Gottfried Wilhelm Leibniz Universität Hannover

Four Empirical Essays on Human Capital and Labour Market Outcomes

Der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktorin der Wirtschaftswissenschaften
- Doctor rerum politicarum -

genehmigte Dissertation

von

Master of Science, Bethlehem Asres Argaw

geboren am 19.11.1985 in Addis Ababa

2017

Referent: Prof. Dr. Patrick A. Puhani

Korreferent: PD Dr. Friedhelm Pfeiffer

Tag der Promotion: 1. März 2017

To my family

Acknowledgments

First and foremost, I would like to thank God for giving me the strength to persevere through the PhD journey and for blessing me with the right people that made the journey less nerve-wracking. His unconditional love and blessings are the engines of my life.

I am very thankful to my dissertation supervisor Prof. Dr. Patrick A. Puhani for his academic support and guidance. I had the opportunity to work closely with him in the project "Does Class Size Matter? Quasi-Experimental Evidence from Germany" where I learned from his invaluable knowledge and experience in a relaxed atmosphere. I also would like to express my sincere gratitude to my co-supervisor PD Dr. Friedhelm Pfeiffer for his continuous support and encouragement throughout my doctoral research. I am deeply grateful for the helpful advise not only on my academic work but also on life.

After I obtained my master's degree from the University of Copenhagen, I moved to Germany to join the Centre for European Economic Research (ZEW). This transition shaped my life and career in a way that was unthinkable for me at the time. I would like to thank the ZEW staff and especially my colleagues at ZEW's research department on Labour Markets, Human Resources and Social Policy for welcoming me into their team and assisting me to make a smooth transition. ZEW opened the door for my doctoral research by providing the academic and financial support any doctoral researcher could ask for. Financial support from the German Ministry of Education and Research (BMBF) under the framework "The Economics of Science" is also gratefully acknowledged.

My sincere gratitude is extended to Olga J. Skriabikova and Michael F. Maier with whom I worked on the paper "Risk Attitude, Job Mobility and Wage Growth". As PhD students, we spent countless, at times frustrating, hours shaping and reshaping our thoughts. I am very thankful for your collaboration, continuous discussions and persistent effort.

I had the privilege to write my paper on "Mother Tongue Education, Reading Skills and Labour Market Outcomes" at the United Nations University World Institute for Development Economics Research (UNU-WIDER) in Helsinki. During my time as a PhD Research Intern at UNU-WIDER, I was able to refocus my research on areas where my research output is needed the most. I met and exchanged knowledge, experience and laughter with incredible people from around the world. I would like to thank UNU-WIDER for the financial and academic support and for giving me the opportunity to meet the UNU-WIDER staff and fellow interns who made my time in Helsinki unforgettable.

Last but not least, none of what you are about to read would have been possible without the endless love, support and prayers of my family and friends. The short trips that I made to Ethiopia - when I felt lost, trapped and homesick - energized and brought me back on track. I would not have made it through the cold winter and the German bureaucracies without the love and support of the wonderful friends I made in Germany. I thank them for being the most important part of my life.

Bethlehem A. Argaw
November 2016
Mannheim

Abstract

This dissertation examines, in four chapters, the determinants and the labour market impacts of human capital using micro data. The first section of the dissertation studies institutional factors that influence individuals' health and educational outcomes. In the first chapter, the impact of Lenten fasting during pregnancy on the health outcomes of children is investigated using data from Ethiopia. The second chapter analyzes the effects of class size in elementary schools on educational outcomes based on administrative data from Germany. The second section of the dissertation deals with the impacts of human capital on success in the labour market. The thrid chapter deals with the influence of mother tongue education in primary schools in Ethiopia on individuals' ability to read and the subsequent labour market outcomes including employment status, type of occupation and earnings. In the last chapter, data from Germany is used to examine how individuals' attitude towards risk influences job mobility decisions and wage growth during the first few years of labour market experience.

Keywords: human capital, labour market oucomes, causal relations

Zusammenfassung

Die Dissertation untersucht in vier Kapiteln Determinanten und ausgewählte Arbeitsmarktergebnisse von Humankapital mit Mikrodaten. Der erste Teil der Dissertation befasst sich mit institutionellen Faktoren, die einen Einfluss auf Gesundheit und Bildungsergebnisse haben können. Im ersten Kapitel wird die Wirkung vom Fasten während einer Schwangerschaft auf die Gesundheit der betroffenen Kinder mit Daten aus Äthiopien erforscht. Das zweite Kapitel analysiert den Effekt der Klassengröße in der Grundschule auf Bildungsergebnisse mit administrativen Daten aus Deutschland. Der zweite Teil der Dissertation untersucht Wirkungen von spezifischen Dimensionen des Humankapitals auf Erfolge im Arbeitsmarkt. Das dritte Kapitel beschäftigt sich mit dem Einfluss, den Unterricht in der Muttersprache in Grundschulen in Äthiopien auf die Lesefähigkeit der Lernenden und den späteren Arbeitsmarkterfolg haben, darunter den Beschäftigungsstatus, den Beruf, und den Arbeitsverdienst. Im letzten Kapitel wird mit Daten aus Deutschland untersucht, wie die Risikoneigung sich auf Arbeitsplatzwechsel und das Wachstum der Arbeitsverdienste in den ersten Jahren nach dem Einstieg in den Arbeitsmarkt auswirkt.

Schlüsselworte: Humankapital, Arbeitsmarktergebnisse, kausale Wirkungen

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Human capital refers to the stock of skills, traits, health and experience that the labour force possess. It is considered an asset to an individual or a country as it directly or indirectly raises productivity in the labour market. It can also be an end in itself to the extent that it brings non-monetary benefits to the individual or a society. This dissertation focuses on the two major components of human capital - health and education - and examines certain institutional factors that influence human capital and its impacts on the labour market.

In Chapter 2, I examine the effects of Lenten fasting during pregnancy on the health outcomes of children in Ethiopia. Chapter 3 studies the effects of class size in elementary schools on the educational outcomes of students in the German state of Hesse. In Chapter 4, I look at the effects of accessing primary education in mother tongue on the reading skills and labour market outcomes of adults in Ethiopia. Chapter 5 examines how individual risk attitudes influence job mobility decisions and wage growth during the early career in Germany.

In this introduction, I provide a brief summary of the four papers that make up this dissertation. I also explain the contribution of the dissertation to the economics literature in general and to the literature on education, labour and development economics in particular.

Lenten Fasting During Pregnancy and Child Health Outcomes

The Great Lent is the most important fasting period for Orthodox Christians in Ethiopia and observing the fast is considered a test of one's Christianity. During the 55 days of the Great Lent, individuals avoid eating any food of animal origin including milk and egg and postpone their first meal until the liturgy is finished around noon. Even though pregnant and lactating women are exempted from fasting, the majority of women in Ethiopia observe the Great Lent due to various socio-cultural reasons. In Chapter 2, I examine the effects of observing the Great Lent during pregnancy on the health outcomes of children in Ethiopia.

Improving the health status of children in developing countries is at the forefront of national governments' and international organizations' policy agenda. The medical and economics literature documents that exposure to severe or mild nutritional deficiency during the prenatal and postnatal periods has negative consequences not only on health outcomes but also on long-term economic outcomes such as education and income. However, much of the available evidence from developing countries focuses mainly on the consequences of severe nutritional shocks to the fetal environment caused by famine or war. Such events are rare and hence put restriction on the policy implications that can be drawn from these studies. Chapter 2 of this dissertation fills the current gap in the literature by identifying the causal effects of prenatal exposure to mild nutritional shortage in the context of a developing country, Ethiopia. The chapter also complements the existing recent evidence on the consequences of Ramadan observance during pregnancy by examining the health impacts of changing not only the timing of maternal food intake - similar to Ramadan - but also reducing its nutritional content.

Ethiopia is a multireligious country where Orthodox Christians, Muslims, Protestants respectively make up 44, 34 and 19 percent of the population according to the 2007 census. This has a methodological advantage since it allows me to find a viable comparison group to disentangle the causal effect of maternal Lenten fasting during pregnancy from seasonal factors that affect child health outcomes. The quasi-exogenous variation in exposure to the Great Lent comes from differences in children's religion and month of birth. Orthodox Christian children whose mothers' pregnancy overlaps with the Great Lent (i.e., they were conceived, in utero or born during the Great Lent) are potentially exposed to nutritional shortage. A comparison is made with two control groups that were not exposed to the Great Lent in the womb. These are Orthodox Christians whose gestation period do not overlap with the Great Lent and children who follow a religion other than Orthodox Christianity. The empirical analysis is based on a sample of children between the age of 6 to 59 months taken from the Ethiopian Demographic and Health survey (EDHS) and the Ethiopian socioeconomic survey (ESS).

The findings show that Orthodox Christian children whose gestation months coincide with the Great Lent have significantly lower height-for-age z-scores compared to children who were not exposed to the Great Lent. Children who were in utero during the entire fasting period on average have about 10 to 21 percentage points higher probability of being too short for their age (stunt) and the height-for-age z-scores of children who were born during the Great Lent is between 0.322 and 0.463 standard deviation lower than children who were not exposed to the Great Lent. The negative consequences of fasting exposure on height-for-age is observed mainly for girls. These results are robust to a range of specification checks.

The results are useful in light of the Ethiopian government's and international organizations' goal to improve the health status of children. Since the Ethiopian Orthodox church allows women to avoid fasting during pregnancy, it is necessary to raise the awareness of the public (including religious leaders) in general and pregnant women in particular regarding the fasting rules of the church and the negative consequences of nutritional shortage during pregnancy. During antenatal care or home visits, health care professionals could monitor the nutritional status of pregnant women and refer fasting pregnant women to consult informed religious leaders when deemed necessary.

Does Class Size Matter? Quasi-Experimental Evidence from Germany

This is a joint work with Prof. Dr. Partick A. Puhani.

The policy of class size reductions has been at the center of educational research and policy debate for a few decades. The impact of class size on educational achievements is found to vary, among other dimensions, across school systems, grade levels, gender and students' socio-economic background. Whereas some empirical studies find a positive effect of smaller classes on short- and long-term outcomes, part of the literature finds no substantial benefits from class size reductions. The benefit of class size reductions in the context of an early school tracking system such as in Germany is scarcely studied. In Chapter 3, we provide evidence on the causal effects of class size using a newly available administrative data on about a quarter of a million students in the German state of Hesse. Although there exists some evidence on the relationship between class size and educational outcomes in Germany, to the best of the authors knowledge, there is only one paper (Wößmann, L. (2005): Educational production in Europe. Economic Policy, 20(43): 445-504) that uses a credible methodology to estimate the causal effects of class size in Germany. The large administrative data we use has important advantages over the TIMSS data used by Wößmann to estimate the class-size effect in Germany.

We measure academic performance based on the observation that students in Germany are tracked into more or less academic middle school types at the end of elementary school. The types of middle school that students are recommended to attend and the school type they eventually attend, to a large extent, depends on their academic performance in Math, German and General Studies in elementary school. We use an indicator for getting a recommendation to attend the higher and more academic school type called *Gymnasium* and the actual choice to attend this type of middle school as the main measures of students educational outcome. In order to identify causal effects, our empirical strategy exploits the random assignment of students into classes of different size as a result of the variation in total enrollment and the maximum students allowed in a class.

The rule-induced class size is used as an instrumental variable for actual class size in a fuzzy regression discontinuity framework.

The results show economically small and statistically insignificant effects of class size in 4^{th} grade on the probability of getting a recommendation or eventually attending the more academic school type in 4^{th} grade. The class-size effect is precisely estimated and the magnitude of the estimates is close to zero. Results are robust to several sensitivity checks. We also do not find statistically significant heterogeneous effects of class size by students gender, nationality, country of origin and language used at home. The potential explanation for the insignificant effect of class size we find could be that the average size of classes in Germany is quite small and hence it might be below the threshold that is relevant for students academic achievement. The result that smaller class size does not substantially improve educational outcomes is in line with the previous evidence for Germany.

Mother Tongue Education, Reading Skills and Labour Market Outcomes

Mother tongue instruction especially during the early grades of schooling has been advocated as a silver bullet to increase educational attainment, to improve learning outcomes and to reduce educational inequality across groups. Proponents of mother tongue education point that school curricula are better communicated when students are taught in their mother tongue. The main arguments put against mother tongue education is that it could reduce proficiency skills in the dominant language on the labour market thereby limiting employment and other economic opportunities in the long-term. The empirical evidence on the causal link between mother tongue instruction, language proficiency and long-term economic outcomes is rather mixed. Chapter 4 examines the causal relationship between mother tongue instruction, reading skills and early labour market outcomes using the 1994 policy change in the language of instruction in Ethiopia's primary schools.

Ethiopia is one of the few African countries that introduced mother tongue-based primary education. Prior to 1994, the language of instruction for most subjects during primary education was the official language (Amharic), although only 30 percent of the population have Amharic as their mother tongue. As from 1994, mother tongue-based education in public schools was mandated throughout primary education. Methodologically, the 1994 language reform offers an advantage since it allows me to exploit the variation across language groups over birth cohorts for identification. I use the quasi-exogenous variation in exposure to the language policy change in a difference-in-differences framework to identify, for the first time, the effects of the 1994 policy change on reading skills and labour market outcomes. The chapter also tests, for the first time, whether or not parents engage more in their child's education outside of schools when formal education is provided in mother tongue. Parental educational engagement is hypothesized to be one of the potential channel through which mother tongue education affects schooling outcomes. The empirical analysis is based on data from the 2011 Ethiopian Demographic and Health Surveys (EDHS) and the 2010 Early Grade Reading Assessment (EGRA).

The results reveal that the reform significantly improved the reading skills of birth cohorts that gained access to mother tongue education after 1994 thereby reducing the literacy gap between Amharic and non-Amharic language speakers by half. When I test for any significant differences in labour market outcomes, I find that individuals who received primary education in their mother tongue are more likely to be employed in skilled jobs and to receive their earnings in cash as opposed to informal means of payments. Two possible channels linking instruction based on mother tongue and an improvement in reading skills are identified. First, school enrollment is significantly higher among birth cohorts that gained access to mother tongue instruction after 1994. This implies that providing education in mother tongue increases the accessibility of schools. Second, parents are more likely to invest in their children's education outside schools when their children are taught in mother tongue. Parental investment takes the form of time investment where parents directly help children with homework and/or financial investment where parents hire a tutor to help their children with homework. The findings underscore the importance of mother tongue education to reduce ethno-linguistic inequality with respect to human capital and labour market outcomes.

Risk Attitudes, Job Mobility and Subsequent Wage Growth

This is a joint work with Michael F. Maier and Olga J. Skriabikova.

About two thirds of lifetime job changes occur during the first ten years of labour market experience. Job changes within the first few years on the labour market are important as decisions that are made in this period can strongly influence labour market prospects of the whole career. Changing jobs is a risky decision since it involves incurring substantial costs without entirely foreseeing future benefits. Hence, individuals attitude towards risk is a crucial factor to explain job changing behaviour and heterogeneous patterns of subsequent wage growth. Chapter 5 formulates a theoretical model and empirically tests, for the first time, the relationship between individual risk attitudes, job mobility and the subsequent wage growth during the early career.

The chapter tests two hypotheses on the relationship between risk attitudes, job mobility and the subsequent wage growth. First, since changing a job is a risky decision, risk-averse individuals make fewer job changes during their early career than more risk-tolerant individuals do. Second, since risk-averse individuals demand more compensation for the risk associated with changing jobs, the observed wage increases are on average higher for risk-averse than for more risk-tolerant individuals. Our empirical equations for job mobility and wage growth are estimated using ordinary least squares and fixed effect models respectively and we control for various factors and undertake different sensitivity checks to arrive at causal interpretation. The empirical analysis is based on data from the German Socio-Economic Panel Survey (SOEP). The SOEP is one of the few household panel surveys that contains a direct measure of individuals attitudes towards risk.

