.150***	2.150***
	(0.295)	(0.295)	(0.295)
Constant	-0.698	-0.801	-0.697
	(0.478)	(0.464)	(0.476)
N	114682	114682	114682
R2	0.225	0.224	0.225
F (1, 87)			3.34
P-value			0.0710

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications include controls for gender, age, wealth, agricultural land, type of residence and period in which they were interviewed. All specifications have fixed effects grid level and clustered at ADM1.
The F-test is testing the significance of the sum of the two drought dummy variables. Degrees of freedom are given in parenthesis, and the p-value in the following row.
Significance levels: * p<0.05, ** p<0.01, *** p<0.001

In specification 1 we observe the two types of drought, and the interaction term for male and moderate drought. In specification 2 we observe the same, only for extreme drought, and in specification number 3 we see all four terms describing drought and the interactions with male. Below each panel is the F-statistic and its p-value displayed in column 3. This F-test indicates that the effect of extreme drought has heterogeneous effects on gender in the relatively humid countries, but not in the two other samples.

From panel A, the full sample, specification number 3 we observe that there are clear heterogeneous effects of moderate drought between the genders. The effect for a male of experiencing moderate drought is interpreted as the sum of the coefficients of moderate drought and the interaction term. Following the reasoning from text box 6.1. we know that heterogeneous effects between the genders exist in specification 1 and 2 if the coefficient on the interaction term with male is significant. From specification 3, we note that relative to females, males attend a quarter of a year more education in the presence of drought.

Running the model with interaction terms for females, as shown in appendix 10.1, confirm that there is an overall small, negative effect for males, but this effect is not significant. For females the negative effect is significant and negative, indicating that they attain around a third of a year less education in the full sample.

From panel B, we observe that the heterogeneous effects between males and females are larger in the more humid countries, with males experiencing a moderate drought attaining more than half a year more education relative to females. The heterogeneous effect of experiencing an extreme drought is even larger. Given that the effect for males of experiencing a drought is likely to be positive, the negative effect of drought on years of education detected in the previous sections is likely to be driven by an adverse effect for females.

From panel C, the drier countries, we observe that the effect of experiencing a moderate drought, relative to no drought is about the same as it was for the humid countries when the drought went from moderate to extreme. Although the estimated coefficient of the male interaction term is positive, while the drought coefficient is negative, the heterogenous effect on gender, represented by the F-statistic is not statistically significant.

The difference in adaptation to the drought levels between humid and dry countries is likely to be explained by the same reasoning as was applied in section 6.1.2 regarding the construction of the SPEI values. It is however interesting to note that all three samples seem to confirm that droughts can contribute to widen the gender gap in education, and that the heterogeneous effect seems to be largest in the more humid countries.

The larger heterogenous effect in the humid countries can possibly be explained by the role of males in agricultural production. The humid countries have possibly not adopted as many adaptation mechanisms as the dry countries, meaning that crops are more affected, and male labor less needed in times of drought. Detecting the true drivers behind this pattern is an avenue of further research.

6.1.4 Heterogeneous effects on place of residence

From the descriptive statistics we observed that there is a large difference in educational attainment between rural and urban areas, with seventy-nine percent in urban areas having any education, compared to fifty-four percent in the rural areas.

Knowing that the type of residence is an important predictor of educational attainment it is interesting to study whether there exists any difference in educational adaptation to droughts between the types of residences. Rural residents are likely to be more affected by droughts, due to a higher dependence on agriculture (Thurlow, et al., 2012). However, Thurlow et al. showed that also residents in urban areas can feel the impact of droughts through both a rise in food prices and because many rely on small-scale plots of land for subsistence farming.

Table 6.4 shows the analysis from section 6.1.3 with interaction terms for urban residence instead of for male.

Table 6.4: Heterogeneous effects on place of residence

Years of education attained	(1)	(2)	(3)
Extreme Drought		0.0216	0.0566
		(0.0719)	(0.0685)
Urban x Extreme Drought		-0.0321	-0.0373
		(0.145)	(0.135)
Moderate Drought	**-0.210***[*]		**-0.235***[**]
	(0.0878)		(0.0838)
Urban x Moderate Drought	0.00868		0.0125
	(0.272)		(0.262)
Female	1.421[***]	-1.421[***]	-1.420[***]
	(0.0934)	(0.0932)	(0.0935)
Urban	1.611[***]	1.617[***]	1.608[***]
	(0.345)	(0.175)	(0.345)
Constant	-0.213	-0.349	-0.214
	(0.461)	(0.433)	(0.458)
N	228872	228872	228872
R^2	0.238	0.238	0.238

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications include controls for gender, age, wealth, agricultural land, type of residence and period in which they were interviewed. All specifications have fixed effects grid level and clustered at ADM1.
Significance levels: * p<0.05, ** p<0.01, *** p<0.001

From table 6.4 we note that the interaction terms between urban residence and drought are not significant, when controlling for the effect of residence on educational attainment. From specification 3 we cannot conclude that the sum of the two interaction terms is non-zero.

Considering the signs of the coefficients of the interaction terms, it is possible that a moderate drought has slightly less adverse effects for residents living in urban areas than for rural residents, while an extreme drought has slightly more adverse effects. The coefficients are very low, and the robust standard errors relatively high, indicating that we cannot detect a causal relationship, nor be confident that the signs of the coefficients represent an indication of relationship.

The results show that there are no apparent differences in how residents of urban and rural areas adapt educational patterns in the presence of a drought. This can indicate that the economic shock induced by lower crops affect urban residents most likely both directly through lower yields from subsistence farming, as well as indirectly through higher food prices.

6.1.4 Adaptability over time

Employing time fixed effects on the decade in which individual were born mean we are only studying variation between those born in the same decade. This is beneficial as it eliminates much of the omitted variable bias arising from individuals from different decades being influenced by different trends.

Consequently, the approach does not allow to study trends across time. Therefore, in the following analysis specification number 5 from table 6.1 is applied to individuals born in different decades separately. The table that summarizes the findings from this is found in appendix 10.1 table A2.

The results indicate that experiencing drought had a negative effect on years of attained education for those born in the 1940s and in the 1970s. For the remaining decades the coefficients on the drought variables are not significantly different from zero. However, the magnitudes of the coefficients indicates that the effect of drought on attained education seem to have been larger for those born before and during the 1980s, compared to the individuals born in the 1990s. This indicates that there has been changes in adaptation to droughts across the decades, with experiences of droughts having a less negative effect on education in recent decades compared to the earlier decades.

Figure 4.5 showed the development in percentage of individuals who had experienced drought. From this we observed two local maximums in the curve of extreme drought, one for those born in the 1940s and one for those born in the 1970s. Noting how this coincides with the decades in which the effect of drought is significant and negative in predicting years of attained education we can infer two different possible pathways in which drought experiences influence educational patterns.

For those born in the 1940s, around 35 percent experienced extreme drought while eligible to attend school. Compared to the level for those born in the 1930s, this is an increase of almost 50 percent. The strong effect of drought for those born in the 1940s can in the light of this, be explained as a consequence of the large relative change from the previous level. Adaptation mechanisms to drought experiences might have been inadequate in accommodating the large change, and consequently the educational patterns were affected. The lack of significant results of drought for those born in the 1960s can thus be due to a lower relative change from the previous decade and increased adaptability.

The peak in percent of drought experiences for those born in the 1970s show that 60 percent of individuals experienced extreme drought while they were eligible to attend school. This high percentage implies that drought was widespread, affecting many individuals at the same time. The data also confirms this, as extreme drought was experienced in all ten countries for those born in the 1970s. Almost 30 percent of these experiences were recorded in Ethiopia.

Observing that experiences of drought for those born in the 1970s induce the largest educational adaptation, it is possible to infer that widespread drought has a larger impact than drought which affects less individuals simultaneously. Widespread drought affecting larger areas, causing loss of crops and higher food prices across large areas, affect household resources negatively, possibly making families prioritize who to send to school in the short term.

The lack of significant results of drought experiences in the decades following 1970 can be due to several mechanisms. Among these are an increase in access to education, and possibly increased adaptation mechanisms in response to the many drought experiences for those born in the 1970s.

Since studying each decade in isolation produced interesting results, an analysis is also applied to consider how heterogeneous effects on genders in response to drought has changes across the decades. This is done by introducing gender interaction terms to samples of those born in the 1950s and those born in the 1990s.

Table 6.5 shows the results from the study of heterogeneous effect between the genders in these two decades.