The estimation results show that individuals who are more risk-averse make fewer job changes compared to individuals who are more risk-tolerant. Being risk-averse in occupational matters reduces the average number of job changes experienced by 0.276, which is one third of a standard deviation. We then document a significant difference in the wage growth associated with the job changes of risk-averse individuals compared to more risk-tolerant individuals. Since risk-averse individuals demand higher compensation for the risk associated with changing jobs, their job changes are on average associated with relatively higher wage gains. A voluntary job is associated with 13 percent increase in wages whereas the job changes of risk-tolerant individuals is associated with a 6 percent increase in wages.

These empirical findings enhance our understanding of job mobility decisions during the early career in that individuals attitudes towards risk is an important behavioural trait that influence decisions in the labour market. In addition to the already established link between risk attitudes and labour market outcomes, such as occupational or educational choice, sector of employment, type of contract, the results of this chapter show that individuals attitudes towards risk plays a crucial role in early-career job mobility decisions. The implication of these findings is that heterogeneity with respect to risk attitudes across groups of individuals, for instance with respect to gender and migration background, matters in the relationship between job mobility and the pattern of wage growth observed during the early career.

Chapter 2

Lenten Fasting During Pregnancy and Child Health Outcomes

BETHLEHEM A. ARGAW (Leibniz Universität Hannover)

The author gratefully acknowledges free access to the Ethiopian Demographic and Health Survey (EDHS) from the DHS program and the Ethiopian Socioeconomic Survey (ESS) from the World Bank's Living Standards Monitoring Studies (LSMS). The views expressed in this article are those of the author and do not necessarily reflect the views of the respective organizations. I would like to thank Patrick Puhani, Friedhelm Pfeiffer and Abraham Asfaw for helpful comments and discussions. Any remaining errors are my sole responsibility.

Abstract

Recent studies based on the observance of Ramadan find that prenatal exposure to nutritional shortage due to maternal fasting has negative consequences on short- and long-term economic outcomes. Contrary to Ramadan, where not much restriction is put on food intake after breaking the fast, Ethiopian Orthodox Christians are vegan for 55 days during the Great Lent. This creates a condition where fasting by pregnant women not only affects the timing of food intake - similar to Ramadan - but also reduces its nutritional content. This paper uses the multireligious context of Ethiopia for identification in a difference-in-differences setup to examine the consequences of in utero exposure to the Great Lent on children's health outcomes. Results show that children whose mothers' pregnancy overlap with the Great Lent have significantly lower height-for-age. Children who were in utero during the entire fasting period on average have about 10 to 21 percentage points higher probability of being too short for their age (stunt) whereas the height-for-age z-scores of children who were born during the Great Lent is between 0.322 and 0.463 standard deviation lower than children who were not exposed to the Great Lent. The results are robust to the inclusion of various controls including seasonality of health outcomes and maternal characteristics and to a range of specification checks.

Keywords: Fasting during pregnancy, Lent, nutritional shortage, child health, Ethiopia

JEL classification: I10, I12, Z12

2.1 Introduction

Empirical studies have documented the negative consequences of exposure to nutritional deficiency during the prenatal and postnatal periods on short- and long-term economic outcomes (e.g. Lumey et al., 2011; Currie and Vogl, 2012, for a review of the literature). In the absence of real world experimental evidence, natural experiments have been widely used to identify the causal effects of nutritional shortage on economic outcomes. Prenatal and postnatal exposure to severe nutritional shock caused by famine and diseases have negative effects on adult height (Chen and Zhou, 2007; Dercon and Porter, 2014), educational attainment (Barreca, 2010; Neelsen and Stratmann, 2011), employment and earnings (Almond, 2006; Nelson, 2010; Scholte et al., 2015), among others. As a relatively mild nutritional shock to the fetus environment, recent studies use maternal fasting during the Islamic holy month of Ramadan as a natural experiment. Compared to individuals who were not exposed to Ramadan fasting during their mother's pregnancy, exposed individuals have lower birth weights (Almond and Mazumder, 2011), lower academic performance (Almond et al., 2014), higher probability of disability (Almond and Mazumder, 2011), poor health in old age (van Ewijk, 2011) and lower labour market productivity (Schultz-Nielsen et al., 2016).

The mechanism linking in utero nutritional shock to economic outcomes identified using Ramadan fasting is through the change in the timing of maternal food intake. During the observance of Ramadan, Muslims abstain from any kind of food or drink until sunset without having any restriction on the type of food intake after breaking the fast. Being important on its own, altering the timing of food intake, however, might not greatly influence pregnant women's nutrition intake. During the Great Lent, Ethiopian Orthodox Christians postpone their first meal or drink until the Liturgy is finished around 3 p.m. and abstain from any kind of animal products for 55 days.[1] Among the very devoted, only one meal per day is allowed. This creates a condition where fasting by pregnant

[1] The exception is on weekends when the Liturgy is finished around 9 a.m.

women not only affects the timing of food intake - similar to Ramadan - but also reduces its nutritional content. The biomedical literature show that skipping meals and the lack of essential vitamins and minerals during pregnancy have detrimental impacts on infant growth and development both before and after birth (e.g. von Schenck et al., 1997; Innis, 2007; Dror and Allen, 2008). We use the observance of the Great Lent among Orthodox Christians in Ethiopia to examine the effects of prenatal nutritional shock on children's health outcomes.

By using the observance of the Great Lent among Orthodox Christians in Ethiopia, the paper contribute to the literature in three ways. First, since Orthodox Christians become vegan during the Great Lent, we can identify the effects of altering both the timing and the content of nutrition intake during pregnancy. Second, unlike the four-weeks fasting period during Ramadan, the Great Lent lasts for almost eight weeks in Ethiopia.[2] This makes it possible to examine the effects of a relatively longer in utero exposure to mild nutritional shock. Third, Ethiopia is very suitable to undertake the analysis since it is a multireligious country where Orthodox Christians, Muslims, Protestants respectively make up 44, 34 and 19 percent of the population (CSA, 2008). This allows us to find a viable comparison group to measure the consequences of maternal fasting on children's health outcomes using the observance of both the Great Lent and Ramadan. In the absence of data on large birth cohorts, finding a reliable control group is a big challenge in previous Ramadan studies based on data from western countries, where Muslims make up a small share of the population (Almond et al., 2014).

The observance of Ramadan follows the lunar calendar which moves forward by about 11 days each year in the Gregorian calendar and hence completing the Gregorian calendar over 32 years. Based on data on large Muslim birth cohorts, previous studies use this variation in the observance of Ramadan to identify the consequences of maternal fasting after controlling for seasonality in health outcomes. The Great Lent, however,

[2] The Great Lent is observed for 48 days among Orthodox Christians in other countries. This consists of the 40 days that Jesus Christ fasted and the Holy Week, which is the week before Easter. In addition to the 48 days, Ethiopian Orthodox Christians fast during the first week preceding the fasting of Jesus Christ.

always falls in the months between February and May. Hence, we rely on a difference-in-differences (DID) strategy to control for seasonal variation in health outcomes. The quasi-exogenous variation in exposure to the Great Lent comes from differences in children's religion and month of birth. Orthodox Christian Children whose mothers' pregnancy overlaps with the Great Lent (i.e., they were conceived, in utero or born during the Great Lent) are potentially exposed to nutritional shortage. A comparison is made with two control groups that were not exposed to the Great Lent. These are Orthodox Christians whose gestation period do not overlap with the Great Lent and children who follow a religion other than Orthodox Christianity. Since we do not observe in our data whether or not Orthodox Christian pregnant women indeed fast during the Great Lent, we interpret the results as the "intention-to-treat" (ITT) effects.

Our estimation sample comes from the 2011 Ethiopian Demographic and Health Survey (EDHS) and the 2014 Ethiopian Socioeconomic Survey (ESS). Both the EDHS and the ESS provide data on the exact date of birth and height of children under the age of five. We measure children's health outcomes using their height conditional on their age and gender. Height-for-age reflects failure to receive adequate nutrition over a long period of time and the score does not vary according to recent and temporary shortage of food intake. Children whose height-for-age z-score is below minus two standard deviations from the median of the reference population are considered too short for their age (stunted). The reference population, defined by the World Health Organization, consists of randomly selected sample of healthy infants and children children living in six countries (Brazil, United States, Norway, Ghana, Oman, and India).

We use the timing of the EDHS and ESS survey interview dates to shed some light on the influence of the Great Lent on household food consumption patterns since the fasting status of women during pregnancy is not observed in the data. We examine the differences in the consumption of non-fasting food item across religion depending on whether or not the survey interviews overlap with the Great Lent. Compared to households that follow a religion other than Orthodox Christianity, we find that Orthodox

Christian households have about 10 to 27 percentage points less likelihood of consuming non-fasting foods such as milk, egg, cheese and meat in survey interviews that overlap with the Great Lent compared to interviews that do not overlap with the Great Lent. More importantly, we find that Orthodox Christian mothers have about 10 to 14 percentage points less probability of feeding their children under the age of two non-fasting food items in survey interviews that overlap with the Great Lent compared to interviews that do not overlap with the Great Lent. Since children under the age of seven are exempted from fasting, the type of food that women feed their children under the age of two indirectly shows the type of food available in the household for consumption during the Great Lent.

The difference-in-differences estimation results show that children whose mothers' pregnancy overlap with the Great Lent have significantly lower height-for-age. Children who were in utero during the entire fasting period on average have about 10 to 21 percentage points higher probability of being too short for their age (stunt) compared to children who were not exposed to the Great Lent. Relative to the average stunting rate of 43 percent, the effect implies a 26 to 49 percent increase. The height-for-age z-scores of children who were in late gestation during the Great Lent is between 0.322 and 0.463 standard deviation lower than children who were not exposed to the Great Lent. This represents declines of 19 and 27 percent compared to the average height-for-age Z-scores. The negative consequences of fasting exposure on height-for-age is observed mainly for girls. The point estimates we find are very similar to Akresh et al. (2012) who show that children who were in utero or born during the 1998-2000 Ethiopian-Eritrean war and were living in the war region in Ethiopia experienced 0.45 standard deviations lower height-for-age z-scores.

The results are robust to various sensitivity checks. We account for the potential selective timing of pregnancy in response to the observance of Great Lent by controlling for observable maternal characteristics such as maternal education, literacy skills, height, weight and health knowledge. We also show that there is no systematic relationship between observable maternal characteristics and whether or not a child's gestation period overlaps with the Great Lent. Furthermore, since we identify the effects of in utero exposure to the Great Lent by using Muslims as part of the control group, excluding Muslims whose gestation period overlaps with Ramadan or all Muslims from our control group does not alter the results. This is mainly because we find that the observance of Ramadan has insignificant effect on Muslim children's health outcomes.

The results are useful in light of the Ethiopian government's and international organizations' goal to improve the health status of mothers and children. The available evidence (e.g. Gebremedhin et al., 2015) show that pregnant women fast during the Great Lent not only due to lack of awareness regarding the rules of the church but also due to fear of rejection by the community including religious leaders. Since the Ethiopian Orthodox church allows women to avoid fasting during pregnancy, it is necessary to raise the awareness of the public (including religious leaders) in general and pregnant women in particular regarding the fasting rules of the church and the negative consequences of nutritional shortage on their new born. During antenatal care or home visits, health care professionals could monitor the nutritional status of pregnant women and refer fasting pregnant women to consult informed religious leaders when deemed necessary.

The rest of the paper is structured as follows. Section 2.2 provides an overview of the literature on the short- and long-term impacts of prenatal exposure to nutritional shortage. It also provides institutional background information on fasting in Ethiopia by focusing on the fasting and food consumption patterns of pregnant women and household members in general during the Great Lent. In Section 2.3, we describe the data and the estimation strategy. Results are presented and discussed in Section 2.4 whereas we provide concluding remarks in Section 2.5.

2.2 Background

2.2.1 Literature Overview

Since the seminal work of the British physician and epidemiologist David J. Barker in the 1990s, it is increasing being recognized that nutritional deficiency during the prenatal period has significant impact on health outcomes from infancy to adulthood. The fetal origin (fetal programming) hypothesis posits that the structure and function of different body organs undergo programming during this critical period. Nutritional shock during this period results in "developmental adaptations that produce permanent structural, physiological, metabolic, and epigenetic changes thereby predisposing an individual to adult cardiovascular, metabolic, and endocrine diseases" (Lau et al., 2011). Poor or inadequate nutrition during early life development could directly impact the development of the brain and other important body organs exposing an individual to physical and mental health problems including cognitive function late in life (Lau et al., 2011; de Rooija et al., 2010).

Following Barker and Osmond (1986)'s initial finding on the association between poor nutrition in utero, infant mortality and Ischaemic heart disease in England and Wales, the fetal origin hypothesis has been tested for different health outcomes including obesity, hypertension and diabetes. Since real-life human experiment is unethical in this context, experimental studies in animals provide support for the hypothesis (Langley-Evans, 2006). In addition, studies based on micronutrient variation in pregnant women find effects of fetal nutrition on childhood cognitive function (Bhate et al., 2008). The economics literature has expanded the topic further by using credible causal identification methods and examining a range of outcomes including educational attainment, employment status and earnings, among others (see Almond and Currie, 2011; Currie and Vogl, 2012, for review).

In the absence of real world experimental evidence, natural experiments have been widely used to identify the causal effects of prenatal and infant exposure to nutritional shortage. Exposure to famine and diseases resulting from natural or human causes are the most commonly used natural experiments in the literature (see Lumey et al., 2011, for review). This includes the Great Finnish Famine of 1866-68 (Doblhammer et al., 2013), the 1918 Influenza Pandemic (Almond, 2006; Nelson, 2010), the Leningrad Blockade of 1941-1944 (Stanner et al., 1997; Sparén et al., 2003), the famines in European countries in the 1940s (Roseboom et al., 2001; Neelsen and Stratmann, 2011; van den Berg et al., 2016), the Chinese Great Leap Forward famine of 1959-1961 (Chen and Zhou, 2007; Almond et al., 2007; Mu and Zhang, 2011), the 1984 famine in Ethiopia (Dercon and Porter, 2014) and Malaria exposure in the US (Barreca, 2010).

Severe nutritional shocks in utero or during early childhood as a result of exposure to famine or disease is found to significantly influence adult health including physical disability (Almond, 2006; Mu and Zhang, 2011), obesity (van den Berg et al., 2016), hospitalization rates (Scholte et al., 2015) , height (Chen and Zhou, 2007; Dercon and Porter, 2014; van den Berg et al., 2016), life expectancy (Doblhammer et al., 2013), mortality (Sparén et al., 2003) and socio-economic outcomes such as educational attainment (Barreca, 2010; Neelsen and Stratmann, 2011; Dercon and Porter, 2014), literacy skills (Almond et al., 2007; Nelson, 2010; Mu and Zhang, 2011), employment outcomes (Almond, 2006; Nelson, 2010; Scholte et al., 2015), among others.

Exposure to famine and diseases during in utero or early childhood results in severe nutritional shocks. Such events are rare and they have devastating impacts on mortality. This creates methodological and practical challenges. The methodological challenge arises duet to the survival bias (Doblhammer et al., 2013). That is, individual who survive severe nutritional shock resulting from exposure to famine and diseases may not be a random sample of the population. Since famine and the outbreak of diseases are rare events, it also imposes limitation on the policy implications that can be drawn from the estimated effects.

Recent studies in the economics literature use maternal fasting during the Islamic holy month of Ramadan as a natural experiment to identify its effects on later life outcomes. In utero nutritional shortage arising from maternal fasting is a relatively mild shock to the fetus environment compared to exposure to famine and diseases. Hence, important policy implication can be drawn by identifying its consequences on health and other economic outcomes. Almond and Mazumder (2011) is among the first studies that used Ramadan exposure as a natural experiment to examine the impact of mild nutritional shortage on birth weight, the sex ratio and adult health outcomes. The authors find lower birth weights and a 6.1 percentage points lower likelihood of a male birth among Arabs exposed to maternal Ramadan observance around the time of conception using natality data from Michigan. The authors also find a 20 percent likelihood of disability among Muslims in Uganda and Iraq when Ramadan overlaps with early pregnancy.

Following Almond and Mazumder (2011), the possible negative consequences of in utero exposure to Ramadan have been examined by various studies in different contexts. For instance, van Ewijk (2011) uses data from Indonesia and finds that individuals who were prenatally exposed to Ramadan are more likely to report symptoms of coronary hear problems and type 2 diabetes higher and have lower general health outcomes in adulthood and at old age. On the other hand, Juerges (2015) does not find any significant effects of Ramadan exposure on the sex-ratio and the birth weight of Muslims born in Germany and hence, questioning the generalizability of results from previous studies. Among the few studies that go beyond health outcomes, Almond et al. (2014) find a 0.05 to 0.08 standard deviation lower test scores for children who were exposed to Ramadan during early pregnancy in England. Similarly, Oosterbeek and van der Klaauw (2013) find that one additional Ramadan week lowers the final grade of Muslim economic students at VU University Amsterdam in the Netherlands by almost 10 percent of a standard deviation.

As our review of the evidence shows that the wave of recent studies that examine the consequences of mild nutritional shock is mainly based on data from developed countries.

As to the author's knowledge, there are two studies in the economics literature that examine the consequences of nutritional shortage in the context of Ethiopia. Both of these studies examine the effects of severe nutritional shortage caused by famine (Dercon and Porter, 2014) and war (Akresh et al., 2012). The first paper estimates the impact of the 1984 Ethiopian Famine, which led to the death of more than half a million people, on the economic outcomes of adults (17-25 years old) who were in utero or infant during the peak of the famine. The authors find that children under the age of 36 months during the famine are 3 cm shorter, less likely to have completed primary school, and more likely to have experienced recent illness as adults. The other major event that occurred in recent history of the country is the war between Ethiopia and Eritrea that took place from 1998 to 2000. Akresh et al. (2012) measure the conflict's impact on children's health outcomes exploiting the quasi-exogenous variation in exposure to the war across time and space. The authors find that children who were exposed to the war in both countries have lower height-for-age Z-scores. The short- and long-term consequences of mild nutritional shock in the context of Ethiopia is yet to be explored.

2.2.2 Institutional Setting

Ethiopia is said to be one of the oldest Christian nations. The country adopted Christianity as the state religion during the 4th century A.D. (Selassie and Tamerat, 1970). Until the early 1970s, the country had a monarchy system and its kings and queens originated from regions where the Orthodox Christian church played a dominant role. After the fall of the last king of Ethiopia, Emperor Haile Selassie, there is a clear distinguish between the state and religion. The majority of Ethiopians follow Orthodox Christianity though the share of its followers declined from 51 percent in 1994 to 44 percent in 2007. The second largest religion is Islam and its share remains stable around 33 percent. The third largest religious groups are protestants whose share increased from 10 percent in 1994 to about 19 percent in 2007. The rest of the population follows either Catholicism or traditional religions (CSA, 1994, 2008).

The Ethiopian Orthodox Christian has approximately 220-293 fasting days in a year. It is, however, the very devout who are expected to strictly follow all the fasting periods. For an ordinary Christian, the fasting days are approximately 110-150 days per year (Knutsson and Selinus, 1970). There are seven official fasting periods and these are the Great Lent (*Abiy Tsom* - for 55 days), the fast of the Apostles (*Yehawariayat Tsom* - its duration various between 14-44 days depending on the date of Easter), the fast of the Prophets (*Yenebiyat Tsom* - for 43 days), the fast of the Assumption of St. Mary (*Filiseta* - for 15 days), the fast of Nineveh (*Tsome Nenewe* for 3 days), the fast of Revelation (*Tsome Gahad* - for one day) and the fast of Salivation (all Wednesday and Fridays except for the 50 days after Easter).[3]

As the name indicates, the Great Lent has significant importance than any other fast and is observed among the majority of Orthodox Christians. It represents the fasting of Jesus Christ and its longer duration is considered a test of one's Christianity.[4] During the 55 days of the Great Lent, individuals avoid eating any food of animal origin including milk and egg and postpone the first meal until the Liturgy is finished around noon.[5] Those exempted from fasting are pregnant and lactating women, individuals with serious illness and children under the age of seven. The significant importance of the Great Lent can also be observed through the meltdown it creates in the demand and supply of animal products. The majority of Orthodox Christian butcher remain closed during the Great Lent (Boylston, 2013; Seleshe et al., 2014). Studies also find reductions in the price of chicken (Dinka et al., 2010), sheep and goat (Teklewold et al., 2009), and the consumption and production of meat in general (Betru and Kawashima, 2009) during the fasting season.

---

[3] The information is accessed from the Ethiopian Orthodox Church Association - Mahibere Kidusan at `http://eotcmk.org/site-en/index.php?option=com_content&task=view&id=56&Itemid=1`.

[4] Although the duration of the fast of the Prophets and the Apostles are comparable to the Great Lent, these fasting periods are compulsory only for the clergy.

[5] Eating pork is forbidden even on non-fasting days. Some fasting observant consume fish during fasting periods.

There is no nationwide survey undertake in Ethiopia to provide evidence on the extent of fasting among pregnant women. Following, we review the available evidence on the severity of the problem. Most of these studies are based on qualitative data and some quantitative research also exist albeit based on small sample size. Knutsson and Selinus (1970) is the first anthropological and nutritional studies to examine the nutritional consequences of fasting among Ethiopian Orthodox Christians. When describing the extent of fasting among pregnant and lactating women, the authors write *"Although the church recognizes the right for pregnant and lactating women to ask, and be given, permission to eat animal food - for instance, to drink milk - in practice this is only occasionally made use of among the illiterate people. Knowledge about the rights of exemption is often not widespread."*. After comparing the diets of children aged between six months and three years during fasting and non-fasting periods, the study finds that Orthodox Christian children consume significantly lower nutrition in terms of calories, protein, calcium and riboflavin nutrients during fasting periods.

Zepro (2015) examines food taboos and misconception among pregnant women in the southern part of the country. The author interviewed 295 pregnant women attending antenatal care (ANC) in public health institutions and finds that 38.3 percent of the interviewed pregnant women reported to fast during pregnancy. Since women attending ANC are likely to be better educated, one can presume that the share of pregnant women who reported to fast during pregnancy would have been higher if women who do not attend ANC were included in the sample. Unfortunately, the author does not provide the distribution of fasting pregnant women across religion. This makes it impossible to get an approximate figure on the share of Orthodox Christians who fast during pregnancy.

With the aim of identifying maternal associated factors of low birth weight, Gebremedhin et al. (2015) interviewed 308 mothers (of which 86 percent are Orthodox Christians) within six hours after they gave birth in a hospital in northern Ethiopia. The author undertook three focus group discussions to study socio-cultural patterns of feeding practices during pregnancy including fasting during pregnancy. The responses of the mothers, some of which are displayed below, show that mothers in the northern part of the country do practice fasting during pregnancy.

> *"Older parents and religious rules will never accept you to eat non fasting foods whatever your pregnancy status. Rather, they strongly advocate to fasting and have strong link with the church. If you did not obey this, it is considered as a sinful act."*
>
> *"If someone disobeys the religious rules (if a woman is not fasting), either the mother will face problems during child birth or the child will be born with ill health. That's why I prefer to stick with it because I do not want to have unhealthy baby."*
>
> *"I was not fasting all the fasting hours during the course of my pregnancy but I never took non fasting foods like meat, eggs and milk during the fasting time because I do not need to breach the rules of my religion and isolate myself from my family for this reason."*
>
> *"I will never ask my religious father to such a permission (to having non fasting foods) while pregnant because such permissions are allowed for conditions beyond your capability and for seriously ill individuals."*

The timing of household surveys undertaken in the country allows us to shed some light on the influence of the Great Lent on household food consumption patterns. In the 2014 Ethiopian Socioeconomic Survey (ESS), data collection on household characteristics took place between February and April. The Great Lent in 2014 was observed between February 25 and April 20. Since the observance of the Great Lent overlaps with the survey interview, we can examine the differences in food consumption patterns across religion in fasting and non-fasting periods. ESS gathers information on the type of food consumed in the household over the past one week preceding the interview.

Panel A in Table 2.1 shows the relationship between religion and the probability of non-fasting food consumption in the household during the past one week preceding the survey interview. The outcome variables are binary indicators showing whether or not at least one member of the household consumed non-fasting foods such as milk, cheese, egg and meat. The variable of interest is the interaction between being Orthodox Christian and the survey interview overlapping with the fasting period. Interviews that were undertaken between February 25 and April 20 overlaps perfectly with the Lent. If the Lenten observance affects the availability of non-fasting foods in Orthodox Christian households, we expect the interaction term to have a negative coefficient.

Table 2.1: Influence of the Great Lent on household food consumption patterns

	Milk	Cheese	Egg	Meat	Water
	(1)	(2)	(3)	(4)	(5)
Panel A: At least one member of the household consumed the food item over the past week preceding the interview					
Orthodox Christian	-0.077***	0.107***	0.043**	0.116***	
	(0.027)	(0.020)	(0.022)	(0.022)	
Lent overlaps with interview	0.139	0.073	-0.142	0.136	
	(0.180)	(0.119)	(0.153)	(0.248)	
Orthodox Christian * Overlap	-0.153***	-0.097**	-0.267***	-0.252***	
	(0.055)	(0.041)	(0.048)	(0.040)	
Mean of dependent variable by religion					
Orthodox Christians	0.222	0.127	0.170	0.245	
Other religion	0.430	0.084	0.163	0.148	
No. of observations	4,895	4,895	4,895	4,895	
Panel B: Mother fed the child the food item the night or day preceding the interview					
Orthodox Christian	-0.137***	0.064**	-0.071***	0.078***	-0.008
	(0.022)	(0.026)	(0.020)	(0.022)	(0.029)
Lent overlaps with interview	-0.212	-0.150	0.132	0.549*	-0.137
	(0.536)	(0.316)	(0.346)	(0.303)	(0.515)
Orthodox Christian * Overlap	-0.140***	-0.102***	0.102***	-0.095***	-0.046
	(0.042)	(0.034)	(0.031)	(0.026)	(0.056)
Mean of dependent variable by religion					
Orthodox Christians	0.113	0.070	0.070	0.059	0.801
Other religion	0.244	0.098	0.070	0.025	0.761
No. of observations	4,929	4,929	4,929	4,929	4,929

Note: Estimation: OLS. Regression includes controls for month of birth, dummies for age in months, gender, rural-urban residence, region of residence, father's and mother's educational and occupation status, maternal health status, household access to safe drinking water and toilet facility. Standard errors, clustered at month of birth, year of birth and their interaction, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011 on panel B and the Ethiopian Socioeconomic Survey (ESS) 2014 on panel A.

As expected, Orthodox Christian households are less likely to consume non-fasting foods during the Great Lent. When fasting period coincides with the survey interview, we find that the consumption of non-fasting foods such as milk, egg, cheese and meat declines among Orthodox Christian households by 10 to 27 percentage points.

Panel B in Table 2.1 shows similar analysis based on the 2011 Ethiopian Demographic and Health Survey (EDHS). Similar to the ESS, data collection in the 2011 EDHS overlaps with the Great Lent. Data collection took place between December 27, 2010 to June 3, 2011. The Great Lent in 2011 was observed between February 28 until April 24. Interviews that were undertaken on March and April overlaps perfectly with the Great Lent which was observed between February 28 until April 24. EDHS allows us to examine the food consumption pattern of children under the age of two. Specifically, women with children under the age of two are asked about the type of food they fed to their children the day or night preceding the interview. If the observance of the Great Lent affects the availability of non-fasting foods in Orthodox Christian households, we expect to observe a negative correlation between being Orthodox Christian and the probability of mothers giving non-fasting food items to their children under the age of two. The outcome variable are indicators for whether or not the mother gave milk, cheese, egg or meat to the child the night or day before the interview.

Similar to the result on the food consumption of household members, we find that children under the age of two in Orthodox Christian households are less likely to be given non-fasting food items by their mother during the Great Lent. Since the availability of plain water does not depend on fasting observance, we find no significant correlation between religion and the probability of mother's giving plain water to their children under the age of two as shown in column (5). Though women are asked about their pregnancy status in the EDHS, there is no information available on their own food consumption pattern. Nevertheless, since children under the age of seven are exempted from fasting, the type of food that women feed their children under the age of two indirectly measures the type of food available in the household for consumption during the Great Lent.

The analysis so far show that Orthodox Christians are less likely to consume non-fasting food during the Great Lent. In addition, even though children under the age of seven are exempted from fasting, we find that Orthodox Christian mothers are less likely to feed non-fasting foods to their children under the age of two during the Great Lent. If mothers are less likely to feed their children non-fasting foods during the observance of the Great Lent, it is likely that the same food restriction is put on pregnant women. Indeed, the available evidence based on qualitative and quantitative data show this to be the case. Due to the meltdown the observance of the Great Lent creates in the economy, even those pregnant women who want to consume non-fasting food items could not easily access them during the fasting period.[6]

The next question we shed some light on is whether or not fasting by pregnant women affects their nutrition intake. There are two features of observing the Great Lent that could be detrimental to maternal and fetal health. The first is postponing the first meal until around 3 p.m. Much of the evidence on the health consequences of skipping meal come from the biomedical literature on Ramadan. Accelerated starvation due to skipping meals is found to lower glucose concentrations, diminished fetal cognitive function, neurological impairments, fetal breathing movements and heart rate accelerations (see citations in Almond and Mazumder, 2011). Though the time until the first meal during Lent is not as long as Ramadan, studies also show that mild starvation due to skipping breakfast during pregnancy results in biochemical changes that are detrimental to the fetal health (Metzger et al., 1982).

[6] It is important to notice that, due to the overlap of the survey interviews with the Great Lent, the EDHS and ESS underestimate the average consumption of non-fasting food items among Orthodox Christians. These surveys are among the commonly used data sources to measure food consumption patterns and calorie consumption at the national and regional level. Since Orthodox Christianity is the main religion in specific groups (e.g. more than 85 percent of the population in Tigray and Amhara regions are Orthodox Christians), studies have to take into consideration the time of survey interviews when comparing food consumption patterns and calorie consumption across groups (e.g. regions).

The second feature of fasting observance among Orthodox Christians in Ethiopia is the complete avoidance of animal products during fasting periods. Avoiding or low consumption of animal products leads to deficiency in essential vitamins and minerals which are necessary for normal fetal growth. Vegan pregnant women are at high risk of B12 deficiency since vitamin B12 is naturally found only in animal products (Dror and Allen, 2008). Avoiding animal products also affects the level of vitamin B12 concentration in breast milk (von Schenck et al., 1997). Vitamin B12 is essential for red blood cell formation, the proper functioning of the brain and nervous system, and DNA synthesis (Dror and Allen, 2008). Lack of vitamin B12 during pregnancy is associated with persistent neurological damage in infants (von Schenck et al., 1997; Black, 2008). von Schenck et al. (1997) shows that the cognitive and language development of infants with vitamin B12 deficiency remained severely low at the age of two years even after Vitamin B12 supplementation.

Maternal vegetarian diet is also found to lower the concentration of essential fatty acid in breast milk and hence in infants (Innis, 2007). DHA and other essential fatty acid support infant growth and development both before and after birth. Innis (2007) report that *"the most important factor determining the secretion of DHA in human milk is the mother's intake of DHA"*. In advanced countries, it is possible for a vegan pregnant woman to obtain Vitamin B12, essential fatty acid or other vitamin and minerals which promote optimal fetal growth through fortified foods or dietary supplements. In the context of Ethiopia, where 62 percent of women are illiterate and only 35 percent of pregnant women use antenatal care (DHS, 2012), neither of the options seem feasible and animal products remain the main sources of essential vitamins and minerals for pregnant women.