Table 6.5: Heterogeneous effects on gender across time

Years of education attained	Panel A (1) Education	1950-1960 (2) Education	(3) Education	Panel B (1) Education	1990-2000 (2) Education	(3) Education
Extreme Drought		-0.184	-0.142		0.272	0.275
		(0.113)	(0.115)		(0.142)	(0.143)
Male x Extreme Drought		**0.559****	**0.419***		**-0.531****	**-0.542****
		(0.182)	(0.170)		(0.174)	(0.186)
Moderate Drought	**-0.184***		-0.129	0.0708		-0.0146
	(0.0917)		(0.0907)	(0.135)		(0.134)
Male x Moderate Drought	**0.588****		**0.417***	-0.151		0.0556
	(0.198)		(0.182)	(0.187)		(0.192)
Female	-0.956***	-1.239***	-0.957***	-0.733***	-0.781***	-0.737***
	(0.201)	(0.154)	(0.200)	(0.200)	(0.168)	(0.201)
Urban	1.743***	1.741***	1.743***	1.278***	1.283***	1.283***
	(0.369)	(0.370)	(0.369)	(0.130)	(0.130)	(0.130)
Constant	1.566***	1.814***	1.589***	0.903***	0.889***	0.848***
	(0.377)	(0.402)	(0.388)	(0.242)	(0.220)	(0.246)
N	16489	16489	16489	41970	41970	41970
R^2	0.173	0.173	0.174	0.163	0.164	0.164
F – statistic		F (1, 151) = 11.17			F (1, 137) = 5.43	
P-value			**0.001****			**0.0212***

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications include controls for gender, age, wealth, agricultural land, type of residence and period in which they were interviewed. All specifications have fixed effects grid level and clustered at ADM1.
The F-test is testing the difference between the effect on males and females, by testing significance of the sum of the interaction terms on male. The degrees of freedom are given in parenthesis, and the p-value in the following row.
Significance levels: * p<0.05, ** p<0.01, *** p<0.001

From table 6.5 we observe that the coefficients indicate the heterogeneous effect on the genders is reversed when comparing those born in the 1950s to 1990s. From panel A we observe that for those born in the 1950s and experience a moderate drought, males attend almost half a year more education, relative to the females in this sample. This difference in educational attainment is increased further if an extreme drought is experienced.

From panel B, those born between 1990 and 2000, we observe a different pattern. We observe that in the presence of drought males attain about a half of a year less education, relative to the females in the sample.

The observed pattern of heterogeneity of effect between genders and decades studied can have multiple explanations, and knowing the exact causes requires more in-depth study. However, it can indicate that the value of male child-labor in agriculture decreased during periods of drought in the first sample. The results from the second sample can indicate that the increased focus on access to education for females have been influential.

From table 6.5 we also note two other trends. Being female had a more negative impact on years of attained education for those born in 1950 than in 1990. The negative effect has been reduced by about a fifth of a year. This indicates a positive development in gender equality in educational opportunities, but also shows that there are still improvements to be made. We also note that the negative impact of living in a rural area on years of education has decreased. This can indicate that infrastructure is improving and access to education is increasing.

6.2 Effect of drought on different levels of education

Investigating the effect of drought on years of attained education is beneficial, as it allows to study the possibility that individuals are kept out of education for limited amount of time, before returning.

However, as previous studies have indicated children of different ages are affected differently it is interesting to consider whether the experience of drought has any implication for the probability of attaining a certain level of education (Bjorkman-Nyqvist, 2013). From section 4.6. we know that the level of educational attainment has been changing over time at the aggregated level in our sample.

Investigating the relation between drought and educational level is done by redefining the five levels of education recorded by the DHS into three binary variables. These binary variables capture whether individuals have any education, whether they have completed primary education and whether they have attended secondary or higher education.

Table 6.6: Percentage of sample according to levels of education

	Some education	Completed primary	Secondary or higher
No	36.19	67.33	77.67
Yes	63.81	32.67	22.33
Total	100.00	100.00	100.00

From table 6.6 we observe that sixty-four percent of the sample has some education, while only twenty-two percent has secondary or higher education. The three binary variables are used as the dependent variable to assess the impact of drought on the probability of attending the different levels of education.

Table 6.7 shows the estimation results from the first set of regressions, where having any education is used as the binary variable, leaving only those with no education in the control group.

Table 6.7: Any education and drought

Any Education	(1)	(2)	(3)	(4)	(5)
Moderate Drought	-0.0104	-0.00740	**-0.0308*****	**-0.0292*****	**-0.0281*****
	(0.0150)	(0.0145)	(0.00709)	(0.00728)	(0.00738)
Extreme Drought	0.000816	0.00106	0.00193	0.00208	0.00244
	(0.0185)	(0.0173)	(0.00616)	(0.00594)	(0.00564)
Female		-0.160***	-0.160***	-0.169***	-0.168***
		(0.0117)	(0.0111)	(0.0115)	(0.0122)
Urban		0.211***	0.191***	0.194***	0.121***
		(0.0206)	(0.0203)	(0.0196)	(0.0144)
N	243841	243841	243841	233811	228993
R^2	0.000	0.063	0.145	0.158	0.182
F (1, 188)	0.53	0.19	12.79	10.79	9.13
P-value	0.4692	0.6660	**0.0004*****	**0.0012****	**0.0029****
Religion	No	No	No	Yes	Yes
Wealth	No	No	No	No	Yes
Agriculture	No	No	No	No	Yes
Decade born	No	No	Yes	Yes	Yes
Age	No	No	No	Yes	Yes
Year interviewed	No	No	No	No	Yes

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level.
The F-test is testing the significance of the sum of the two drought dummy variables. Degrees of freedom are given in parenthesis, and the p-value in the following row. For specification 5 the F-statistic has degrees of freedom F (1, 157).
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6.7 shows that experiencing of at least a moderate drought decreases the probability of having any education by roughly three percent in specification 3-5. The F-test indicates that experiencing an extreme drought, relative to a situation of no drought reduces the probability of having any education by almost three percent.

The analysis is also performed with interaction terms for male, to investigate whether there exist heterogeneous effects also here. The results from this analysis indicate that the negative effect on the probability of having any education is partly driven by a more adverse effect for females relative to that of males. The table with these results can be found in appendix 10.1 table A3.

Experiences of drought has no significant effect on the probability of completing primary education or secondary and higher education. No heterogeneous effects on gender are observed for the impact of drought on the probability of these two levels of education. The table with results for these two levels of education is found in appendix 10.1 table A4.

The results from this section can indicate that either drought experiences hinder individuals from ever attending school, by reducing the probability by three percent, or it only keeps individuals out of school for a limited amount of time, before returning. This reasoning can explain why there is no observed effect of drought on the probability of attending the higher levels of education. The result from this section justifies why the main approach in this thesis is to investigate how droughts affect the years of education attained, rather than level attained.

To be better suited to evaluate how droughts affect the level of attained education data on when during their time as eligible to attend school individuals experienced droughts should be used.

6.3 Effect of drought on literacy

Investigating the effect of droughts on educational level and years of education obtained is important but might undermine the importance of the quality of education for long-term development (Hanushek, 2013; Nordstrom & Cotton, 2020). Therefore, investigating the effect of droughts on literacy can give an indication of the possible impacts on the quality of education in the presence of droughts.

The DHS defines literacy as individuals who attended secondary level schooling or who can read a whole sentence or parts of a sentence (Demographic and Health Survey Program, 2021). As the aim is to investigate the relationship between schooling, literacy and drought, the literacy variable is for the purpose defined solely based on whether an individual can read parts of a sentence.[9]

[9] Including also the part of DHS´ definition of literacy of having attended secondary schooling in construction of the variable would mean we could not study the effect of secondary schooling on literacy, as literacy would be in both the dependent variable and the explanatory variable, introducing mechanical correlation.

In this part of the analysis, surveys conducted prior to year 2000 are excluded as these do not ask the literacy question in the same way as the more recent surveys do. The resulting dataset is used to define a binary variable of whether an individual can read or not. This yields a dataset of 216 000 individuals, where fifty-five percent of respondents are defined as literate.

A simple correlation analysis shows that there is a significant, negative correlation between having experienced drought in the years between five and fifteen years of age and being literate, corroborating the findings of Nordstrom and Cotton who found that drought can yield less learning in the presence of drought (Nordstrom & Cotton, 2020).

To investigate the relationship between drought prevalence and literacy a linear probability model is employed. The equation used to study the effect of drought on literacy is the following:

$$\mathrm{P(Literacy_{it} = 1) = \beta_o + \beta_1 Drought + \beta_i X_i \dots\dots + \alpha_i + \lambda_t + u_i} \tag{6.1}$$

As previously, X_i are the control variables, indicated in the specifications, and α_i and λ_t are the entity and time fixed intercepts. Table 6.8 shows the results from this estimation.

Table 6.8: Literacy and drought

Literacy	(1)	(2)	(3)	(4)
Moderate Drought	**-0.0257***	**-0.0249****	**-0.0208****	**-0.0204****
	(0.0128)	(0.00757)	(0.00641)	(0.00630)
Extreme Drought	0.00477	0.00669	0.0153**	0.0148**
	(0.0182)	(0.00906)	(0.00525)	(0.00496)
Female	-0.205***	-0.224***	-0.221***	-0.218***
	(0.0138)	(0.0149)	(0.0142)	(0.0140)
Urban	0.229***	0.116***	0.112***	0.113***
	(0.0207)	(0.0161)	(0.0149)	(0.0148)
Constant	0.647***	0.680***	0.307***	0.433***
	(0.0129)	(0.0221)	(0.0571)	(0.0578)
N	215884	213486	204812	204812
R^2	0.073	0.151	0.171	0.172
Religion	No	Yes	Yes	Yes
Wealth	No	Yes	Yes	Yes
Agriculture	No	Yes	Yes	Yes
Decade born	No	No	Yes	Yes
Age	No	Yes	Yes	Yes
Year interviewed	No	No	No	Yes

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

From table 6.8 we note that experiencing a drought lowers the probability of being literate by around two percent. This can be due to lower quality of education during times of drought, but it can also be a direct result of individuals attending less education, as observed in section 6.1. To investigate these linkages an analysis of the role of education on the probability of literacy is performed.

Education is one of the most important tools to achieve literacy. Controlling for attained education in table 6.8 would however not be desired as we have already observed a direct relationship between drought and education, making education an outcome of the treatment variable. Therefore, the link from attained education to literacy is studied in isolation in table 6.9.

The equation used to study the effect of education on literacy is the following:

$$\mathrm{P}(\mathrm{Literacy}_{it} = 1) = \beta_o + \beta_1 \mathrm{EducationalLevel}_i + \beta_i \mathrm{X}_i \ldots \ldots + \alpha_i + \lambda_t + u_i \quad (6.2)$$

Table 6.9 shows the results of this estimation. All four specifications in table 6.9 include control variables for gender, residence, religion, wealth, agricultural land, age, and survey period. They all have fixed effects on the constructed grid cell level, and time fixed effect on decade born.

Table 6.9: The effect of education on literacy

Literacy	(1)	(2)	(3)	(4)
Started Education	0.672***			0.605***
	(0.00952)			(0.0116)
Completed Primary		0.375***		0.0933**
		(0.0186)		(0.0283)
Secondary Higher			0.344***	0.135***
			(0.0218)	(0.0265)
Female	-0.103***	-0.183***	-0.193***	-0.0954***
	(0.00702)	(0.0125)	(0.0128)	(0.00736)
Urban	0.0407***	0.0723***	0.0742***	0.0229***
	(0.00614)	(0.0103)	(0.0109)	(0.00516)
Constant	0.167***	0.275***	0.300***	0.106**
	(0.0416)	(0.0597)	(0.0617)	(0.0395)
N	204812	204812	204812	204812
R^2	0.474	0.288	0.254	0.505

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level and decade born. The control variables included are gender, residence, religion, wealth, agricultural land, age, and survey period
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In table 6.9 we observe how the different levels of education all contribute to a higher probability of being literate. By exploiting the way the three binary variables of education are constructed, yielding an increase in the control group for each level of education, specification 1-3 show the effect of the binary variables separately.

In specification number 4 all three educational levels are included. The estimated coefficients from this specification allows us to observe the marginal gain in terms of probability of literacy of reaching a higher educational level.

We note that the marginal effect on the probability of literacy of having started any education is much higher than the marginal effect of attaining the higher levels of education on the probability of being literate. This indicate that some individuals do learn to read without formal education, and for those attending education, most learn to read during the first years of education.

From regression table 6.8 and 6.9 we have observed that experiencing a drought reduces the probability of being literate of around two percent, and that having started any education is the greatest contributor to increasing the probability of literacy.

Running an analysis on the relationship between drought and literacy, while controlling for educational attainment would give an indication of how well the educational institutions perform during drought events. The results from this are found in table A5 in the appendix 10.1. These results indicate that when controlling for educational attainment, going from a situation of moderate drought to a situation of extreme drought has a positive and highly significant effect on the probability of being literate of around one percent. The coefficient of the moderate drought variable is negative, but not significant.

This can indicate that the negative impact of less education is induced when a moderate drought occurs, but if the drought becomes even worse communities could have had time to adapt to new conditions, and the added effect on learning outcomes can be slightly positive.

Results from this section imply that increasing the likelihood of individuals attending school is important, as this leads to a higher probability of being literate. This is an important notion, considering that results from regression table 6.2 indicated that a drought could decrease the likelihood of having any education by three percent. Preventing the adverse effects of droughts, by adopting measures for better adaptability, could therefore increase school attendance, and thereby increase literacy. This could be an important part of reaching a higher long-term growth rate for developing countries.

There are however several shortcomings of the linear probability model. The fitted prediction line of OLS can predict less than zero and above one as the change in probability of occurrence, when this is not feasible. Also, as the linear probability model with OLS fits a straight line to a curve that is likely to be non-linear, it usually overstates the predicted changes in the dependent variable when the predicted change is close to zero or one (Stock & Watson, 2015). This means that the predicted reduction of two percent in probability of literacy due to drought occurrences is likely to be overestimated.

6.4 An event study of the Ethiopian famine 1983-1986

The analysis conducted until this point consider whether individuals have experienced moderate or extreme drought during the years they were eligible to attend school. Some of the results show a significant and negative effect of drought on educational attainment. The analysis has not considered in isolation the effects of long-term drought occurrences, that can cause famines.

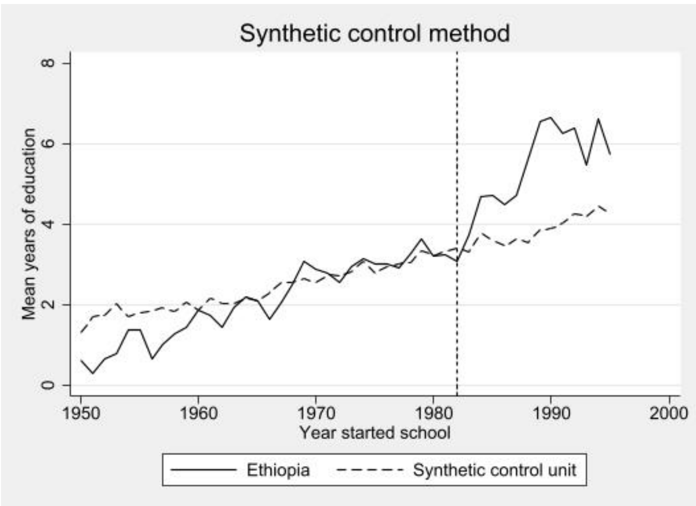
In the years from 1984 to 1986 Ethiopia experienced one of the worst famines in recent history, affecting nearly eight million people, and causing up to one million fatalities (Dercon & Porter, 2014; Vestal, 1985). The reasons behind famines experienced in Ethiopia are several, ranging from poorly led governments, to conflicts and severe and long-lasting droughts (Dercon & Porter, 2014). This makes it difficult to consider the effect of the long-lasting drought in isolation. Therefore, this event study considers the long-term impacts on education from a severe famine, causing among other effects a sharp rise in food prices, which is at least partially attributable to a long-lasting drought.

First, some of the methods from this chapter are applied to Ethiopia in isolation, to evaluate whether droughts affect Ethiopia differently than observed using the full sample. The tables with results from this can be found in appendix 10.1 table A6.

Experiencing moderate drought reduces the years of attended education by almost 0.4 in Ethiopia, depending on specification. This is larger than the overall trend observed in section 6.1.1. Experiencing a moderate drought reduces the probability of being literate with almost four percent, and the added effect of experiencing an extreme drought is increasing the probability of being literate with almost two percent. The probability of having any education is reduced by around six percent if a moderate drought is experienced. It seems that the effects of experiencing a drought while eligible to attend school are larger in Ethiopia, relative to the results from the full sample.

To investigate the effects of the persistent drought in the 1980s in Ethiopia the method of synthetic controls, developed by Abadie et al. is applied (Abadie & Gardeazabal, 2003). This method uses information from untreated countries to generate a control group that has as similar trend to that of Ethiopia prior to the famine as possible. Applying a difference-in-difference approach is inferior to the method of synthetic controls, as no countries in the sample have a similar trend to that of Ethiopia prior to the famine. A difference-in-difference approach is however used to evaluate the validity of results from this method. This analysis can be found in the appendix 10.1 figure A11 and table A12. The difference-in-difference analysis confirms the trend that is found using the method of synthetic controls.

Figure 6.1: Mean years of education in Ethiopia and the synthetic control group



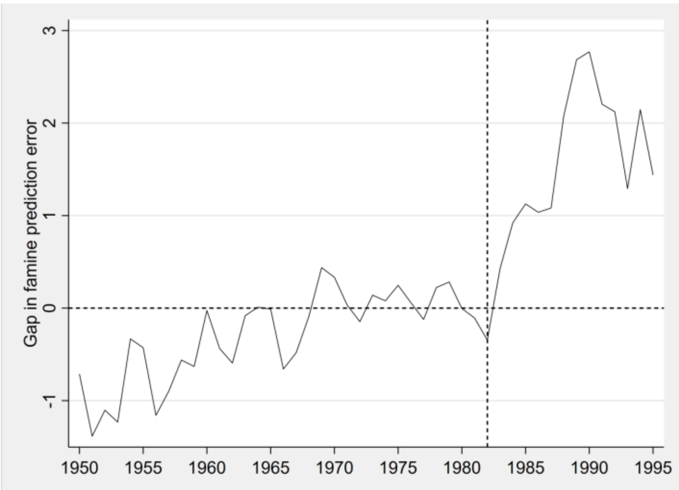
Note: Mean years of education in Ethiopia and the constructed synthetic control group, calculated as the mean years of educational attainment of those who started education in the same year. The figure is developed using Stata.

From figure 6.4 we observe the trend across time in mean years of education reported in Ethiopia, relative to that of the generated synthetic control group. All the other nine countries in the sample are used in a weighted average to generate a synthetic control group that nearly follows the trend of Ethiopia prior to 1982. In order to replicate the trend of Ethiopia prior to the famine most weight is put on the observations from Mali, Niger and Burkina Faso in generation of the synthetic control group.