2.3 The Data and Method

2.3.1 The Data

Two data sources are used in the empirical analysis. The first is the Ethiopian Socioeconomic Survey (ESS). The survey collects data on a range of household and community level characteristics. The first wave was implemented in 2011-12 and the second wave was implemented in 2013-14. We base our analysis on the second wave because the first wave was only collected in rural areas and hence it is not nationally representative. Our second data source is the Ethiopian Demographic and Heath Survey (EDHS) which is a national representative repeated cross sectional survey of around 20,000 households and their members. We use the latest EDHS which was undertaken in 2011. Both data sets contain a wealth of information on socioeconomic characteristics at the individuals and household level. EDHS also gathers detail information on maternal characteristics including maternal literacy skills, health status, and health knowledge.[7]

Exact date of birth is one of the vital information required to identify individuals' exposure to the Great Lent. Compared to the available surveys such as the national census and labour force surveys, ESS and EDHS are ideal to undertake this study since they contain information on exact date of birth.[8] In both data sets, information on exact date of birth is only available for children aged between 0 and 59 months. Hence, our analysis examines the effects of in utero nutritional shock on children's health outcome. As in most developing countries, only a limited number of individuals in Ethiopia have birth cards which contain information on their exact date of birth. This is mainly because most births still occur at home (DHS, 2012). Hence, data on children's exact date of birth is collected during the survey interview by asking members of the household, in most cases, the child's mother. Measurement error in children's date of birth is likely to be smaller since parents are more likely to recall the date of birth of their small children

[7] Source for ESS: `http://go.worldbank.org/ZK2ZDZYDD0` and EDHS: `http://dhsprogram.com/`.

[8] Though the Young Lives Survey collects data on exact date of birth, we do not use it in this paper since it covers very few birth cohorts and the survey is not nationally representative.

compared to recalling their older children's date of birth. There are 11,067 children between the age of 5 and 59 months with valid data on exact date of birth. Three-fourth of our estimation sample comes from the EDHS.

Information on three key variables is required to calculate whether or not a child was in utero during the Great Lent. These are exact date of birth, the Great Lent fasting days from 2005 to 2014 and an average length of the gestation period. Following the medical literature, we assume that the average of length of pregnancy is about 266 days. Based on this information, we define a treatment variable which takes three values indicating whether or not a child was conceived during the Great Lent, in utero during the entire Great Lent or born during the Great Lent. Orthodox Christian Children who were conceived after the Great Lent and born before the next Great Lent and children following a religion other than Orthodox Christianity are in the control group.

We measure children's health outcomes using their height conditional on age and gender. Height-for-age is an indicator of linear growth retardation and cumulative growth deficits. Height-for-age reflects failure to receive adequate nutrition over a long period of time and does not vary according to recent and temporary shortage of food intake. In order to determine the nutritional status of children, the height of the child in question is compared with an international reference population. The reference population is defined by the World Health Organization and is composed of randomly selected sample of healthy infants and children children living in six countries (Brazil, United States, Norway, Ghana, Oman, and India). Height-for-age z-score is defined as the difference between the height of a child and the median height of the reference population for the same age and gender divided by the standard deviation of the reference population. Children whose height-for-age z-score is below minus two standard deviations from the median of the reference population are considered short for their age (stunted)

Our estimation sample becomes 10,372 after dropping children who have missing data on Height-for-age z-scores. Table 2.2 provides summary statistics for key variables.

Table 2.2: Summary statistics

	Overall	Differences across religion	
		Orthodox Christian	Other religion
	(1)	(2)	(3)
Height-for-age z-score	-1.70	-1.74	-1.68
	(1.52)	(1.40)	(1.60)
1 if child is stunt	0.43	0.43	0.42
Age in months	32.88	33.06	32.76
	(15.67)	(15.60)	(15.76)
1 if child is a girl	0.49	0.49	0.49
1 if rural residence	0.87	0.82	0.91
1 if father completed atleast primary education	0.14	0.12	0.16
1 if father completed atleast secondary education	0.07	0.08	0.06
1 if mother completed atleast primary education	0.10	0.12	0.08
1 if father works in the non-agriculture sector	0.21	0.24	0.20
1 if mother works in the agriculture sector	0.29	0.37	0.23
1 if mother works in the non-agriculture sector	0.27	0.27	0.27
Observations	10,372	3,696	6,676

Note: Standard deviation are in parentheses.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011 and the Ethiopian Socioeconomic Survey (ESS) 2014.

The nutritional status of children in Ethiopia is quite low compared to children of similar age in the reference population. On average, children are 1.70 standard deviations below the average height-for-age of a child in the reference population. Forty three percent of children in the country are stunted. There is no significant difference in children nutritional status between Orthodox Christians and other religion followers. There are, however, differences across religion in terms of household characteristics. About 82 percent of Orthodox Christians live in rural areas while the figure is at 91 per cent for households that follow other religions. Orthodox Christians are also better educated and have better outcomes in the labour market. About eight percent of Orthodox Christian fathers completed at least secondary education and 24 percent works in the non-agriculture sector. The respective figures for non-Orthodox Christian fathers are at 6 and 20 percent. Similarly, 12 percent of Orthodox Christian mothers completed at least primary education and 64 percent participate in the labour market. The respective figures for non-Orthodox Christian mothers are 8 and 43 percent.

As mentioned in Section 2.2.2, there is no nationwide survey undertake in Ethiopia to provide evidence on the extent of fasting among pregnant women. The available information both on the ESS and EDHS do not allow us to determine whether the mother was fasting while she was pregnant of the child under consideration. Hence, the effects that we identify in this paper are based on the assumption that fasting during pregnancy is a common practice among Orthodox Christians, for which supportive evidence is provided in Section 2.2.2. The implication on the interpretation of the estimation results is discussed in Section 2.3.2.

2.3.2 Identification Strategy

Previous studies relied on Ramadan fasting among Muslims as a "natural experiment" to identify the short and long-term impacts of fasting during pregnancy (Almond and Mazumder, 2011; van Ewijk, 2011; Oosterbeek and van der Klaauw, 2013; Almond et al., 2014). During the holy month of Ramadan, Muslims around the world abstain from taking any food or drink from sunrise until sunset. Since Ramadan is based on the lunar calendar, its timing changes throughout the Gregorian calendar year. This creates variation in fasting exposure even after controlling for confounding seasonal factors using, for example, month of birth. The mechanism linking prenatal nutrition and health outcomes identified using Ramada fasting is through the timing of prenatal nutritional intake. This is because maternal food intake during Ramadan is restricted from sunrise until sunset, but it is "unlimited" until dawn. Being important on its own, timing of prenatal nutrition, however, might not necessarily result in the reduction of pregnant women's nutrition intake.

Ethiopian Orthodox Christians become vegan for 55 days during the Great Lent. This creates a condition where fasting by pregnant women not only affects the timing of food intake - similar to Ramadan - but also reduces the nutritional content. It does so for a relatively longer period of time (55 days vs 30 days). Hence, the mechanisms that link prenatal environment to health outcomes in our analysis is both altering the timing of prenatal nutrition and a reduction in the type of caloric intake of pregnant women.

Unlike Ramadan, the months that the Great Lent falls change only slightly throughout the Gregorian Calendar year. This makes it impossible to disentangle the effects of in utero fasting exposure from confounding seasonal factors if we only rely the variation in the timing of the Great Lent. Instead, we use the multireligious context of Ethiopia for identification. In Ethiopia, Orthodox Christians, Muslims, Protestants respectively make up 44, 34 and 19 percent of the population according to the 2007 census (CSA, 2008). Given this country context, the effects of prenatal nutritional shortage on children's health outcomes is estimated in a difference-in-differences framework.

The quasi-exogenous variation in exposure to maternal fasting comes from differences among children in their religion and the exact date of birth. The first factor that determines fasting exposure is children's date of birth. A child's treatment status depends whether or not the child's month of gestation coincides with the Great Lent. We determine a child's treatment status using information on exact date of birth, the dates of each Great Lent between 2004 and 2014 and the average length of human gestation, which is about 38 weeks from the day of conception.[9] The share of children who were conceived, in utero and born during the Great Lent are 15.5, 59.4 and 13.6 percent respectively.

The second factor that determines fasting exposure is maternal religion. The health outcome of children whose mother is Orthodox Christians are compared with that of children in the control group whose mother follows a religion other than Orthodox Christianity. Since Muslims also fast during the holy month of Ramadan, our control group could be contaminated. We undertake two specification tests to deal with the observance of Ra-

[9] Even though the average length of human gestation used by medical doctors is 40 weeks, Mongelli et al. (1996) shows that only four percent of women deliver at 40 weeks. The median time from ovulation to birth is about 268 days (38 weeks and 2 days) (Jukic et al., 2013). van Ewijk (2011) also uses a gestation period of 266 days. The dates of Great Lent is obtained from `http://www.oocities.org/hjsmithh/Easter/EasterB.html`.

madan by restricting the control group to Muslim children whose mothers' pregnant do not overlap with Ramadan and by restricting the control group to individuals who follow a religion other than Islam. These religions include protestant, catholic and traditional religions and individuals following these religions comprise 26 percent of our estimation sample.

We use the following equation to get the difference-in-differences estimates:

$$Y_i = \beta_0 + \beta_1 OC_i + \beta_2 Exposed_i + \beta_3 OC_i * Exposed_i + \gamma X_i + \epsilon_i \tag{2.1}$$

where Y_i is a measure of child health outcomes. OC_i is a binary variable that takes the value one if the child's mother is Orthodox Christian, and zero if the mother has a religion other than Orthodox Christianity. $Exposed_i$ is an indicator for whether or not the child's month of gestation coincides with the Great Lent. The exposure variable takes the value (0) if the child's gestation period does not coincide with the Great Lent, (1) if the child was conceived during the months of the Great Lent, (2) if the child was in utero the months of the Great Lent and (3) if the child was born during the months of the Great Lent. X_i is a vector of child-, parental- and household-level control variables including gender, dummies for age in months, month of birth fixed effects, ethnicity, place (urban or rural), region of residence, paternal educational and occupational status and household access to basic sanitation facilities.

We estimate equation (2.1) using median regression to deal with the potential influence of outliers in children's health outcomes. β_3 is the coefficient of interest. It is the difference-in-differences estimate which shows the reduced-form effects of potential in utero exposure to nutritional shortage on children's health outcomes. Since we do not observe in our data whether or not Orthodox Christian pregnant women indeed fast during the Great Lent, we interpret the results as the "intention-to-treat" (ITT) effects. To the extent that not all Orthodox Christian pregnant women observe the Great Lent, the effects we identify in this setting are lower bound estimate of the average treatment effect.

2.4 Estimation Results

2.4.1 Impacts on Children's Nutritional Status

Table 2.3 shows estimation results for the effects of exposure to the Great Lent on children's height-for-age. Children's nutritional status is measured using the continuous variable height-for-age z-score in column (1)-(3) and using a binary indicator for being too short for age (stunt) in column (4)-(6). For each outcome, we report estimates for all children and separate estimates for boys and girls. Baseline regressions are shown in Panel A whereas child, household and geographical controls are included in Panel B. For the sake of clarity, only the difference-in-differences estimates are reported.

Table 2.3: Effects of exposure to the Great Lent on children's height-for-age

	Height-for-age z-score			Child is stunt		
	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Baseline						
Orthodox * Conceived during Lent	-0.050	0.220	-0.200	0.029	-0.084	0.118*
	(0.214)	(0.319)	(0.330)	(0.060)	(0.088)	(0.063)
Orthodox * In utero during Lent	-0.010	0.190	-0.110	0.024	-0.087	0.111**
	(0.172)	(0.261)	(0.261)	(0.041)	(0.068)	(0.049)
Orthodox * Born during Lent	-0.340	-0.010	-0.490*	0.116*	-0.012	0.222***
	(0.211)	(0.303)	(0.295)	(0.068)	(0.085)	(0.078)
Panel B: Including controls						
Orthodox * Conceived during Lent	0.069	0.208	-0.192	0.033	-0.025	0.097*
	(0.125)	(0.162)	(0.209)	(0.059)	(0.085)	(0.057)
Orthodox * In utero during Lent	0.016	0.167	-0.150	0.030	-0.038	0.095**
	(0.083)	(0.139)	(0.110)	(0.039)	(0.066)	(0.044)
Orthodox * Born during Lent	-0.349**	-0.208	-0.397**	0.117*	0.004	0.209***
	(0.143)	(0.157)	(0.178)	(0.067)	(0.086)	(0.068)
No. of observations	10,363	5,270	5,093	10,363	5,270	5,093

Note: Estimation: Median regression in columns (1)-(3) and OLS in columns (4)-(6). Dependent variable: Height-for-age z-score (HAZ) in columns (1)-(3) and a binary indicator for being stunt (HAZ<-2.00) in columns (4)-(6). Controls include child's month of birth, dummies for age in months, gender, rural-urban residence, region of residence, survey year, paternal educational and occupation status, household access to safe drinking water and toilet facility. Standard errors, clustered at month of birth, year of birth and their interaction, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011 and the Ethiopian Socioeconomic Survey (ESS) 2014.

Children whose gestation months overlap with the Great Lent have significantly lower height-for-age z-score compared to children whose gestation period do not overlap with the Great Lent. The negative effect on height-for-age is mainly observed for girls. The results are robust when we add controls for child level (month of birth, dummies for age in months and gender), geographical (rural-urban residence, region of residence) and household background characteristics (paternal education and occupation status and household access to safe drinking water and toilet facility).

The height-for-age of Orthodox Christian girls whose birth coincides with the Great Lent is about 0.490 standard deviation lower than girls who were not exposed to the Great Lent. When we use a binary indicator for having a height-for-age z-score lower than minus two, we find that Orthodox Christian girls who were born during the Great Lent on average have about 21 percentage points more probability of being stunt compared to girls who were not exposed to the Great Lent. Similarly, girls who were in utero during the entire fasting period on average have about 10 to 21 percentage points higher probability of being too short for their age compared to girls who were not exposed to the Great Lent. Relative to the average stunting rate of 43 percent, the effect implies a 26 to 49 percent increase.[10] The point estimates we find are very similar to Akresh et al. (2012) who examined the effects of in utero and early childhood exposure to the 1998-2000 Ethiopian-Eritrean war. The authors find that children born during the Ethiopia-Eritrea war and were living in the war region in Ethiopia experienced 0.45 standard deviations lower height-for-age z-scores.

The different results we found across gender could be because of the following reasons. First, the medical literature show that boys are more fragile to the fetal environment than girls (e.g. Kraemer, 2000). If mortality is high among the severely undernourished boys, those who survive after birth are relatively healthier. Hence, this positive selection at birth could downward bias any effect of in utero nutritional shortage on the height-for-age of the surviving boys. Second, there could be preferential treatment among Ethiopian

[10] The prevalence of stunting in Ethiopia is similar across gender.

parents for boys which can compensate for any negative consequences of nutritional shock that occurred during pregnancy. The existing studies show that Ethiopian parents prefer to have sons (Short and Kiros, 2002) and are more likely to seek health care services (such as immunization) for their sons compared to their daughters (WHO, 2010).

The impacts of nutritional shocks to the fetus environment could vary depending of the gestational periods at the time of exposure to maternal fasting. The empirical evidence in the economics and bio-medical literature finds negative impacts on health outcomes at various gestational periods. Almond and Mazumder (2011)'s results highlight the importance of the first and second trimester as critical period for birth weight and disability outcomes whereas other studies show that nutritional shocks during the late gestation period (Almond and Currie, 2011) and during all stages of the prenatal period (van Ewijk, 2011) influence short-and long-term measures of health outcomes. To test for heterogeneous effects of fasting exposure by gestation period, we divide the "in utero" treatment variable into three trimesters of pregnancy. The results are shown in Table 2.4.