From figure 6.4 we observe that the trend of the synthetic control group and the trend of Ethiopia follows one another relatively closely prior to 1982, while for those who started school after 1982 the trend in mean years of education in Ethiopia is consistently larger than for the synthetic control group. There are however several shortcomings of this method, making it difficult to determine the true causes of this observed trend.

Most notably, we know that several of the other countries have also experienced famines during the period, making the synthetic control group biased. Also, Ethiopia has experienced several famines across time, making this famine not a shock in isolation. Therefore, the magnitude of the effect is not described. However, it is possible to conclude from figure 6.4 that relative to the synthetic control group constructed, Ethiopia has a larger increase in mean length of education, and that this increase started with the commence of the infamous famine in the 1982.

Figure 6.2: The gap in the predicted error



Note: The evolution in the gap between the mean years of education in Ethiopia and the mean years of education in the constructed synthetic control group. The figure is developed using Stata.

Figure 6.2 shows the evolution in the gap between the mean years of education in Ethiopia and the synthetic control group, visualizing better the difference in trends between the synthetic control and Ethiopia as observed in figure 6.1. The better the fit of the synthetic control group, the smaller the gap in the prediction should be prior to the famine. This gap is higher than desired for the years from 1950 to around 1972, indicating that the synthetic control group does not follow the trend of Ethiopia adequately. However, the larger gap in the prediction post famine, can indicate that Ethiopia benefited from the famine in terms of educational outcomes. Note that other possible elements that could influence educational attainment at country level, such as investments in the educational sector, infrastructure improvements, aid inflows, and conflicts are not controlled for.

Considering the shortcomings of this method the findings should be interpreted with caution. However, the results from this thesis indicate opposite directions of effects of short-term and long-term drought occurrences. It is possible that droughts have a negative impact on households once it commences due to a strain on household resources, but that the long-term effect is that parents prioritize education as a long-term investment, as they have observed the vulnerability in agricultural production. This pattern was observed in a study of long-term adaptation mechanisms to drought in Kenya (Opiyo, et al., 2015). From figure 2.1 we observed how land used for agriculture in Ethiopia declined sharply during the 1990s, possibly motivated by long-lasting drought episodes and higher levels of education.

This part of the analysis should be extended using a more appropriate control group, including important control variables, and considering shocks that are more isolated, to be able to conclude on the direction and magnitude of effect with more certainty.

6.5 Robustness evaluation

This section evaluates the performance of the empirical strategies by testing several of their underlying assumptions.

As mentioned in section 5.4, the Hausman test is run to rule out the use of random effects. All Hausman tests performed on specifications reject the hypothesis of random effects being superior to fixed effects with the constructed dataset.

The issues of heteroscedasticity and autocorrelation are addressed by using clustered standard errors in all specifications, generating heteroscedasticity robust standard errors.

Restricting the sample to individuals who have lived in the place of current residence always violates the central assumption of random sampling for OLS to yield unbiased and consistent estimates. To estimate the effect of this possible source of bias the main specifications employed in part 6.1.1 are imposed also on the sample prior to this restriction. The results from this are found in table 6.10.

Table 6.10: Comparing restricted and full sample

Years of education	Panel A: prior to restriction					Panel B: after the restriction				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Moderate Drought	**-0.268****	**-0.204****	**-0.175*****	**-0.182*****	**-0.151*****	-0.124	-0.0993	**-0.252*****	**-0.249*****	**-0.232*****
	(0.0813)	(0.0751)	(0.0412)	(0.0392)	(0.0359)	(0.130)	(0.125)	(0.0590)	(0.0593)	(0.0571)
Extreme Drought	**-0.224***	**-0.209***	**-0.142****	**-0.109****	**-0.0819***	-0.0265	-0.00866	0.0327	0.0371	0.0481
	(0.0965)	(0.0907)	(0.0433)	(0.0367)	(0.0363)	(0.177)	(0.164)	(0.0591)	(0.0626)	(0.0582)
Female		-1.271***	-1.438***	-1.385***	-1.315***		-1.443***	-1.440***	-1.451***	-1.420***
		(0.0695)	(0.0705)	(0.0689)	(0.0629)		(0.0884)	(0.0880)	(0.0888)	(0.0933)
Urban		3.180***	3.100***	3.106***	1.349***		2.848***	2.677***	2.687***	1.604***
		(0.130)	(0.126)	(0.127)	(0.0967)		(0.211)	(0.206)	(0.191)	(0.163)
Constant	5.445***	5.302***	1.200***	-2.988***	-4.743***	4.600***	4.839***	1.941***	1.199**	-0.214
	(0.0559)	(0.0707)	(0.258)	(0.617)	(0.766)	(0.0909)	(0.111)	(0.201)	(0.374)	(0.444)
N	591907	591907	591907	591907	584181	243718	243718	243718	233690	228872
R^2	0.001	0.107	0.157	0.162	0.244	0.000	0.091	0.162	0.180	0.238

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level and decade born. The control variables included are gender, residence, religion, wealth, agricultural land, age, and survey period
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

From table 6.10, panel A we observe a larger estimated coefficient of extreme drought compared to the effect observed in the restricted sample, presented in panel B. In specification 5, panel A, the sum of the effect on the two drought dummy variables almost equal the coefficient of moderate drought in specification 5, panel B.

This indicates that by restricting the sample to those who have not moved since childhood we are not introducing large bias on the overall effect of drought on educational outcomes. This can be due to individuals reporting having moved residence while having moved short enough to still be within an area that was or was not affected by drought. In the restricted sample we seem to be detecting a slightly lower overall effect than the one detected in the full sample which strengthens the confidence in the results from this thesis.

Analysis is run on subsets of countries in section 6.1.2, without large differences in the overall adaptability of education to drought being detected. This also strengthens the confidence in the validity of estimates.

Clustering on country level rather than ADM1 level is done as an alternative. The results from this are found in appendix 10.1 table A7. Clustering on country rather than ADM1 level yields the same estimated coefficients of the three specifications which yield significant results in section 6.1.1. However, the heteroscedasticity robust standard errors are marginally smaller in specification number 1, 3 and 4, indicating marginally more precise results for these three specifications when clustering on country level. Despite this, clustering on ADM1 level is preferred as ADM1 level is believed to be better suited to capture the geographical spread of drought than countries are, especially for the larger countries in the sample.

Appendix 10.2 also shows a robustness check of the linear probability model by comparing it to the use of ordered logistic and ordered probit models. This can be found in table A10.

7 Weaknesses and possible extensions

This thesis has only used one drought measure, the SPEI, to evaluate the effect of droughts on educational attainment. Alternative measures, such as the SPI or the SPEI global drought monitor could also have been used, despite being calculated slightly differently, as stated in section 4.2.2. To extend on the results from this thesis different drought measures could be included.

With the use of a more frequent drought measure it would be possible to detect whether there exist heterogeneous effects of droughts depending on the time during the year when a drought commences. This could be used to predict the importance of alternative value of labor in educational choices, as it would be possible to detect whether droughts have larger effect on educational outcomes in seasons where child labor is valued differently to other seasons, for instance during the main harvest season.

Two threshold values of drought were generated based on the available descriptions of consequences for crops and livelihoods of the different SPEI values. Other threshold values could be generated to investigate how this choice affects results. Also, other types of extreme weather events, such as floods could be included and investigated to supplement the results presented in this thesis.

As described in section 4 and 6, restrictions had to be imposed on the sample to yield an appropriate dataset for analysis. Omitting all individuals who have moved residency since childhood is in violation of one of the central assumptions underlying the empirical method applied. Those who choose to not move might differ in other aspects from the individuals who choose to move. Furthermore, living in a place with devastating droughts over time could induce forced migration, leaving out the individuals who have been most exposed to droughts in the sample. Although section 6.5 demonstrated that this likely had little implication for the direction and magnitude of the results, a possible extension is to use more randomized data to replicate the analysis.

Another source of potential bias in the analysis is the geographical displacement procedure used by the DHS to ensure confidentiality of individuals´ responses. The DHS regards confidentiality highly, and to be granted access to the geocoded datasets, a separate application must be submitted. These datasets provide a longitude and latitude coordinate associated with each interviewed cluster. These coordinates are geographically displaced by up to two kilometers in urban areas and up to five kilometers in rural areas, with a further one percent of rural clusters being displaced even further, and up to ten kilometers from its true location (Burgert, Colston, Roy, & Blake, 2013).

Using ArcGIS clusters were placed within the 0.5x0.5-degree grid cells and based on this procedure individuals have either experienced drought, or they have not during childhood. Considering that the DHS clusters are somewhat displaced it is possible that some of the clusters identified within one grid cell, truly pertain to a neighboring grid. Further study could consider the possibility of omitting the clusters which fall within a proximity to the border of a grid by two and five kilometers, respectively in rural and urban areas to ensure that this does not influence results. However, as droughts usually affect larger areas simultaneously the implications for the validity of results are assumed to be small.

The findings of this thesis do not necessarily extend to explain how extreme weather events will affect educational attainment in the future in this region, nor in other regions. Firstly, the frequency, spatial coverage and severity of droughts seem to be changing around the world (Masson-Delmotte, et al., 2021). This, together with rapidly changing technologies, might induce different adaptation mechanisms to such events. Secondly, as observed by considering the most recent subsample of individuals, the adverse effects of droughts on educational attainment seem to be subsiding.