Table 2.4: Impacts of fasting exposure by gestation period

	Height-for-age z-score			Child is stunt		
	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
Conceived during Lent	0.078	0.210	-0.200	0.035	-0.025	0.095
	(0.122)	(0.331)	(0.187)	(0.062)	(0.085)	(0.058)
1st Trimester	0.077	0.174	-0.160	-0.041	-0.091	0.020
	(0.134)	(0.136)	(0.159)	(0.056)	(0.081)	(0.047)
2nd Trimester	0.038	0.165	-0.100	0.050	-0.022	0.114*
	(0.106)	(0.112)	(0.143)	(0.057)	(0.077)	(0.064)
3rd Trimester	-0.022	0.128	-0.322**	0.046	-0.013	0.135*
	(0.134)	(0.163)	(0.140)	(0.055)	(0.076)	(0.076)
Born during Lent	-0.337**	-0.195*	-0.463**	0.104	0.003	0.210***
	(0.145)	(0.113)	(0.186)	(0.069)	(0.086)	(0.068)
No. of observations	10,363	5,270	5,093	10,363	5,270	5,093

Note: Estimation: Median regression in columns (1)-(3) and OLS in columns (4)-(6). Dependent variable: Height-for-age z-score (HAZ) in columns (1)-(3) and a binary indicator for being stunt (HAZ<-2.00) in columns (4)-(6). Regression includes controls for child's month of birth, dummies for age in months, gender, rural-urban residence, region of residence, survey year, paternal educational and occupation status, household access to safe drinking water and toilet facility. Standard errors, clustered at month of birth, year of birth and their interaction, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011 and the Ethiopian Socioeconomic Survey (ESS) 2014.

The health effects of Lenten exposure is larger when it occurs during mid to late gestation period. Although all the coefficients show negative effects on the height-for-age of girls, the impact is only statistically significant when fasting exposure occurs during the third trimester or around birth. The height-for-age z-scores of girls who were in late gestation during the Great Lent is between 0.322 and 0.463 standard deviation lower compared to girls who were not exposed to the Great Lent. This represents declines of 19 and 27 percent compared to the average height-for-age Z-scores. Girls whose second trimester gestation period coincide with the Great Lent are also significantly more likely to be stunt.[11]

2.4.2 Selective Timing of Pregnancy

We undertake various tests to check the robustness of the results. Our first robustness check deals with the possible systematic differences in the timing of pregnancy. Well-educated or health-conscious mothers may be aware of the possible negative effect that fasting during pregnancy may have on their children and hence they may avoid pregnancies that overlap with the Great Lent. Children in the control group could be more healthier than children in the treated group, not because they were not exposed to fasting, but because they have better educated mothers. We minimize the consequences of selective time of pregnancy by controlling for various measures of observable maternal characteristics. One of the advantage of DHS is that it includes vast information to capture the education level, literacy skills, occupational status, health and health knowledge of mothers.

Table 2.5 shows the robustness of the results when we control for various measures of maternal characteristics.

[11] When we use height in centimetre instead of height-for-age z-scores, the results in Appendix Table A2.3 show a decline in height of about 1 to 1.9 centimetres when fasting exposure occurs during the third trimester or around birth. This result is lower than the 3 centimetres reduction in height that Dercon and Porter (2014) find for children who were under the age of 36 months during the peak of the 1984 Ethiopian famine .

Table 2.5: Controlling for maternal characteristics

	Height-for-age z-score			Child is stunt		
	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Maternal education, literacy and occupational status						
Orthodox * Conceived during Lent	0.020	0.229	-0.225	0.033	-0.018	0.095
	(0.124)	(0.155)	(0.206)	(0.059)	(0.086)	(0.058)
Orthodox * In utero during Lent	-0.003	0.177	-0.181	0.030	-0.036	0.097**
	(0.099)	(0.138)	(0.121)	(0.037)	(0.066)	(0.045)
Orthodox * Born during Lent	-0.367**	-0.171	-0.517***	0.116*	0.003	0.209***
	(0.151)	(0.186)	(0.192)	(0.066)	(0.086)	(0.068)
Panel B: Including maternal health status						
Orthodox * Conceived during Lent	0.062	-0.132	-0.062	0.038	0.010	0.087
	(0.168)	(0.196)	(0.155)	(0.061)	(0.084)	(0.060)
Orthodox * In utero during Lent	0.004	-0.033	-0.153	0.024	-0.035	0.089*
	(0.133)	(0.174)	(0.135)	(0.042)	(0.066)	(0.050)
Orthodox * Born during Lent	-0.263*	-0.383**	-0.481***	0.103	-0.010	0.191***
	(0.150)	(0.178)	(0.149)	(0.068)	(0.086)	(0.068)
Panel C: Including maternal health knowledge						
Orthodox * Conceived during Lent	0.065	-0.139	-0.068	0.043	0.007	0.098
	(0.157)	(0.199)	(0.151)	(0.061)	(0.084)	(0.060)
Orthodox * In utero during Lent	0.003	-0.078	-0.106	0.028	-0.037	0.098**
	(0.128)	(0.173)	(0.133)	(0.041)	(0.067)	(0.048)
Orthodox * Born during Lent	-0.288**	-0.392**	-0.451***	0.109	-0.011	0.205***
	(0.145)	(0.190)	(0.174)	(0.067)	(0.086)	(0.068)
No. of observations	7,621	3,883	3,738	7,621	3,883	3,738

Note: Estimation: Median regression in columns (1)-(3) and OLS in columns (4)-(6). Dependent variable: Height-for-age z-score (HAZ) in columns (1)-(3) and a binary indicator for being stunt (HAZ<-2.00) in columns (4)-(6). All regressions include controls for child's month of birth, dummies for age in months, gender, rural-urban residence, region of residence, survey year, paternal educational and occupation status, household access to safe drinking water and toilet facility. Standard errors, clustered at month of birth, year of birth and their interaction, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011.

In panel A, we control for maternal education, literacy skills and occupational status. Maternal education is measured using a binary variable indicator whether or not the mother has completed at least primary education. Maternal literacy skills is a binary indicator for being able to read and write. Maternal occupational status takes the value zero is the mother is not working, one if the mother works in the agriculture sector, two if the mother works in the non-agricultural sector. In panel B, we include controls for maternal health status. We use an objective measures of health status using the mother's height (in centimeter) and weight (in kilogram). Finally, panel C controls for maternal health knowledge where health knowledge is measured by a binary variable indicating whether or not the mother knows the ovulation cycle. Since the detailed measures of maternal characteristics are not available in the ESS, this part of our analysis is based on the EDHS data. The end results show that controlling for maternal characteristics barely changes the difference-in-differences estimates.

To provide further evidence that selective timing of pregnancy is unlikely to confound the difference-in-differences estimates, Table 2.6 tests for any significant correlation between our explanatory variables of interest and maternal characteristics. Specifically, we use maternal education, literacy skills and health knowledge as outcome variables in equation (2.1). If mothers with better education or health knowledge are more likely to alter the timing their pregnancy, we should observe a significant negative correlation between a child's exposure to the Great Lent and maternal characteristics. The results show that whether or not a child is conceived, in utero or born during the Great Lent is not related to the mother's observable characteristics.

Table 2.6: Selection to Lent exposure with respect to maternal characteristics.

	Educational level		Literacy skills		Health knowledge	
	(1)	(2)	(3)	(4)	(5)	(6)
Orthodox * Conceived during Lent	-0.021	-0.043	0.043	0.015	0.000	0.022
	(0.039)	(0.035)	(0.057)	(0.046)	(0.030)	(0.032)
Orthodox * In utero during Lent	-0.016	-0.035	0.002	-0.020	0.028	0.037
	(0.037)	(0.034)	(0.044)	(0.039)	(0.032)	(0.033)
Orthodox * Born during Lent	-0.036	-0.038	0.020	0.017	-0.000	0.011
	(0.046)	(0.044)	(0.069)	(0.067)	(0.053)	(0.057)
Controls[1]		√		√		√
No. of observations	10,363	10,363	7,968	7,968	7,958	7,958

Note: Estimation: OLS. The dependent variables are displayed on top of each column. Controls include month of birth, dummies for age in months, gender, rural-urban residence, region of residence, paternal educational and occupation status, household access to safe drinking water and toilet facility. Standard errors, clustered at month of birth, year of birth and their interaction, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011 and the Ethiopian Socioeconomic Survey (ESS) 2014.

2.4.3 Consquences of Ramadan Exposure

As discussed in Section 2.3.2, we identify the effects of in utero exposure to the Great Lent by using Muslims as part of the control group. Since Muslims also fast during the holy month of Ramadan, our control group could be contaminated. We undertake two specification tests to deal with the observance of Ramadan. First, we restrict the control group to Muslim children whose mothers' pregnant do not overlap with Ramadan. If children whose gestation period coincide with Ramadan have negative health outcome, dropping Muslims who were exposed to Ramadan could still contaminate the control group. Our control group may consist of the relatively healthy Muslims who were not exposed to Ramadan. The multi-religious of Ethiopia allows use to deal with this by restricting the control group to individuals who follow a religion other than Islam. These religions include protestant, catholic and traditional religions and individuals following these religions comprise 26 percent of our estimation sample.

Estimation results are reported in Table 2.7. In column (1)-(3), Muslims who were exposed to Ramadan are dropped from the control group whereas all Muslims are dropped from the control group in column (4)-(6). The negative effect of exposure to the Great Lent becomes significant for boys born during the fasting period. The height-for-age z-score of boys born during the Great Lent is about 0.49 standard deviation lower than boys who were not exposed to the Great Lent. The negative effect of girls exposure to fasting on the likelihood of stunting becomes insignificant due to the relatively small size. However, the magnitude of the effects are similar to the results shown in Table 2.3. As we show on Appendix Table A2.4, we do not find any statistically significant negative effect of Muslims's in utero exposure to the observance of Ramadan on children's height-for-age. This provides further evidence that the observance of Ramadan among Muslims does not significantly alter the effects of in utero exposure to the Great Lent.

Table 2.7: Possible selection bias due to Ramadan exposure

	Drop Muslims exposed to Ramadan			Drop all Muslims		
	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Height-for-age z-score						
Orthodox * Conceived during Lent	-0.001	-0.181	-0.172	-0.082	-0.157	-0.372
	(0.188)	(0.207)	(0.241)	(0.188)	(0.206)	(0.252)
Orthodox * In utero during Lent	0.080	-0.092	-0.164	0.053	0.049	-0.356*
	(0.149)	(0.186)	(0.190)	(0.158)	(0.178)	(0.198)
Orthodox * Born during Lent	-0.381**	-0.491**	-0.532**	-0.395**	-0.512***	-0.660***
	(0.173)	(0.200)	(0.233)	(0.165)	(0.194)	(0.237)
Panel B: Child is stunt						
Orthodox * Conceived during Lent	0.076	0.055	0.124	0.077	0.059	0.124
	(0.082)	(0.114)	(0.101)	(0.081)	(0.114)	(0.102)
Orthodox * In utero during Lent	0.049	0.033	0.097	0.049	0.021	0.111
	(0.052)	(0.087)	(0.075)	(0.054)	(0.091)	(0.076)
Orthodox * Born during Lent	0.178**	0.131	0.236**	0.180**	0.134	0.238**
	(0.084)	(0.108)	(0.109)	(0.084)	(0.109	(0.109)
No. of observations	4,716	3,298	2,348	4,210	2,099	2,111

Note: Estimation: Median regression in Panel A and C and OLS in Panel B and D. All regression include controls for month of birth, dummies for age in months, gender, rural-urban residence, region of residence, paternal educational and occupation status, household access to safe drinking water and toilet facility and maternal characteristics including education, literacy skills, occupation, health status and health knowledge. Standard errors, clustered at month of birth, year of birth and their interaction, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011 and the Ethiopian Socioeconomic Survey (ESS) 2014.

2.4.4 Impacts on Sex-Ratio and Birth Weight

Our empirical analysis so far shows that exposure to the Great Lent during pregnancy negatively affect the nutritional status of children. Following, we shed some light on the consequences of maternal fasting on the sex ratio and birth weight. As we discussed above, the medical literature shows that boys are more fragile to the fetal environment than girls. Nutritional shock during pregnancy could lead to higher mortality among boys, and hence reducing the number of male live births. Furthermore, the negative consequences of maternal fasting during pregnancy could already be observed when comparing birth weights of exposed and non-exposed pregnancies. However, as the fetal origin hypothesis posits, the negative health effects of nutritional shock during pregnancy might not necessarily be observed at birth.

We measure birth weight based on the mother's subjective assessment of the birth weight of her children born in the past five years compared to other children of the same age. The information is only available in the EDHS. The variable takes five values: the child's birth weight is very large, larger than average, average, smaller than average, or very small. For ease of interpretation, we define a binary variable which takes the value one if the child's birth weight is smaller than average or very small and the variable takes zero if the child's birth weight is average, larger than average or very large. To examine the effect on the sex-ratio, we use an outcome variable indicating that the child is a boy. Table 2.8 shows estimation results.

Table 2.8: Effects of Lent exposure on sex-ratio and birth weight

	Sex-ratio	Below average birth weight		
	Child is male	All	Boys	Girls
	(1)	(2)	(3)	(4)
Panel A: Baseline				
Orthodox * Conceived during Lent	0.013	0.005	0.044	-0.030
	(0.087)	(0.059)	(0.067)	(0.067)
Orthodox * In utero during Lent	0.010	-0.043	-0.034	-0.049
	(0.066)	(0.051)	(0.059)	(0.058)
Orthodox * Born during Lent	0.048	0.031	0.082	-0.012
	(0.087)	(0.059)	(0.076)	(0.076)
No. of observations	7,968	7,951	4,054	3,897
Panel B: Including controls				
Orthodox * Conceived during Lent	-0.031	0.018	0.097	-0.048
	(0.086)	(0.065)	(0.063)	(0.082)
Orthodox * In utero during Lent	-0.018	-0.033	-0.013	-0.059
	(0.065)	(0.054)	(0.044)	(0.074)
Orthodox * Born during Lent	0.029	0.014	0.036	-0.024
	(0.086)	(0.060)	(0.058)	(0.087)
No. of observations	7,621	7,605	3,877	3,728

Note: Estimation: OLS. Dependent variable: A binary indicator for being male in columns (1) and a binary indicator for smaller than average birth weight assessed by the mother in columns (2)-(4). In panel B, we control for month of birth, dummies for age in months, gender, rural-urban residence, region of residence, survey year, parental educational and occupation status, household access to safe drinking water, toilet facility and maternal literacy skills, health status and health knowledge. Standard errors, clustered at month of birth, year of birth and their interaction, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011.

Column (1) shows results for male-to-female sex-ratio whereas the effect of fasting exposure on birth weight is shown in column (2) for all children, separately for boys in column (3) and girls in column (4).

We find no statistically significant differences in the sex-ratio and birth weight of children whose gestation periods overlap with the Great Lent and children whose gestation periods do not overlap with the Great Lent. These results should be interpreted with caution for the following reasons. First, our estimation sample consists of children who are alive at the time of the survey since the surveys only collect information on exact date of birth for children who are alive at the time of the survey. The results could differ if we use birth records which include all live births. Second, we measure birth weight using mothers' subjective assessment of the birth weight of their children compared to other children of the same age. Despite these data shortcomings, our result adds to the debate on the relationship between nutritional shortage during pregnancy, the sex ratio and birth weight.

There is less consensus reached in the medical and economics literature on the relationship between nutritional shortage during pregnancy and the sex ratio (e.g. Stein et al., 2004; Cramer and Lumey, 2010; Neelsen and Stratmann, 2011; van den Berg et al., 2016). Based on one million births to Muslim mothers in Germany, Juerges (2015) find zero effect of Ramadan observance during pregnancy on the fraction of male births and birth weight. The significant negative effect that Almond and Mazumder (2011) finds on birth weight is very small in magnitude compared to the effect that in utero fasting exposure have on other economic outcomes. Hence, the effects of in utero exposure to nutritional shortage on the sex-ratio and outcomes observed at birth seem to be context specific.

2.5 Conclusion

According to the David Barker's fetal origin hypothesis, nutritional deficiency during the gestation period have negative consequences on health outcome because the structure of different body organs including the brain, lung, liver, etc. undergo programming during this critical period. In Ethiopia, Orthodox Christians children whose gestation period coincides with the observance of the Great Lent have significantly lower nutritional status compared to children whose gestation period do not overlap with the Great Lent. This is attributed to in utero nutritional shortage due to maternal fasting during the Great Lent.