Sub-Saharan Africa has different characteristics from the rest of the world, both in terms of population, weather, availability of technology and dependence on agricultural production (The World Bank, 2021). Therefore, the results uncovered in this thesis will most likely not be replicated in other regions of the world. Another possible extension is to use DHS household data to investigate whether other effects of drought on educational outcomes exist in Latin-America or Asia, where the DHS also conducts surveys.

8 Discussion and conclusion

This thesis has used geocoded data to investigate linkages between drought episodes and educational attainment. Education is one of the most powerful tools to combat poverty, reach the SDGs, and enable higher long-term economic growth. Climate change, bringing with it more extreme weather poses a critical threat to agricultural dependent populations in sub-Saharan Africa, possibly lowering their prospects for income generating activities. This thesis has investigated the possible implications this has for educational attainment. The findings from this thesis contributes to the somewhat limited existing literature on the topic of extreme weather events and education.

By combining data from the DHS with climatic data from the SPEI, significant results were found, indicating that episodes of drought during childhood decreases the years of attained education with about a quarter of a year for the full sample. Applying the same analysis to subsamples of countries based on differing climatic conditions show that the effect is replicated in more humid countries, while slightly more adverse in drier countries. These results are largely explained by the construction of the SPEI as deviations from long-term means. This means that according to values recorded by the SPEI a drought in an already dry region is therefore more severe, in absolute terms, than the same drought detected in a humid region.

Experiencing a drought decreases the probability of having any education by three percent. Both this effect, and the negative, significant effect of droughts on years of attained education are strongly heterogeneous for the genders, implying that females where most adversely affected by droughts. However, when only considering a more recent sample, this effect seems to have been reversed. As results show that the probability of being literate increases by sixty percent if the individual has attended any education, it indicates that individuals do learn something in school, and highlights the importance of ensuring universal access to education.

The broader question of this thesis considers how households adapt to exogeneous shocks to household income. The heterogeneous effects on gender can be explained in the context of the substitution and income effects. When deciding on school enrollment for their children, parents trade off the value the children can produce in income generating activities, with the expected future value generated by education. A drought usually yields less crops to cultivate, and therefore lower household income. This can lead to prioritizations of whom to send to school, and consequently less education.

On the other hand, less crops to cultivate imply that the alternative value of child labor in agriculture is reduced, meaning families have less to lose by sending their children to school. It is possible to imagine that these two effects, along with several others, are at play when families decided on school attendance for their children.

Observing how females attained less education in the presence of drought, while males attained more it is reasonable to conclude that the size of the two effects differed between the genders. At least in earlier decades it seems households prioritized ensuring education for males in the presence of household income shocks. Furthermore, the relative value of male labor in agriculture relative to education shrinks when there are less crops to cultivate. The same cannot be claimed for the value of female labor if primarily used for home production. These results are in line with the findings on heterogeneous effects of weather shocks on educational outcomes between the genders found in both Ethiopia and Uganda (Bjorkman-Nyqvist, 2013; Mani, et al., 2013). A possible avenue for further research is studying these mechanisms more in depth.

Results from this thesis also show that the experience of drought can decrease the likelihood of literacy by two percent, indicating that learning outcomes are lower during times of stress on household income. There exist several possible explanations behind this effect. A recent study of the effects of the COVID-19 pandemic on rural household in Uganda concludes that households lowered their expenditures on food consumption by as much as forty-four percent and exhausted their savings in response to the imposed lockdown (Mahmus & Riley, 2020).

These findings indicate that many households in rural Africa have scarce resources, and therefore changes to their disposable income have large implications. Mitigating the adverse effects of extreme weather on crops is therefore important to ensure adequate nutrition and resources to prioritize education. The adverse effects of droughts on crops can at least partially explain why results show that individuals experiencing droughts have a lower probability of having any education and being literate.

As it is observed that adaptations to drought seems to be stronger during earlier decades, than for the more recent decades, it can indicate that the increased focus on equal and adequate access to education has been influential. Briefly studying the effects of a severe famine in Ethiopia seem to indicate that persistent drought episodes can increase educational attainment, possibly through increasing the expected payoff from education, relative to that of a career in a vulnerable agricultural sector. Therefore, the long-term effects of persistent droughts can have positive implications for the level of investment in education. This is contrary to the findings from earlier decades on negative short-term effects of experiencing drought on the years of attained education.

9 Bibliography

Abadie, A., & Gardeazabal, J. (2003). The Economic Costs of Conflict: A Case Study of the Basque Country. *The American Economic Review*, Volum 93(1), p 113-131.

Abel, G., Brottrager, M., Cuaresma, J., & Muttarak, R. (2019). Climate, conflict and forced migration. *Global Environmental Change*.

Agamile, P., Dimova, R., & Golan, J. (2021). Crop Choice, Drought and Gender: New Insights from Smallholders' Response to Weather Shocks in Rural Uganda. *Journal of Agricultural Economics*, Volum 72(3), p 829-856.

Alesina, A., Hohmann, S., Michalopoulos, S., & Papaioannou, E. (2021). Intergenerational mobility in Africa. *Econometrica*, Volum 89(1), p 1-35.

ArcGIS Pro. (2021, September 20). *Spatial Join (Analysis)*. Retrieved from esri: https://pro.arcgis.com/en/pro-app/latest/tool-reference/analysis/spatial-join.htm

Barrios, S., Bazoumana, O., & Strobl, E. (2008). The impact of climatic change on agricultural production: Is it different for Africa? *Food policy*, Volum 33(4), p. 287-298.

Beguería, S., Latorre, B., Reig, F., & Vicente-Serrano, S. M. (2021, May 19). *Global SPEI database* . Retrieved from Global SPEI database : https://spei.csic.es/database.html

Begueria, S., Vicente-Serrano, S. M., & Angulo-Martinez, M. (2010). A mulitscalar global drought dataset: The SPEIbase: a new gridded product for the analysis of drought variability and drought impacts. *American Meteorological Society*, Volum 91(10), p.1351-1354.

Bjorkman-Nyqvist, M. (2013). Income shocks and gender gaps in Education: Evidence from Uganda. *Journal of Development Economics*, Volum 105, p. 237-253.

Burgert, C. R., Colston, J., Roy, T., & Blake, Z. (2013). *Geographic Displacement Procedure and Georeferenced Data Release Policy for the Demographic and Health Surveys. DHS Spatial Analysis Reports.* Calverton, Maryland: USAID.

Burgert, R. C., Zachary, B., & Colston, J. (2013). *Incorporating geographic information into demographic and health surveys: A field guide to GPS data collection.* Calverton, Maryland : The United States Agency for International Development.

Crespo Cuaresma, J., & Lutz, W. (2007). *Human Capital, Age Structure and Economic Growth: Evidence from a New Dataset. IIASA Interim Report.* Laxenburg, Austria: IIASA.

Demographic and Health Survey Program. (2021, August 29). *Literacy*. Retrieved from Guide to DHS Statistics DHS-7: https://dhsprogram.com/data/Guide-to-DHS-Statistics/Literacy.htm

Dercon, S., & Porter, C. (2014). Live Aid Revisited. *Oxford University Press*, Volum 12(4), p. 927-948.

Ellis, J., & Galvin, K. (1994). Climate Patterns and Land-use Practices in the Dry Zones of Africa. *BioScience* , Volume 44 (5), p. 340-349.

ESA Globcover Project. (2021, September 24). *ESA*. Retrieved from GlobCover: http://due.esrin.esa.int/page_globcover.php

FAO, IFAD, UNICEF, WFP and WHO. (2021). *The State of Food Security and Nutrition in the World 2021.* Rome, Italy: FAO, IFAD, UNICEF, WFP and WHO.

Filmer, D., & Pritchett, L. (2001). Estimating wealth effects without expentidure data-or tears: An application to educational enrollments in states of India. *Demography*, Volum 38(1), p. 115-132.

Glory, E., & Nsikak-Abasi, E. (2013). Child labour in agriculture among poor rural households: some issues and facts. *European journal of Physical and Agricultural Sciences*, Volum 2 (1), p. 13-19.

Haile, G. G., Tang, Q., Hosseini-Moghari, S.-M., Liu, X., Gebremicael, T. G., Leng, G., . . . Yun, X. (2020). Projected Impacts of Climate Change on Drought Patterns over East Africa. *Earth´s Future*, Volum 8 (7), p. n/a.

Hanushek, E. (2013). Economic growth in developing countries: The role of human capital. *Economics of education review*, Volume 37, p. 204-212.

Hertel, T. W., & Rosch, S. D. (2010). Climate Change, Agriculture and Poverty. *Applied Economic Perspectives and Policy*, Volum 32(3), p. 355-385.

Hongbin, L., Junsen, Z., & And, Z. (2008). The quantity-quality trade-off of children in a developing country: Identification using Chinese twons. *Demography*.

Humanitarian Data Exchange . (2021, September 23). *Humanitarian Data Exchange* . Retrieved from Dataset, Administrative Divisions: https://data.humdata.org/search?ext_administrative_divisions=1&q=&ext_page_size=25

International Labour Organization. (2010). *Accelerating action against child labour – Global Report under the follow-up to the ILO Declaration on Fundamental Principles and Rights at Work.* Geneva: International Labour Organization.