Using data from the 2011 Ethiopian Demographic and Health Survey and the 2014 Ethiopian Socioeconomic Survey, we examine the differences in the consumption of non-fasting food item across religion depending on whether or not the survey interviews overlap with the Great Lent. Compared to households that follow a religion other than Orthodox Christianity, we find that Orthodox Christian households have about 10 to 27 percentage points less likelihood of consuming non-fasting foods such as milk, egg, cheese and meat in survey interviews that overlap with the Great Lent compared to interviews that do not overlap with the Great Lent. More importantly, we find that Orthodox Christian mothers have about 10 to 14 percentage points less probability of feeding their children under the age of two non-fasting food items in survey interviews that overlap with the Great Lent compared to interviews that do not overlap with the Great Lent. Since children under the age of seven are exempted from fasting, the type of food that women feed their children under the age of two indirectly shows the type of food available in the household for consumption during the Great Lent.

Our analysis shows that Orthodox Christian children whose gestation period overlaps with the Great Lent have significantly lower nutritional status. Children who were in utero during the entire fasting period on average have about 10 to 21 percentage points higher probability of being too short for their age (stunt) compared to children who were not exposed to the Great Lent. The height-for-age z-scores of children who were in late gestation during the Great Lent is between 0.322 and 0.463 standard deviation lower than children who were not exposed to the Great Lent. The negative consequences of fasting exposure on height-for-age is observed mainly for girls and only for boys who were born during the Great Lent. The results are robust to controlling for seasonality in health outcomes and various measures of family background characteristics including parental education, occupation, maternal health status and knowledge and household access to safe drinking water and sanitation facilities. Potential selective timing of pregnancy in response to the observance of Great Lent also does not confound the results.

The results are useful in light of the Ethiopian government's and international organizations aim to improve the health status of mothers and children. The country's national health policy program has given a significant focus to improving nutritional status during early childhood (MoH, 2010, 2015). One of the health intervention to improve the nutritional status of children is to raise the awareness and change the attitude of the public (including religious leaders) in general and pregnant women in particular regarding the fasting rules of the church and the negative consequences it could have on their new born. The Ethiopian Orthodox church does allow pregnant women to restrain from fasting. The available evidence (e.g. Gebremedhin et al., 2015) show that pregnant women fast during the Great Lent not only due to lack of awareness regarding the rules of the church but also due to fear of rejection by the community including religious leaders. Health professional in public and private health care facilities could play an important role by paying special attention to the nutritional intake of women during fasting periods. During antenatal care or home visits, health care workers could monitor the nutritional status of pregnant women and refer fasting pregnant women to consult informed religious leaders when deemed necessary.

The possible negative consequences of fasting observance on the nutritional status of mothers and children is also being recognized by international donors. Under their

ENGINE[12] program, USAID and Save the Children recently held a discussion with religious leaders of the Ethiopian Orthodox Church on the role of the church to improve nutrition by addressing fasting practices of pregnant and lactating women and children under the age of two. Soon after the workshop, the organizations led out a guideline endorsed by the Patriarch of the church to raise the awareness not only of the public but also that of religious leaders regarding the fasting rules of the church.[13] The impact of the guidelines on the fasting practices of Orthodox Christians remains to be seen.

As to the author's knowledge, there is no data available on the fasting practice of pregnant women in all of the large scale household surveys. This pauses a critical challenge to examine the extent of fasting during pregnancy and to trace its change over time. Since the timing of existing surveys such as the EDHS and ESS coincide with the observance of the Great Lent, questions on the pregnancy and fasting status of women could easily be incorporated. Furthermore, due to lack of data on exact of birth for adults, the current paper exclusively focused on the health status of children under the age of five. The consequences of in utero fasting exposure on adult outcomes including health, educational attainment and labour market outcomes could be examined when data on the exact date of birth of adults become available.

[12] ENGINE stands for Empowering New Generations to Improve Nutrition and Economic opportunities

[13] A press release titled "Patriarch of the Ethiopian Orthodox Tewahedo Church Endorses Nutrition Sermon Guide Encouraging Better Nutrition for Mothers and Children" was made on January 12, 2016 and it can be accessed at `https://www.usaid.gov`.

2.6 References

Akresh, R., Lucchetti, L., and Thirumurthy, H. (2012). Wars and child health: Evidence from the eritreanethiopian conflict. *Journal of development economics*, 99(2):330–340.

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

Almond, D. and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *Journal of Economic Perspectives*, 25(3):153–72.

Almond, D., Edlund, L., Li, H., and Zhang, J. (2007). Long-term effects of the 1959-1961 china famine: Mainland china and hong kong. *National Bureau of Economic Research, Working Paper No. w13384*.

Almond, D. and Mazumder, B. (2011). Health capital and the prenatal environment: The effect of ramadan observance during pregnancy. *American Economic Journal: Applied Economics*, 3(4).

Almond, D., Mazumder, B., and van Ewijk, R. (2014). In utero ramadan exposure and children's academic performance. *Economic Journal*.

Barker, D. J. and Osmond, C. (1986). Infant mortality, childhood nutrition, and ischaemic heart disease in england and wales. *The Lancet*, 327(8489):1077–1081.

Barreca (2010). The long-term economic impact of in utero and postnatal exposure to malaria. *Journal of Human Resources*, 45(4):865–892.

Betru, S. and Kawashima, H. (2009). Patterns and determinants of meat consumption in urban and rural ethiopia. *Livestock Research for Rural Development*, 21(9):143.

Bhate, V., Deshpande, S., Bhat, D., Joshi, N., Ladkat, R., Watve, S., Fall, C., de Jager, C. A., Refsum, H., and Yajnik, C. (2008). Vitamin b12 status of pregnant indian women and cognitive function in their 9-year-old children. *Food and nutrition bulletin*, 29(4):249–254.

Black, M. M. (2008). Effects of vitamin b12 and folate deficiency on brain development in children. *Food and nutrition bulletin*, 29.2(2 suppl1):S126–S131.

Boylston, T. (2013). *Food, life and material religion in Ethiopian Orthodox Christianity.* John Wiley and Sons.

Chen, Y. and Zhou, L.-A. (2007). The long-term health and economic consequences of the 19591961 famine in china. *Journal of health economics*, 26(4):659–681.

Cramer, J. S. and Lumey, L. H. (2010). Maternal preconception diet and the sex ratio. *Human biology*, 82(1):103–107.

CSA (1994). Summary and statistical report of the 2007 population and housing census. Technical report, Federal Democratic Republic of Ethiopia Population Census Comisssion.

CSA (2008). Summary and statistical report of the 2007 population and housing census. Technical report, Federal Democratic Republic of Ethiopia Population Census Comisssion.

Currie, J. and Vogl, T. (2012). Early-life health and adult circumstance in developing countries. *NBER Working Paper No. 18371.*

de Rooija, S. R., Woutersa, H., Yonkerb, J. E., Painterc, R. C., and Roseblooma, T. J. (2010). Prenatal undernutrition and cognitive function in late adulthood. *In Proceedings of the National Academy of Sciences*, 107.39(39):16881–16886.

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

DHS (2012). Ethiopian demographic and health survey 2011. Technical report, Central Statistical Agency and ICF International, Addis Ababa, Ethiopia and Calverton, Maryland, USA.

Dinka, H., Chala, R., Dawo, F., Bekana, E., and Leta, S. (2010). Major constraints and health management of village poultry production in rift valley of oromia, ethiopia. *American-Eurasian Journal of Agriculture and Environmental Science*, 9(5):529–533.

Doblhammer, G., van den Berg, G. J., and Lumey, L. H. (2013). A re-analysis of the long-term effects on life expectancy of the great finnish famine of 1866-68. *Population Studies*, 67:309–322.

Dror, D. K. and Allen, L. H. (2008). Effect of vitamin B12 deficiency on neurodevelopment in infants: current knowledge and possible mechanisms. *Nutrition Reviews*, 66(5):250–255.

Gebremedhin, M., Ambaw, F., Admassu, E., and Berhane, H. (2015). Maternal associated factors of low birth weight: a hospital based crosssectional mixed study in tigray, northern ethiopia. *BMC Pregnancy Childbirth*, 15(222).

Innis, S. M. (2007). Human milk: maternal dietary lipids and infant development. *Proceedings of the Nutrition Society*, 66:397–404.

Juerges, H. (2015). Ramadan fasting, sex-ratio at birth, and birth weight: No effects on muslim infants born in germany. *Economics Letters*, 137:13–16.

Jukic, A., Baird, D., Weinberg, C., McConnaughey, D., and Wilcox, A. (2013). Length of human pregnancy and contributors to its natural variation. *Human Reproduction*, 28(10):2848–2855.

Knutsson, K. E. and Selinus, R. (1970). Fasting in ethiopia: An anthropological and nutritional study. *The American Journal of Clinical Nutrition*, 23(7):956–969.

Kraemer, S. (2000). The fragile male. *BMJ: British Medical Journal*, 321(7276):1609.

Langley-Evans, S. C. (2006). Developmental programming of health and disease. *Proceedings of the Nutrition Society*, 65(1):97–105.

Lau, C., Rogers, J. M., Desai, M., and Ross, M. G. (2011). Fetal programming of adult disease. *Obstet Gynecol*, 117:978–85.

Lumey, L., A.D., S., and Susser, E. (2011). Prenatal famine and adult health. *Annual Review of Public Health*, 32(237-62).

Metzger, B. E., Vileisis, R. A., Ravnikar, V., and Freinkel, N. (1982). "accelerated starvation" and the skipped breakfast in late normal pregnancy. *The Lancet*, 319(8272):588–592.

MoH (2010). Health Sector Development Program (HSDP) IV 2010/11-2014/15. Technical report, Federal Ministry of Health.

MoH (2015). Health Sector Transformation Plan (HSTP) 2015/16 - 2019/20. Technical report, Federal Ministry of Health.

Mongelli, M., Wilcox, M., and Gardosi, J. (1996). Estimating the date of confinement: ultrasonographic biometry versus certain menstrual dates. *American journal of obstetrics and gynecology*, 174(1):278–281.

Mu, R. and Zhang, X. (2011). Why does the great chinese famine affect the male and female survivors differently? mortality selection versus son preference. *Economics and Human Biology*, 9(1):92–105.

Neelsen, S. and Stratmann, T. (2011). Effects of prenatal and early life malnutrition: Evidence from the greek famine. *Journal of Health Economics*, 30(3):479–488.

Nelson, R. E. (2010). Testing the fetal origins hypothesis in a developing country: evidence from the 1918 influenza pandemic. *Health economics*, 19(10):1181–1192.

Oosterbeek, H. and van der Klaauw, B. (2013). Ramadan, fasting and educational outcomes. *Economics of Education Review*, 34:219–226.

Roseboom, T. J., Meulen, J. H. V. D., Ravelli, A. C., Osmond, C., Barker, D. J., and Bleker, O. P. (2001). Effects of prenatal exposure to the dutch famine on adult disease in later life: an overview. *Molecular and cellular endocrinology*, 185(1):93–98.

Scholte, R. S., van den Berg, G. J., and Lindeboom, M. (2015). Long-run effects of gestation during the dutch hunger winter famine on labor market and hospitalization outcomes. *Journal of Health Economics*, 39:1730.

Schultz-Nielsen, M. L., Tekin, E., and Greve, J. (2016). Labor market effects of intrauterine exposure to nutritional deficiency: Evidence from administrative data on muslim immigrants in denmark. *Economics & Human Biology*, 21:196 – 209.

Selassie, S. H. and Tamerat, T. (1970). *The Church of Ethiopia: a Panorama of History and Spiritual Life*. Ethiopian Orthodox Church, Addis Ababa.

Seleshe, S., Jo, C., and Lee, M. (2014). Meat consumption culture in ethiopia. *Korean Journal of Food Science from Animal Resources*, 34(1):7–13.

Short, S. E. and Kiros, G.-E. (2002). Husbands, wives, sons, and daughters: Fertility preferences and the demand for contraception in ethiopia. *Population Research and Policy Review*, 21(5):377–402.

Sparén, P., Vågerö, D., Shestov, D. B., Plavinskaja, S., Parfenova, N., Hoptiar, V., Paturot, D., and Galanti, M. R. (2003). Long term mortality after severe starvation during the siege of leningrad: prospective cohort study. *British Medical Journal*, 328:11–15.

Stanner, S. A., Bulmer, K., Andres, C., Lantseva, O. E., Borodina, V., Poteen, V. V., and Yudkin., J. S. (1997). Does malnutrition in utero determine diabetes and coronary heart disease in adulthood? results from the leningrad siege study, a cross sectional study. *BMJ: British Medical Journa*, 315(7119):1342–1348.

Stein, A. D., Barnett, P. G., and Sellen, D. W. (2004). Maternal undernutrition and the sex ratio at birth in ethiopia: evidence from a national sample. In *Proceedings of the Royal Society of London B: Biological Sciences*, volume 271 of *Suppl 3*, pages S37 – S39.

Teklewold, H., Legese, G., Alemu, D., and Negasa, A. (2009). Determinants of livestock prices in ethiopian pastoral livestock markets: Implications for pastoral marketing strategies. In *Contributed Paper prepared for presentation at the International Association of Agricultural Economists Conference Beijing, China, August 16-22*.

van den Berg, G. J., Pinger, P. R., and Schoch, J. (2016). Instrumental variable estimation of the causal effect of hunger early in life on health later in life. *The Economic Journal*, 126(591):465–506.

van Ewijk, R. (2011). Long-term health effects on the next generation of ramadan fasting during pregnancy. *Journal of Health Economics*, 30:1246 1260.

von Schenck, U., Bender-Götze, C., and Koletzko, B. (1997). Persistence of neurological damage induced by dietary vitamin B-12 deficiency in infancy. *Archives of Disease in Childhood*, 77(137-139).

WHO (2010). *Gender, Women and Primary Health Care Renewal: A Discussion Paper*. World Health Organization.

Zepro, N. B. (2015). Food taboos and misconceptions among pregnant women of shashemene district, ethiopia, 2012. *Science Journal of Public Health*, 3(3):410–416.

2.7 Appendix

Table A2.1: Effects of Lent exposure on children's height-for-age z-score

	All		Boys		Girls	
	(1)	(2)	(3)	(4)	(5)	(6)
1 if Orthodox Christian	0.107	0.033	0.026	0.198	0.234**	0.188
	(0.080)	(0.122)	(0.132)	(0.172)	(0.102)	(0.129)
1 if Conceived during Lent	-0.352	-0.291*	-0.473***	-0.154	-0.159	-0.302*
	(0.122)	(0.166)	(0.153)	(0.201)	(0.212)	(0.169)
1 if In utero during Lent	-0.213	-0.166	-0.405**	-0.118	-0.021	-0.197
	(0.168)	(0.163)	(0.197)	(0.209)	(0.261)	(0.200)
1 if Born during Lent	-0.054	-0.161	-0.184	-0.029	0.079	0.010
	(0.169)	(0.160)	(0.178)	(0.202)	(0.299)	(0.174)
Orthodox * Conceived in Lent	0.069	0.065	0.208	-0.139	-0.192	-0.068
	(0.125)	(0.157)	(0.172)	(0.199)	(0.209)	(0.151)
Orthodox * In utero during Lent	0.016	0.128	0.167	-0.078	-0.150	-0.106
	(0.083)	(0.145)	(0.139)	(0.173)	(0.110)	(0.132)
Orthodox * Born during Lent	-0.349**	-0.288**	-0.208	-0.392**	-0.397**	-0.451***
	(0.143)	(0.145)	(0.158)	(0.190)	(0.178)	(0.174)
1 if Girl	0.014	0.025				
	(0.035)	(0.037)				
1 if HH lives in rural area	-0.353***	-0.350***	-0.419***	-0.297***	-0.340***	-0.215**
	(0.083)	(0.073)	(0.091)	(0.100)	(0.085)	(0.094)
1 if Father with primary ed.	0.109**	0.024	0.109	-0.007	0.143*	0.093
	(0.512)	(0.070)	(0.104)	(0.076)	(0.075)	(0.076)
1 if Father with secondary ed.	0.512***	0.273***	0.455	0.286**	0.595***	0.208***
	(0.072)	(0.065)	(0.110)	(0.118)	(0.089)	(0.094)
1 if Father works in non-agri.	-0.009	0.026	-0.010	0.097	-0.123**	-0.097
	(0.052)	(0.055)	(0.076)	(0.063)	(0.063)	(0.094)
1 if HH has access to safe water	0.029	0.039	0.118***	0.129**	0.023	0.021
	(0.038)	(0.046)	(0.045)	(0.054)	(0.052)	(0.047)
1 if HH has access to toilet	0.050	0.021	0.248	0.050	-0.039	-0.024
	(0.037)	(0.041)	(0.046)	(0.051)	(0.051)	(0.047)
1 if Mother with primary edu.		0.171**		0.098		0.294***
		(0.072)		(0.117)		(0.093)
1 if Mother is literate		-0.049		-0.138		-0.057
		(0.059)		(0.096)		(0.068)
1 if Mother works in agri.		0.087*		0.165***		-0.048
		(0.047)		(0.056)		(0.055)
1 if Mother works in non-agri.		-0.057		-0.115*		-0.028
		(0.049)		(0.065)		(0.056)
Mother height (centimeter)		0.040***		0.029***		0.042***
		(0.004)		(0.005)		(0.005)
Mother weight (kilogram)		0.016***		0.024***		0.015***
		(0.003)		(0.004)		(0.004)
1 if Mother knows her ovu. cycle		0.021		0.076		-0.063
		(0.034)		(0.245)		(0.059)
No. of observations	10,363	7,621	5,270	3,883	5,100	3,738

Note: Estimation: Median regression. Dependent variable: Height-for-age z-score (HAZ). All regressions include controls for month of birth, dummies for age in months, region of residence and survey year. Bootstrap standard errors are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011 and the Ethiopian Socioeconomic Survey (ESS) 2014.