International Labour Organization, ILOSTAT database. (2021, January 29). *The World Bank*. Retrieved from Employment in agriculture (% of total employment) (modeled ILO estimate) - Sub-Saharan Africa: https://data.worldbank.org/indicator/SL.AGR.EMPL.ZS?locations=ZG

Knight, J., & Shi, L. (1996). Educational attainment and the rural-urban divide in China. *Oxford Bulletin of Economics and Statistics*, Volume 58(1), p. 83-117.

Kotsdam, A., & Flatø, M. (2020). Rainfall shocks and intimate partner violence in sub-Saharan Africa. *Journal of Peace Research* .

Linard, C., Gilbert, M., Snow, R., Noor, A., Tatem, A. J., & Schumann, G. J.-P. (2012). Population Distribution, Settlement Patterns and Accessibility across Africa in 2010. *PLoS One*, Volum 7(2), p.e31743-e31743.

Mahmus, M., & Riley, E. (2020). Household response to an extreme shock: Evidence on the immediate impact of the Covid-19 lockdown on economic outcomes and well-being in rural Uganda. *World Development*.

Mani, S., Hoddinott, J., & Strauss, J. (2013). Determinants of Schooling: Empirical Evidence from Rural Ethiopia. *Journal of African Economies,*, Volum 22 (5), p. 693-731.

Masson-Delmotte, V., Zhai, A., Pirani, S. C., Péan, S., Berger, N., Caud, Y., . . . B, Y. (2021). *Climate Change 2021 The Physical Scence Basis. Contribution of Working Group I to the Sixth*

Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press.

Murray, U., Paola, T., Demeranville, J., & Hurst, P. (2010). Breaking the rural poverty cycle: Getting girls and boys out of work and into school. Washington DC.

National Drought Mitigation Center. (2021, August 2). *U.S. Drought Monitor*. Retrieved from Drought Classification: https://droughtmonitor.unl.edu/About/AbouttheData/DroughtClassification.aspx

Nordstrom, A., & Cotton, C. (2020). Impact of a Severe Drought on Education: More Schooling but Less Learning. *Environment and development economics*, Volum 24(4), p. 413-436.

Opiyo, F., Wasonga, O., Nyangito, M., & Schilling, J. (2015). Drought Adaptation and Coping Strategies Among the Turkana Pastoralists of Northern Kenya. *International Journal of Disaster Risk Science*, Volum 6(3), p. 295-309.

Peace Research Institute Oslo. (2021, August 5). *Download the PRIO-GRID data*. Retrieved from PRIO-GRID 2.0: https://grid.prio.org/#/download

Randell, H., & Gray, C. (2016). Climate variability and educational obtainment: Evidence from rural Ethiopia. *Global Environmental Change*, Volum 41, p. 111-123.

Rutstein, S. O., & Johnson, K. (2004). *The DHS Wealth Index.* Maryland: DHS comparative reports.

Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., & Ziese, M. (2015). *GPCC Full Data Reanalysis Version 7.0 at 0.5°: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data.* Putgarten, Germany: Global Precipitation Climatology Centre at Deutscher Wetterdienst.

Shah, M. (2017). Drought of Opportunities: Contemporaneous and Long Term Impacts of Rainfall Shocks on Human Capital. *Journal of Political Economy*, Volum 125(2), p. 527-561.

Stock, J. H., & Watson, M. W. (2015). *Introduction to Econometrics.* (Third Edition). Harlow, England: Pearson Education Limited.

Sumida, S., & Kawata, K. (2021). An analysis of the learning performance gap between urban and rural areas in sub-Saharan Africa. *South African Journal of Education*.

Tacoli, C., & Satterthwaite, D. (2013). Gender and urban change. *Environment & Urbanization*, Volum 25(1), p. 3-8.

The DHS Program. (2021, August 2). *DHS Model Questionnaires* . Retrieved from The DHS Program: https://dhsprogram.com/Methodology/Survey-Types/DHS-Questionnaires.cfm

The World Bank. (2021, September 3). *Education, School enrollment, primary*. Retrieved from The World Bank: https://data.worldbank.org/topic/education?locations=ZG-1W

The World Bank. (2021, September 21). *Sub-Sarahan Africa*. Retrieved from The World Bank Data: https://data.worldbank.org/country/ZG

Thurlow, J., Zhu, T., & Xinshen, D. (2012). Current Climate Variability and Future Climate Change: Estimated Growth and Poverty Impacts for Zambia. *Review of Development Economics*, Volum 16. p. 394–411.

Tollefsen, A. F., Bahgat, K., Nordkvelle, J., & Buhaug, H. (2015). *PRIO-GRID v.2.0 Codebook.* Oslo, Norway: Peace Research Institute Oslo.

Tollefsen, A. F., Strand, H., & Buhaug, H. (2012). PRIO-GRID: A unified spatial data structure. *Journal of Peace Research*, Volum 49(2), p. 363-374.

UNESCO Institute for Statistics. (2021, August 10). *Agricultural land (% of land area) - Sub-Saharan Africa*. Retrieved from The World Bank: https://data.worldbank.org/indicator/AG.LND.AGRI.ZS?locations=ZG

UNESCO Institute for Statistics. (2021, September 15). *School enrollment, primary (gross), gender parity index (GPI) - Sub-Saharan Africa*. Retrieved from The World Bank: https://data.worldbank.org/indicator/SE.ENR.PRIM.FM.ZS?locations=ZG

United Nations. (2021, September 21). *United Nations*. Retrieved from United Nations Sustainable Development : https://sdgs.un.org/goals/goal4

United Nations. (2021, September 20). *United Nations Climate Action*. Retrieved from What is climate change?: https://www.un.org/en/climatechange/what-is-climate-change

Vestal, T. M. (1985). Famine in Ethiopia: Crisis of Many Dimensions. *Africa Today*, Volume 32 (4), p. 7-28.

World Vision. (2021, August 11). *1980s Ethiopia famine: Facts, FAQs, and how to help*. Retrieved from World Vision: https://www.worldvision.org/disaster-relief-news-stories/1980s-ethiopia-famine-facts

10 Appendix

10.1 Additional tables

Table A 1: All gender interaction terms

Full sample	(1)	(2)	(3)
	Years of Education		
Female x Moderate Drought	**-0.278***** (0.0716)		**-0.311***** (0.0677)
Male x Moderate Drought	-0.110 (0.0791)		-0.102 (0.0876)
Female x Extreme Drought		0.0315 (0.0800)	0.0759 (0.0768)
Male x Extreme Drought		-0.0175 (0.0645)	-0.00618 (0.0700)
Female	-1.270*** (0.126)	-1.443*** (0.0973)	-1.271*** (0.126)
Urban	1.604*** (0.162)	1.604*** (0.163)	1.604*** (0.162)
Constant	-0.274 (0.458)	-0.341 (0.440)	-0.284 (0.452)
N	228872	228872	228872
***R*²**	0.238	0.238	0.238

Wet region	(1)	(2)	(3)
	Years of Education		
Female x Moderate Drought	**-0.264*** (0.113)		-0.143 (0.115)
Male x Moderate Drought	**0.298*** (0.122)		0.187 (0.127)
Female x Extreme Drought		**-0.454***** (0.125)	**-0.427**** (0.124)
Male x Extreme Drought		**0.309*** (0.135)	0.270 (0.141)
Female	-1.778*** (0.214)	-2.041*** (0.174)	-1.778*** (0.215)
Urban	0.824*** (0.111)	0.823*** (0.110)	0.823*** (0.110)
Constant	-0.0946 (0.480)	0.118 (0.454)	-0.0456 (0.459)
N	61611	61611	61611
R2	0.291	0.292	0.292

Dry region	(1)	(2)	(3)
	Years of Education		
Female x Moderate Drought	**-0.401***** (0.0901)		**-0.408***** (0.0896)
Male x Moderate Drought	-0.167 (0.110)		-0.186 (0.127)
Female x Extreme Drought		-0.0456 (0.0485)	0.0167 (0.0480)
Male x Extreme Drought		0.00945 (0.0798)	0.0376 (0.0893)
Female	-0.940*** (0.143)	-1.126*** (0.130)	-0.940*** (0.143)
Urban	2.150*** (0.295)	2.150*** (0.295)	2.150*** (0.295)
Constant	-0.698	-0.801	-0.697

	(0.478)	(0.464)	(0.476)
N	114682	114682	114682
R2	0.225	0.224	0.225

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level and include controls for gender, age, wealth, agricultural land, residency, religion and period interviewed.
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A 2: Effect of drought on separate samples

	Decade born: Effect of experiencing a drought					
Years of education	**1940-1950**	**1950-1960**	**1960-1970**	**1970-1980**	**1980-1990**	**1990-2000**
Moderate Drought	**-0.256****	0.0399	0.261	**-0.426****	-0.167	-0.0000651
	(0.0921)	(0.0802)	(0.147)	(0.132)	(0.0860)	(0.104)
Extreme Drought	0.236	0.0110	0.102	0.189	-0.161	0.0666
	(0.147)	(0.101)	(0.0717)	(0.105)	(0.0886)	(0.113)
Female	-0.561***	-1.415***	-1.798***	-1.803***	-1.464***	-0.603***
	(0.166)	(0.148)	(0.134)	(0.122)	(0.125)	(0.138)
Urban	1.306***	1.739***	1.765***	1.802***	1.425***	1.279***
	(0.251)	(0.372)	(0.305)	(0.247)	(0.131)	(0.130)
Constant	-0.931	1.851***	1.876***	2.700***	3.614***	0.859**
	(0.851)	(0.385)	(0.278)	(0.272)	(0.258)	(0.256)
N	**4330**	**16489**	**37962**	**60481**	**67406**	**41970**
R^2	0.140	0.172	0.186	0.193	0.180	0.162
P-value	0.9018	0.6718	**0.0238***	0.1411	**0.0150***	0.6190

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level and include controls for gender, age, wealth, agricultural land, residency, religion and period interviewed.
The given p-value is that of the F-statistic testing the joint significance of the two drought dummy variables.
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A 3: Heterogeneous effects of drought on any education

	Any Education		
	(1)	(2)	(3)
Extreme Drought		-0.00325	0.00186
		(0.00862)	(0.00827)
Male x Extreme Drought		0.00458	0.00146
		(0.0122)	(0.0119)
Moderate Drought	**-0.0354*****		**-0.0361*****
	(0.00824)		(0.00823)
Male x Moderate Drought	0.0213		0.0205
	(0.0116)		(0.0105)
Female	-0.148***	-0.166***	-0.148***
	(0.0126)	(0.0110)	(0.0126)
Urban	0.121***	0.121***	0.121***
	(0.0144)	(0.0144)	(0.0144)
Constant	0.362***	0.353***	0.362***
	(0.0532)	(0.0501)	(0.0524)
N	228993	228993	228993
R^2	0.182	0.182	0.182
P-value (extreme)			0.1496
P-value (moderate)			0.0013**

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level and include controls for gender, age, wealth, agricultural land, residency, religion and period interviewed.
The given p-value is that of the F-statistic testing the joint significance of the drought dummy variable and the interaction term.
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A 4: Probability of having completed primary and secondary

	Completed primary				
	(1)	(2)	(3)	(4)	(5)
Moderate Drought	0.00337	0.00478	-0.00733	-0.00615	-0.00578
	(0.00987)	(0.00965)	(0.00677)	(0.00672)	(0.00631)
Extreme Drought	-0.00267	-0.00160	0.000739	-0.000693	0.000864
	(0.0125)	(0.0119)	(0.00501)	(0.00534)	(0.00506)
Female		-0.0862^{***}	-0.0864^{***}	-0.0978^{***}	-0.0931^{***}
		(0.00767)	(0.00773)	(0.00698)	(0.00697)
Urban		0.194^{***}	0.182^{***}	0.184^{***}	0.115^{***}
		(0.0182)	(0.0181)	(0.0170)	(0.0143)
Constant	0.314^{***}	0.322^{***}	0.136^{***}	0.341^{***}	0.286^{***}
	(0.00708)	(0.00956)	(0.0154)	(0.0428)	(0.0434)
N	243841	243841	243841	233811	228993
R^2	0.000	0.031	0.060	0.069	0.091

	Secondary or higher education				
	(1)	(2)	(3)	(4)	(5)
Moderate Drought	-0.00800	-0.00705	-0.0115	-0.0125	-0.0116
	(0.0102)	(0.0100)	(0.00767)	(0.00749)	(0.00710)
Extreme Drought	-0.00871	-0.00786	-0.000963	-0.00343	-0.00235
	(0.0116)	(0.0111)	(0.00393)	(0.00378)	(0.00377)
Female		-0.0648^{***}	-0.0641^{***}	-0.0756^{***}	-0.0717^{***}
		(0.00641)	(0.00652)	(0.00570)	(0.00585)
Urban		0.190^{***}	0.178^{***}	0.178^{***}	0.112^{***}
		(0.0159)	(0.0160)	(0.0152)	(0.0135)
Constant	0.221^{***}	0.217^{***}	0.0621^{***}	0.305^{***}	0.247^{***}
	(0.00816)	(0.0103)	(0.0135)	(0.0475)	(0.0477)
N	243841	243841	243841	233811	228993
R^2	0.000	0.031	0.071	0.076	0.100
Religion	No	No	No	Yes	Yes
Wealth	No	No	No	No	Yes
Agriculture	No	No	No	No	Yes
Decade born	No	No	Yes	Yes	Yes
Age	No	No	No	Yes	Yes
Year interviewed	No	No	No	No	Yes

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level.
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A 5: Effect of drought on literacy when controlling for education

Literacy	(1) Literal	(2) Literal	(3) Literal	(4) Literal
Started Education	0.672***			0.604***
	(0.00956)			(0.0117)
Completed Primary		0.375***		0.0933**
		(0.0186)		(0.0282)
Secondary Higher			0.344***	0.135***
			(0.0218)	(0.0265)
Moderate Drought	-0.00355	**-0.0194*** **	**-0.0179** **	-0.00400
	(0.00365)	(0.00527)	(0.00587)	(0.00315)
Extreme Drought	**0.00956*** **	**0.0125** **	**0.0142** **	**0.00928*** **
	(0.00270)	(0.00401)	(0.00512)	(0.00257)
Female	-0.103***	-0.183***	-0.193***	-0.0954***
	(0.00702)	(0.0125)	(0.0128)	(0.00737)
Urban	0.0407***	0.0723***	0.0742***	0.0229***
	(0.00614)	(0.0103)	(0.0109)	(0.00516)
Constant	0.167***	0.281***	0.305***	0.106**
	(0.0414)	(0.0601)	(0.0620)	(0.0394)
N	204812	204812	204812	204812
R^2	0.475	0.288	0.254	0.505

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level and decade born, in addition to and controls for gender, age, wealth, agricultural land, residency, religion and period interviewed.
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A 6: Analysis replicated for Ethiopia

Years of education Ethiopia	(1)	(2)	(3)	(4)	(5)
Moderate Drought	-0.0187	-0.0625	-0.0625	-0.287	**-0.339***
	(0.253)	(0.221)	(0.221)	(0.146)	(0.127)
Extreme Drought	-0.201	-0.114	-0.114	-0.119	-0.0683
	(0.123)	(0.100)	(0.100)	(0.0909)	(0.0861)
Female		-1.830***	-1.830***	-2.031***	-1.685***
		(0.203)	(0.203)	(0.204)	(0.209)
Urban		5.804***	5.804***	5.289***	4.093***
		(0.368)	(0.368)	(0.273)	(0.337)
Constant	3.698***	3.476***	3.476***	6.451***	3.641***
	(0.230)	(0.205)	(0.205)	(0.470)	(0.498)
N	49511	49511	49511	49299	49299
***R*²**	0.001	0.236	0.236	0.296	0.339

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level and include the same control variables as in regression table 6.1.
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Literacy Ethiopia	(1)	(2)	(3)	(4)
Moderate Drought	-0.0144	**-0.0437*****	**-0.0365****	**-0.0359****
	(0.0226)	(0.00894)	(0.0109)	(0.0105)
Extreme Drought	-0.00605	-0.00173	**0.0183***	**0.0182***
	(0.00950)	(0.0115)	(0.00789)	(0.00778)
Female	-0.294***	-0.319***	-0.280***	-0.277***
	(0.0285)	(0.0300)	(0.0265)	(0.0261)
Urban	0.474***	0.315***	0.275***	0.273***
	(0.0336)	(0.0362)	(0.0303)	(0.0310)
Constant	0.554***	0.654***	0.228***	0.328***
	(0.0298)	(0.0526)	(0.0512)	(0.0632)
N	49485	49485	49273	49273
***R*²**	0.179	0.250	0.274	0.275

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level and include the same control variables as in regression table 6.8.
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Any education Ethiopia	(1)	(2)	(3)	(4)	(5)
Moderate Drought	-0.0169	-0.0192	**-0.0698****	**-0.0648****	**-0.0621****
	(0.0389)	(0.0336)	(0.0157)	(0.0151)	(0.0152)
Extreme Drought	-0.0217	-0.0115	0.0219	0.0184	0.0192
	(0.0160)	(0.0142)	(0.0120)	(0.0105)	(0.0107)
Female		-0.249***	-0.210***	-0.226***	-0.220***
		(0.0296)	(0.0301)	(0.0289)	(0.0292)
Urban		0.454***	0.398***	0.367***	0.244***
		(0.0280)	(0.0393)	(0.0273)	(0.0303)
N	49511	49511	49511	49299	49299
***R*²**	0.001	0.147	0.298	0.311	0.333

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell level and include the same control variables as in regression table 6.7.
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A7: Clustering by country

	Clustering on country				
	(1) Education	(2) Education	(3) Education	(4) Education	(5) Education
Moderate Drought	-0.124	-0.0993	**-0.252****	**-0.249*****	**-0.232****
	(0.152)	(0.148)	(0.0541)	(0.0515)	(0.0581)
Extreme Drought	-0.0265	-0.00866	0.0327	0.0371	0.0481
	(0.311)	(0.296)	(0.0965)	(0.0969)	(0.0896)
Female		-1.443***	-1.440***	-1.451***	-1.420***
		(0.245)	(0.253)	(0.245)	(0.234)
Urban		2.848***	2.677***	2.687***	1.604**
		(0.536)	(0.506)	(0.465)	(0.467)
Constant	4.600***	4.839***	1.941***	1.199	-0.214
	(0.126)	(0.270)	(0.277)	(0.554)	(0.737)
N	243718	243718	243718	233690	228872
***R*²**	0.000	0.091	0.162	0.180	0.238
Religion	No	No	No	Yes	Yes
Wealth	No	No	No	No	Yes
Agriculture	No	No	No	No	Yes
Decade born	No	No	Yes	Yes	Yes
Age	No	No	No	Yes	Yes
Year interviewed	No	No	No	No	Yes