Table A2.2: Effects of Lent exposure on children's probabiltiy of child stunting

	OLS			Probit		
	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
1 if Orthodox Christian	-0.055*	-0.013	-0.101***	-0.152*	-0.044	-0.293***
	(0.032)	(0.063)	(0.032)	(0.089)	(0.174)	(0.094)
1 if Conceived during Lent	0.101	0.063	0.110	0.271*	0.164	0.308
	(0.061)	(0.061)	(0.086)	(0.163)	(0.166)	(0.232)
1 if In utero during Lent	0.043	0.019	0.050	0.116	0.058	0.134
	(0.054)	(0.062)	(0.083)	(0.145)	(0.173)	(0.227)
1 if Born during Lent	0.052	0.042	0.031	0.141	0.119	0.090
	(0.045)	(0.060)	(0.063)	(0.123)	(0.165)	(0.169)
Orthodox * Conceived in Lent	0.043	0.007	0.098	0.122	0.021	0.299*
	(0.061)	(0.084)	(0.060)	(0.165)	(0.229)	(0.165)
Orthodox * In utero during Lent	0.028	-0.037	0.098**	0.078	-0.115	0.300**
	(0.041)	(0.067)	(0.048)	(0.115)	(0.186)	(0.139)
Orthodox * Born during Lent	0.109	-0.011	0.205***	0.299	-0.039	0.598***
	(0.067)	(0.086)	(0.068)	(0.181)	(0.236)	(0.188)
1 if Girl	-0.011			-0.033		
	(0.020)			(0.055)		
1 if HH lives in rural area	0.111***	0.071	0.149***	0.318***	0.206	0.445***
	(0.034)	(0.047)	(0.046)	(0.099)	(0.141)	(0.137)
1 if Father with primary edu.	-0.009	-0.001	-0.019	-0.024	-0.005	-0.043
	(0.023)	(0.030)	(0.031)	(0.061)	(0.083)	(0.091)
1 if Father with secondary edu.	-0.101***	-0.073	-0.130***	-0.334***	-0.251	-0.439***
	(0.034)	(0.052)	(0.043)	(0.111)	(0.169)	(0.146)
1 if Father works in non-agri.	0.022	-0.009	0.054	0.065	-0.021	0.155
	(0.026)	(0.035)	(0.036)	(0.074)	(0.098)	(0.102)
1 if HH has access to safe water	-0.005	-0.033	0.009	-0.013	-0.088	0.016
	(0.016)	(0.024)	(0.022)	(0.045)	(0.068)	(0.064)
1 if HH has access to toilet	-0.014	-0.033	-0.002	-0.042	-0.093	-0.011
	(0.018)	(0.025)	(0.025)	(0.050)	(0.068)	(0.072)
1 if Mother with primary educ.	-0.043	-0.054	-0.030	-0.141	-0.185	-0.098
	(0.038)	(0.057)	(0.047)	(0.115)	(0.170)	(0.145)
1 if Mother is literate	-0.002	0.001	-0.002	0.001	0.009	-0.007
	(0.022)	(0.037)	(0.033)	(0.063)	(0.103)	(0.099)
1 if Mother works in agri.	-0.029	-0.057*	0.003	-0.081*	-0.152*	0.005
	(0.018)	(0.031)	(0.024)	(0.048)	(0.241)	(0.068)
1 if Mother works in non-agri.	0.012	0.029	0.003	0.033	0.082	0.014
	(0.019)	(0.029)	(0.026)	(0.034)	(0.080)	(0.075)
Mother height (centimeter)	-0.011***	-0.011***	-0.011***	-0.030***	-0.031***	-0.032***
	(0.001)	(0.002)	(0.002)	(0.004)	(0.006)	(0.006)
Mother weight (kilogram)	-0.004***	-0.006***	-0.003	-0.013***	-0.018***	-0.009*
	(0.001)	(0.002)	(0.002)	(0.004)	(0.005)	(0.005)
1 if Mother knows her ovul. cycle	-0.024	-0.044*	-0.009	-0.068	-0.131*	-0.023
	(0.017)	(0.024)	(0.023)	(0.046)	(0.068)	(0.066)
No. of observations	7,621	3,883	3,738	7,621	3,883	3,738

Note: Estimation: OLS regression in column (1)-(3) and Probit model in column (4)-(6). Dependent variable: A binary indicator for being stunt (HAZ<-2.00). All regressions include controls for month of birth, dummies for age in months, region of residence and survey year. Standard errors, clustered at month of birth, year of birth and their interaction, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011 and the Ethiopian Socioeconomic Survey (ESS) 2014.

Table A2.3: Impacts of fasting exposure on height (in cm) by gestation period

	All		Boys		Girls	
	(1)	(2)	(3)	(4)	(5)	(6)
Conceived during Lent	-2.400	0.011	-0.100	0.939	-4.200**	-0.694
	(1.678)	(0.519)	(2.737)	(0.675)	(1.883)	(0.755)
1st Trimester	-1.900	0.454	-0.500	1.139*	-2.600	-0.497
	(1.596)	(0.552)	(2.577)	(0.601)	(2.016)	(0.545)
2nd Trimester	-1.100	0.089	0.600	0.626	-3.100	-0.426
	(1.707)	(0.441)	(2.279)	(0.565)	(2.288)	(0.560)
3rd Trimester	-0.100	-0.045	-0.600	0.354	-0.200	-1.051*
	(1.297)	(0.492)	(1.796)	(0.799)	(1.922)	(0.600)
Born during Lent	-2.700	-1.375**	-2.600	-0.292	-2.500	-1.910***
	(1.651)	(0.646)	(2.177)	(0.643)	(2.383)	(0.718)
Controls[1]		√		√		√
Median [sd] of dependent variable	85.30 [11.19]		86.10 [10.99]		84.70 [11.34]	
No. of observations	10,363		5,270		5,093	

Note: Estimation: Median regression. Dependent variable: Height in centimetre. Regression includes controls for child's month of birth, dummies for age in months, gender, rural-urban residence, region of residence, survey year, paternal educational and occupation status, household access to safe drinking water and toilet facility. Standard errors, clustered at month of birth, year of birth and their interaction, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011 and the Ethiopian Socioeconomic Survey (ESS) 2014.

Table A2.4: Effect of Ramadan exposure on children's height-for-age

	Drop Orthodox Christians exposed to Lent			Drop all Orthodox Christians		
	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Height-for-age z-score						
Muslim * Conceived during Ramadan	-0.039	-0.366	0.436	-0.008	-0.320	0.462
	(0.303)	(0.386)	(0.415)	(0.232)	(0.370)	(0.437)
Muslim * In utero during Ramadan	0.043	-0.327*	0.492	0.013	-0.375*	0.486
	(0.172)	(0.180)	(0.328)	(0.160)	(0.209)	(0.315)
Muslim * Born during Ramadan	0.157	-0.122	0.257	0.119	-0.088	0.277
	(0.263)	(0.290)	(0.466)	(0.262)	(0.253)	(0.468)
Panel B: Child is stunt						
Muslim * Conceived during Ramadan	-0.048	0.099	-0.209*	-0.046	0.102	-0.203*
	(0.117)	(0.151)	(0.120)	(0.117)	(0.150)	(0.119)
Muslim * In utero during Ramadan	0.030	0.090	-0.011	0.028	0.089	-0.011
	(0.048)	(0.071)	(0.067)	(0.047)	(0.070)	(0.068)
Muslim * Born during Ramadan	0.078	0.119	0.098	0.079	0.123	0.100
	(0.053)	(0.086)	(0.092)	(0.053)	(0.086)	(0.092)
No. of observations	7,097	3,625	3,472	6,671	3,410	3,261

Note: Estimation: Median regression in panel A and OLS in panel B. All regression include controls for month of birth, dummies for age in months, gender, rural-urban residence, region of residence, paternal educational and occupation status, household access to safe drinking water and toilet facility and maternal characteristics including education and occupation status. Standard errors, clustered at month of birth, year of birth and their interaction, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011 and the Ethiopian Socioeconomic Survey (ESS) 2014.

Chapter 3

Does Class Size Matter? Quasi-Experimental Evidence from Germany

BETHLEHEM A. ARGAW (Leibniz Universität Hannover)

PATRICK A. PUHANI (Leibniz Universität Hannover; CReAM, University College London; SEW, University of St. Gallen; IZA, Bonn)

The authors gratefully acknowledges access to the administrative linked teacher and student panel data (in German: *Lehrer- und Schülerdatenbank, LUSD*) for the school years 2007/08 until 2012/13 from the Hessisches Statistisches Landesamt. The views expressed in this article are those of the author and do not necessarily reflect the views of the Hessisches Statistisches Landesamt. While working on this paper, Bethlehem Argaw was a researcher in the Labour Markets, Human Resources and Social Policy Department of the Centre for European Economic Research (ZEW). The authors would like thank Maresa Sprietsma for helpful discussions and comments. Any remaining errors are our sole responsibility.

Abstract

We use a newly available administrative data on about a quarter of a million students in the German state of Hesse to estimate the causal effect of class size on educational outcomes. Our identification strategy relies on the quasi-random assignment of students to different class sizes based on maximum class size rules. In Germany, students are tracked into more or less academic middle school types at about age ten based, to a large extent, on academic achievement in elementary school. We find economically negligible and statistically insignificant effects of class size in elementary school on receiving a recommendation or the actual choice to attend the more academic middle school type.

Keywords: class size, panel administrative data, education production

JEL classification: I21, I28

3.1 Introduction

The policy of class size reductions has been at the center of educational research and policy debate for a few decades. The impact of class size on educational achievements is found to vary, among other dimensions, across school systems, grade levels, gender and students' socio-economic background. Whereas some empirical studies find a positive effect of smaller classes on short- and long-term outcomes (Card and Krueger, 1996; Angrist and Lavy, 1999; Krueger, 2003; Dustmann et al., 2003), part of the literature finds no substantial benefits from class size reductions (Hoxby, 2000; Levin, 2001; Dobbelsteen et al., 2002; Hanushek, 2003; Wößmann, 2005). Estimation bias arising from non-random sorting of students into classes of different sizes within and across schools has been well studied using randomized experiments (Krueger, 1999) or quasi-randomized experiments (Angrist and Lavy, 1999; Hoxby, 2000). Recently available administrative datasets, mainly from Scandinavian countries, revived the class size debate and improved the precision of previous results (Browning and Heinesen, 2007; Leuven et al., 2008; Fredriksson et al., 2013).

In this paper, we provide evidence on the causal effect of class size using a newly available administrative data from the German state of Hesse. The benefit of class size reductions in the context of an early school tracking system such as in Germany is scarcely studied. Although there exists some evidence on the relationship between class size and educational outcomes in Germany, to the best of our knowledge, Wößmann (2005) is the only study that uses a credible identification strategy to estimate the causal effect of class size in Germany. The author uses the TIMSS data from 15 Western European countries and finds no substantial benefit of smaller classes in lower middle schools in most countries including Germany. Our paper complements Wößmann (2005) by using a large administrative data which has important advantages over the TIMSS data to estimate the class-size effect in Germany.

One of the main advantage of the administrative teacher and student panel data (in German: *Lehrer- und Schülerdatenbank, LUSD*) relates to the measurement of class size and total enrollment. For a substantial share of the German sample in TIMSS, information on class size and total enrollment is not available and hence has to be imputed. When the data is available, class size and total enrollment are measured based on reports by teachers and school headmasters. Measurement errors arising from subjective reporting and the imputation method could thus bias the estimated class-size effect. In addition, total enrollment in TIMSS is measured towards the end of a school year instead of at the beginning. This is problematic because class sizes are determined based on enrollment size at the beginning of the school year and enrollment size could change non-randomly within a school year, for instance, if poor performing students either repeat grades or change schools. The administrative data, on the other hand, enables us to count the exact number of students in each class and total enrollment at the beginning of each school year.

Since the education system in Germany is decentralized across regions, there is no single maximum class size rule that governs all schools. The administrative data we use in this paper comes from the German state of Hesse and hence we can use the maximum class size rule for elementary schools as it is written in the Hessian school law (which is 25 students per class) to identify the causal effect of class size.[1] A further advantage of the administrative data is that it enables us to estimate the effect of exposure to smaller classes in elementary school. Previous evidence from other countries shows that smaller classes are most beneficial during primary or pre-primary education (Angrist and Lavy, 1999; Hoxby, 2000; Ding and Lehrer, 2010). Although the currently available evidence shows that smaller classes in German lower middle schools do not improve achievements, it still remains an open question whether or not class sizes in German primary schools influence educational outcomes.

[1] Wößmann (2005) relies on the pattern in the TIMSS data and uses a maximum class size rule of 30 for Germany. This figure is generated as the class size that best fits the relationship between actual class size and total enrollment in the German TIMSS sample.

The methodological challenge in establishing a cause-and-effect relationship between class size and educational outcomes arises due to the non-random sorting of students between and within schools. To the extent that school and family background characteristics relevant for students' academic performance remain unobservable and are correlated with class size, OLS gives a biased estimate of the class-size effect. In order to identify causal effects, our empirical strategy follows Angrist and Lavy (1999) and exploits the random assignment of students to different class sizes as a result of maximum class size rules. The assigned class size, which is generated based on total enrollment and the maximum class size rule, creates an exogenous variation in actual class size which is uncorrelated with other factors that influence students' academic performance.

We measure academic performance based on the observation that students in Germany are tracked into more or less academic middle school types at the end of elementary school. The type of middle schools that students are recommended to attend and the school type they eventually attend, to a large extent, depends on their academic performance in Math, German and General Studies in elementary school. We use an indicator for getting a recommendation to attend the higher and more academic school type called *Gymnasium* and the actual choice to attend this type of middle school as the main measures of students' educational outcome similar to Dustmann et al. (2016). We also use the probability to defer tracking until grade seven by attending the support stage called *Förderstufe* as additional outcome measure.

The instrumental variable estimation results show economically small and statistically insignificant effects of class size in 4^{th} grade on the probability of getting a recommendation or eventually attending the more academic school type in 5^{th} grade. In addition, class size does not significantly influence the probability of attending the support stage thereby deferring school tracking until grade seven. The class-size effect is precisely estimated and the magnitude of the estimates is close to zero. Results are robust to restricting the sample close to the discontinuity created by the maximum class size rule and to generating the instrument for class size based on total enrollment measured in grade one instead of grade four. We find a marginally significant but economically negligible negative effect of larger class size for female non-European nationals. In line with the previous evidence, we can, therefore, rule out any tangible benefit of smaller classes in Germany.