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications. All specifications have fixed effects on grid cell. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A 8: Correlation matrix between drought and education

Variables	(1)	(2)	(3)
(1) Years of education	1.000		
(2) Moderate Drought	**-0.079***	1.000	
	(0.000)		
(3) Extreme Drought	**-0.136***	0.310*	1.000
	(0.000)	(0.000)	

**** p<0.01, ** p<0.05, * p<0.1*

Table A 9: Variables used in regression analysis

Variable	Usage	Description	Source
Years of education	Dependent variable	How many years the individual has attended any education	DHS
Highest level of education	Dependent variable	The highest level of attained education	DHS
Literacy	Dependent variable	Defined as the ability to read parts of a sentence	DHS
Extreme drought while in school	Main variable of interest	Dummy variable capturing incidences of SPEI below -1,6 during the years individuals are eligible to attend school	SPEI through PRIO-GRID
Moderate drought while in school	Main variable of interest	Dummy variable capturing incidences of SPEI below -0,8 during the years individuals are eligible to attend school	SPEI through PRIO-GRID
Gender	Explanatory variable	Recoded to female, where female = 1 is female	DHS
Residency	Explanatory variable	Recoded to urban where urban = 1 is urban residence	DHS
Wealth index	Explanatory variable	Index ranging from 1-5, where 1 is poorest and 5 is richest. Computed as a relative measure for each survey round.	DHS
Religion	Explanatory variable	Recording religion of individuals.	DHS
Age	Explanatory variable	Age of respondent at time of survey.	DHS
Agricultural land	Explanatory variable	Gives the percentage of land used for agriculture within each grid cell. Variable is recoded into ten ranges, from 0-9, where 0 means the cell has 0 - 10 percent agricultural land, and 9 indicates that the cell has between 90 and 100 percent land used for agriculture.	European Space Agency Global Cover Portal through PRIO-GRID
First administrative levels	Clustering variable		Humanitarian Data Exchange

10.2 Further explanations

Construction of new grid cells

The drought variables are recorded in grid cells, sourced from SPEI. These cells to not stay within country or administrative borders. Individual country residence is recorded by the DHS. However, as grid cells cross borders, two individuals living on either side of a country border will be placed in the same grid, regardless of their country residence. The procedure of splitting grid cells between countries is important when employing fixed effects on grid level, as otherwise one would be assuming that an individual on either side of a country border experiences the same kind of local development. Hence fixing on the original grid cell level provided by PRIO could yield biased estimates.

To ensure that grid cells do not overlap borders used for analysis, new grid cells were constructed. This was done by generating a new set of grid cells, which are a combination of a country code and the grid code it is ensured that grids are split between countries. Using this grid code to repeat the process on ADM1 levels as well, ensures that each grid pertains to one country, and one ADM1 level. After the process, some grid cell levels still pertained to more than one ADM1 border, due to codes used for generation adding up to the same numbers. In these cases, codes were manually recoded to ensure every new grid cell had a unique code. This process ensured grid cells were nested within only one country and one ADM1 level over time.

Explanatory variables and the role of agricultural dependence

The signs and magnitude of the majority of control variables are not included in the regression tables. Therefore, a short description of their role is given here.

Increasing wealth predicts higher education, higher age predicts less education, and consequently the more recent an individual was born the more education it attains. The level of agricultural land in the area is not significant in explaining the years of education in any of the specifications.

All types of religion are significant in explaining educational attainment. Relative to being catholic, being a protestant yields almost half a year more in school, while being Muslim, animist, having no religion or reporting “other religion” all yield around a year less schooling compared to individuals with catholic believes.

Also, the period in which individuals where interviewed is important in the specifications. Relative to being interviewed in the period between 1990 and 1995 individuals who are interviewed between 2015 and 2020 are likely to attend school for almost 1,5 years longer.

The decade in which individuals are born is an important time fixed effect, proving highly significant. We observe that relative to being born in the 1930ies being born in each preceding decade yields higher years of education.

In an alternative specification interaction terms between levels of agricultural land and drought occurrence are included to check whether the hypothesis that drought shocks affect education in agricultural heavy areas more than in non-agricultural areas hold. These interaction terms are not significant in explaining years of education attained. This might indicate that the potential effect of drought on education goes through other channels than changes in agricultural produce.

Robustness of the linear model

The linear model is selected for analysis in sections 6.2 and 6.3 mainly as it is easier to include fixed effects to this model, compared to the alternatives of an ordered logistic or ordered probit model. The ordered models also assume some form of linearity between the dependent variable and the effect of the treatment variable (Stock & Watson, 2015). Below is a table showing the results from employing an ordered logistic and ordered probit model to the same analysis as performed in section 6.2.

Table A 10: Comparing the linear model to the ordered logistic and probit models

Educational attainment	**Linear model**		**Ordered logistic model**		**Ordered probit model**	
Moderate Drought	-0.0183 (0.0378)	-0.0103 (0.0363)	-0.124 (0.100)	-0.0653 (0.0790)	-0.0672 (0.0563)	-0.0380 (0.0446)
Extreme Drought	-0.0285 (0.0475)	-0.0230 (0.0437)	-0.500*** (0.117)	-0.455*** (0.123)	-0.300*** (0.0670)	-0.279*** (0.0717)
Female		-0.455*** (0.0270)		-0.582*** (0.0586)		-0.358*** (0.0335)
Urban		0.950*** (0.0677)		1.559*** (0.134)		0.896*** (0.0759)
N	243840	243840	243840	243840	243840	243840

Note: Standard errors are clustered at the ADM1 level, and heteroscedasticity robust. They are given in parentheses for all specifications.
Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

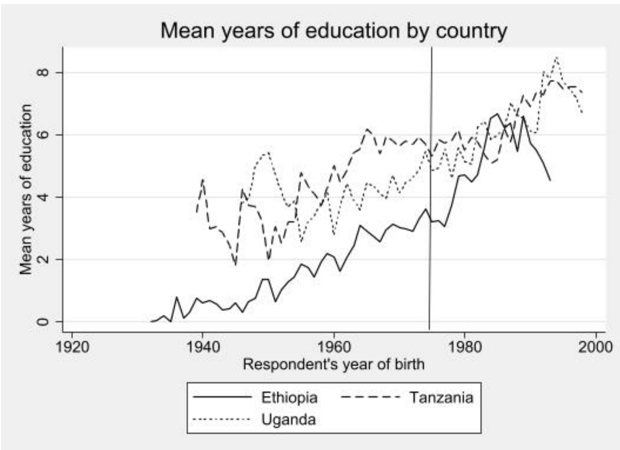
Although the selected specification for the linear model yields non-significant results of drought on the levels of educational attainment, the sign of the results of the ordered probit and logit model are in line with the general findings. Being negative, they indicate that if individuals experience drought they are more likely to be in the lower levels of educational attainment. This indicates that the linear model behaves well. Since the results of the linear model is without significance, it is possible to conclude that the selected model is more conservative, and hence the significant results that do emerge can be trusted more.

Difference-in-difference approach to evaluate the effect of famine in Ethiopia

One of the central underpinning of the difference in difference approach is that the treated and untreated groups have the same trend prior to the shock. This is the reason why the main method applied to consider the effects of the famine is the synthetic control method.

However, a difference in difference analysis is performed to evaluate the validity of the results from the synthetic control analysis. This approach compares the trend in educational outcomes for people in Ethiopia before and after the famine, with this trend in other countries, that were less affected by famines. Tanzania and Uganda are selected as the comparisons with Ethiopia, as the observations in the dataset indicate that these two countries did not experience a persistent drought. The observations are limited to those that attended school after 1960.

Figure A 11: Mean years of education over time in Ethiopia, Uganda and Tanzania



Note: Mean years of education in Ethiopia, Uganda and Tanzania, calculated by the mean years of education of those born in the same year. The figure is developed using Stata.

Table A 12: Difference-in-difference results

Number of observations: 66589

	Before famine	**After famine**
Control:	9853	11588
Treated:	31781	13367

Outcome variable	**Years of education**
Before	
Control	5.206
Treated	3.078
Difference	-2.127
After	
Control	6.711
Treated	5.612
Difference	-1.099
Diff-in-diff	**1.028**

The results from the difference in difference analysis, in table A12, indicate that the mean level of education in Ethiopia was more than two years behind that of Uganda and Tanzania prior to 1982. After 1982 the difference in means between the two countries is reduced to around one year. All differences are significant at one percent level.

From figure A11 we observe that those born around 1975, indicated by the vertical line, experience a small reduction in the mean years of education reported, possibly due to the short-term effect of the famine. However, the mean years of education is trending sharply upwards for those that start education after the famine. This trend is not observed for Uganda and Tanzania.

This can indicate that the effect of the famine in Ethiopia had a positive effect on years of education attained, relative to that of Uganda and Tanzania. There exist however many other possible reasons for the observed trends in mean educational attainment, both in Ethiopia and the countries used as controls. The trend observed using the difference-in-difference approach supports the findings from section 6.4.