The rest of this article is structured as follows. In Section 3.2, we provide a brief overview of the literature. We then describe the German school tracking system and the recently available panel administrative data in Section 3.3.2. Section 3.4.2 discusses the empirical strategy and presents the main estimation results followed by sub-group analysis and robustness checks. Concluding remarks are provided in Section 3.5.

3.2 Overview of the Literature

Much of the debate on the effectiveness of class size reduction to improve students' educational outcomes revolves around getting an unbiased and precise estimate of the benefit of smaller classes. The non-random sorting of students into classes of different size within and between schools biases the least square estimates of the class size effect. Ideally, the causal effect of class size could be identified using randomized experiments. The Project STAR (Student-Teacher Achievement Ratio) in Tennessee is one such experiment which assigned first graders into classes of different sizes. Based on data from the Project STAR, Krueger (1999) finds significant improvement in mathematics and reading test scores of students who were exposed to smaller classes from kindergarten through third grade. In a subsequent study, Krueger and Whitmore (2001) find a positive effect of small classes on long-term outcomes such as the probability to take the ACT or SAT college entrance examinations, especially for minorities. Randomized experiments, however, have shortcomings and are not common in the class size literature.[2]

[2] The so-called Hawthorne effect which occurs if school administrators intentionally put extra effort to positively influence the performance of students in smaller classes in order to increase the likelihood of the experiment's broad implementation could lead to biased estimates (Hoxby, 2000). Furthermore, selective attrition and non-random change in treatment status over the experiment periods could bias the results of multi-period randomized experiments such as Project STAR (Ding and Lehrer, 2010).

Based on observational data sets, other studies have relied on quasi-experimental designs. There are two widely used quasi-experimental approaches in the class size literature. The first approach is pioneered by Angrist and Lavy (1999) and uses the exogenous source of variation in class size induced by the Maimonides' rule of 40 students in Israel's schools. Two otherwise similar school entry cohorts are assigned into classes of different size as a result of the variation in total enrollment and the maximum students allowed in a class. In a fuzzy regression discontinuity framework, Angrist and Lavy (1999) use the rule-induced class size as an instrumental variable for actual class size. They find significant improvement in reading and mathematics test scores for students taught in smaller classes in fourth and fifth grade. In another quasi-experimental approach, Hoxby (2000) uses the variation in class size across school entry cohorts arising from natural fluctuation in population size. The randomness in the timing of birth coupled with school entry age rules results in adjacent school entry cohorts starting school in classes of different size. Based on long panel data from Connecticut, Hoxby (2000) does not find any significant positive effect of small classes on math, reading and writing scores in fourth and sixth grade.

Quasi-experimental designs, especially based on maximum class size rules, have been applied to estimate the class size effect in different countries, at different school levels and using survey and/or administrative datasets. For instance, Browning and Heinesen (2007) for Denmark; Piketty (2004), Gary-Bobo and Mahjoub (2006) for France; Levin (2001) and Dobbelsteen et al. (2002) for the Netherlands; Bonesroenning (2003) and Leuven et al. (2008) for Norway; Fredriksson et al. (2013) for Sweden; Urquiola (2006) for Bolivia, Wößmann (2005), Wößmann and West (2006) for several Western countries. The evidence shows mixed results. Small class size is causally linked to higher test scores in Bolivia, France and Sweden, lower grade repetition in France, higher years of education and wages in Sweden and Denmark. On the contrary, no significant improvement in test scores due to smaller classes is found in the Netherlands and Germany, whereas the evidence is mixed for Norway and the US.

The evidence on the causal effect of class size in Germany is quite scarce. One of the reasons for the relatively scare evidence in Germany, and in European countries in general, is the limited data sets suitable for analyzing the benefits of smaller classes. Using the TIMSS data from 15 Western European countries including Germany, Wößmann (2005) exploits the maximum class size rule and the average class size variation between two adjacent classes in the TIMSS survey to account for the non-random sorting of students into classes of different size and finds no substantial benefit of smaller classes in most countries including Germany. Compared to the administrative data set we use in this paper, the main advantages of the TIMSS data is that it contains information on students' achievement on standardized tests and family background characteristics which are comparable across countries. However, this comes at a cost of small sample size and measurement errors for some of the countries considered in the study including Germany thereby making it difficult to reach country specific conclusion.

Administrative data sets are becoming accessible mainly in Scandinavian countries and are being widely used for empirical research in recent years. For instance, Leuven et al. (2008) uses Norwegian administrative data that contains nationwide test scores and finds that the effect of class size in elementary school is almost zero. Using administrative panel data from Denmark, Browning and Heinesen (2007) find a small negative effect of larger classes in eighth grade on the probability of completing secondary education. On the contrary, Fredriksson et al. (2013) uses Swedish administrative data matched with self-reported data on cognitive and non-cognitive skills and finds that individuals exposed to smaller classes in elementary school have substantially higher cognitive and non-cognitive skills, years of completed education and earnings as adults. We add the evidence from Germany to this recent literature that relies on administrative data to contribute to the class size debate.

3.3 Institutional Background and the Data

3.3.1 Secondary School Tracking in Germany

The education system in Germany is decentralized across the 16 federal states (*Länder*). In most federal states, including the state of Hesse, students are tracked into more or less academic middle school types at about age ten, i.e., after four years in elementary school.[3] The most academic school type, called *Gymnasium*, lasts about nine years and it is the only school type that provides direct access to tertiary academic education (university or university of applied science). The intermediate (*Realschule*) and lower (*Hauptschule*) school types take six and five years respectively and provide qualification for entry into dual-education where vocational schools are combined with apprenticeship training. Some share of students avoids streaming and enter comprehensive schools (*Gesamtschule*). There is also a possibility to postpone tracking until grade seven by attending the support stage (*Förderstufe*).

The trend in the distribution of students in each school type across grades is shown in Appendix Table A3.1. About 45 percent of six graders in Hesse attend the academic-oriented school type. The share declines slightly as years in middle school increases. This could be because students are downgrading to the non-academic school tracks or there is an increase in the share of birth cohorts that attend the higher track. Of the remaining 55 percent who attend the non-academic school types, students are equally distributed across the three school types - the intermediate school, comprehensive school and the support stage. It is worth noting that the share of students that avoid tracking by attending the comprehensive schools is increasing over the years (from 16 percent to 19 percent between the 2007/08 and 2011/12 school years). On the other hand, the share of students in grade 9 that attends the lowest middle school type has declined from 17 percent to 12 percent. This trend illustrates the ongoing debate in Germany on the timing and the extent of tracking after elementary school. In most states, there is

[3] Some states postpone middle school tracking until age 12 (end of grade 6) (Mühlenbweg, 2007).

a push towards abolishing the lower school track and expanding comprehensive schools and/or the so-called regional schools (*Regionalschule*) that merge the intermediate and the lower school types. The lower school track is either abolished or merged with the other schools types in Berlin, Hamburg, Saarland, Schleswig-Holstein and Rhineland-Palatinate (Fuchs, 2009; Wiechmann, 2011; Bartl, 2012).

The type of middle school that students attend is largely determined based on their academic achievement in elementary school. At the end of elementary school, students receive a school recommendation from primary school teachers based on their academic achievement in Mathematics, German and General Studies. The binding nature of the recommendation for the type of middle school students eventually attend varies across federal states. In Hesse, the track recommendation is not binding and hence parents make the final tracking decision. As a rule, it is possible to modify the initial school track choice. However, this is very uncommon mainly due to differences in curriculum between the school types and the physical segregation of schools (Mühlenbweg and Puhani, 2010).

3.3.2 The Hessian Administrative Data

We use the administrative teacher and student panel data (in German: *Lehrer- und Schülerdatenbank, LUSD*) that covers all students in the German state of Hesse for the school years 2007/08 until 2012/13.[4] It contains various measures of student-, teacher-, subject-, classroom- and school-level characteristics. In this paper, we use the information on student characteristics such as age, gender, country of origin, nationality, language used at home, the grade and type of middle school recommendation and the actual type of middle school attended. We restrict the sample to students who are in 4^{th} grade in each school year and we follow them into 5^{th} grade when they make transition into different types of middle schools. The administrative panel data contains six school years. However, we lose observations for one school year since the outcome variable in

[4] The administrative data also covers previous school years 2002/03 to 2006/07. However, a time-consistent student identifier, which allows us to follow students as they make transition into middle schools, was introduced in the 2007/08 school year.

school year t is measured based on data on middle school type in the following school year. Our estimation sample contains 258,098 students during the five school years.

We measure academic performance based on the observation that students in Germany are tracked into more or less academic middle school types at the end of elementary school similar to Dustmann et al. (2016). Unfortunately, the LUSD does not contain data on academic test scores. However, the type of middle schools that students are recommended to attend and the school type they eventually attend is a good proxy since it, to a large extent, depends on students' academic performance in Math, German and General Studies in elementary school. The first outcome variable takes the value one if a 4^{th} grader in year t gets a recommendation to attend the more academic school type (*Gymnasium*) and it takes the value zero if a student gets a recommendation to attend any of the non-academic school types. The second educational measure is defined in a similar notion but based on the actual middle school type that students attend in 5^{th} grade instead of the school recommendation. Whereas the recommendation is by statue strongly correlated with academic achievements in elementary school, the actual school type attended also captures the influence of parents in the tracking decision. As an additional outcome measure, we define an indicator for attending the support stage thereby deferring school tracking until grade seven.[5]

Table 3.1 shows the descriptive statistics. Column (1) shows averages and standard deviations for the full sample whereas column (2) limits the sample to observations with valid data on teacher track recommendation, which we refer to as the TR (Teacher Recommendation) sample.

[5] The school type recommendation data is missing for about one third of students in fifth grade. For every educational outcome, we provide separate estimates based on the full sample and the sample of students with non-missing data on school track recommendation, which we label as TR (track recommendation) sample. We do so in order to check for estimation bias that may result from non-random sample selection.

Table 3.1: Sample summary statistics

	Full Sample	TR Sample	Discontinuity Sample (+/- 5) Full Sample	Discontinuity Sample (+/- 5) TR Sample
Educational outcomes				
Higher track recommendation	-	0.52	-	0.52
Higher track attendance	0.45	0.54	0.46	0.54
Lower track recommendation	-	0.12	-	0.12
Lower track attendance	0.03	0.03	0.03	0.03
Support stage attendance	0.16	0.11	0.15	0.10
Student characteristics				
Age	9.86 (0.50)	9.84 (0.49)	9.85 (0.50)	9.83 (0.49)
Male	0.51	0.51	0.51	0.51
Non-European nationals	0.07	0.07	0.07	0.07
Non-European country of birth	0.06	0.06	0.06	0.05
Non-German language use at home	0.16	0.15	0.17	0.15
School characteristics				
Class size	20.37 (3.82)	20.45 (3.71)	21.08 (4.26)	21.21 (4.21)
Enrollment size	59.69 (26.49)	60.10 (26.67)	60.83 (26.18)	60.97 (26.67)
Number of school years	5	5	5	5
Number of schools	1,156	1,101	887	759
Number of classes	13,317	10,482	4,514	3,537
Number of students	258,098	168,063	90,217	59,060
Number of students with valid data on country of birth	153,240	97,506	62,026	33,882
Number of students with valid data on language used at home	138,367	87,949	52,218	30,842

Note: Standard deviations are shown in brackets when applicable. TR (Track Recommendation) sample refers to observations with valid data on school track recommendation. Data on country of origin and language use at home are only available for school year 2010/11 to 2012/13.
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*)

About half of fourth grade students receive a recommendation to attend the more academic school type. There is an overlap between the share of students who are recommended to attend the more academic track and the share who actually attend the same school type. Puhani (2016) shows that parental decision overlap by over 90 percent with the recommendation provided by teachers even though parents have the final say in the tracking decision. On the contrary, the share of students who attend the lower school type is half the share of students who receive a recommendation to attend the same school type. Finally, about one tenth of students in grade five attend the support stage where they can postpone school tracking until grade seven.

A typical classroom in fourth grade consists of 20 students on average. This is comparable to the average class size in the US and most Western European countries. Only six percent of students are taught in classes that exceed the maximum class size of 25. The average enrollment size is about 60 with a standard deviation of 26. About six percent of students are enrolled in schools with enrollment size of more than 100. Figure 3.1 shows the distribution of class size in the full estimation sample.[6]

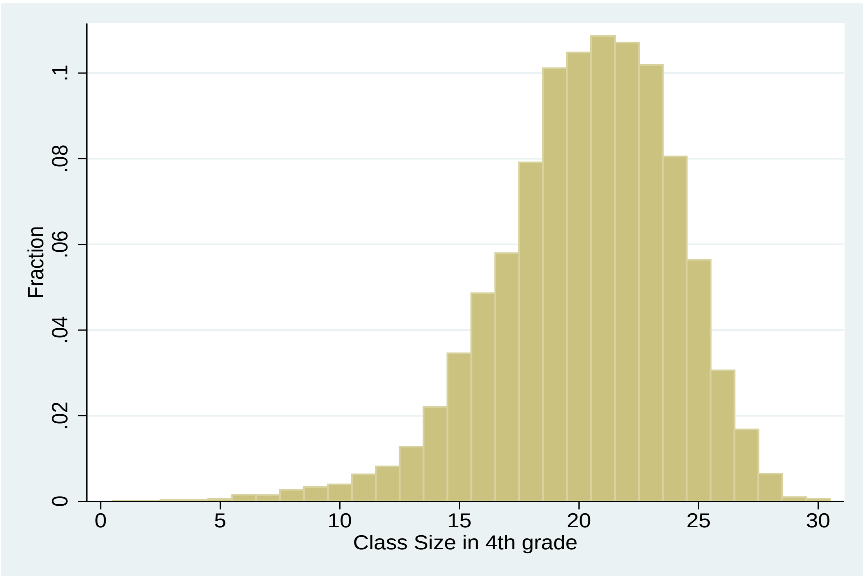


Figure 3.1: Distribution of class size.

[6] See Appendix Figure A3.1 for the distribution of total enrollment.

3.4 Empirical Strategy and Estimation Results

3.4.1 Identification Based on Maximum Class Size Rule

To disentangle the effect of class size from other factors that influence educational outcomes, our empirical strategy relies on maximum class size rules. The approach - pioneered by Angrist and Lavy (1999) - exploits the exogenous variation in actual class size arising from enrollment size and the rule that governs the maximum number of students to be taught in one class. This rule creates a discontinuity in the relationship between actual class size and total enrollment. As long as the factors influencing educational outcomes are smooth around the discontinuity, which we show to be the case, the assigned class size is a valid instrument for actual class size and hence can be used to identify the causal effect of class size on educational outcomes.

The assigned class size is obtained using the following formula.

$$ACS_{st} = E_{st}/[int(\frac{E_{st}-1}{25})+1] \tag{3.1}$$

where ACS_{st} is the assigned 4^{th} grade class size in school (s) at school year (t), E_{st} is the total enrollment in school (s) at school year (t) and 25 is the maximum class size rule in Hesse.

Figure 3.2 shows the relationship between total enrollment, actual class size and the assigned class size generated using equation (3.1).

Figure 3.2: Assigned class size and actual class size by total enrollment.

The figure shows the up-and-down pattern in the relationship between class size and total enrollment. There is a jump in actual and assigned class size whenever total enrollment reaches integer multiples of the maximum class size rule. Schools with total enrollment just below the threshold have relatively larger classes compared to schools with total enrollment just above the threshold. The deviation of the actual class size from the assigned class size indicates that a small number of schools does not strictly follow the maximum class size rule. The assigned class size is, therefore, used as an instrument for the actual class size following a fuzzy regression discontinuity design.

The quasi-experimental identification strategy based on the maximum class size rule relies on the assumption that all factors' influences relevant to students' educational outcomes, besides class size, are smooth around the point of discontinuity. School principals and/or parents, however, could be aware of the expected class size and might sort students into classes of different size based on the expected class size. If such manipulation of the rule occurred non-randomly, the estimated class-size effect would be biased. Non-random manipulation of the rule, however, becomes difficult as enrollment gets closer to the threshold. We therefore restrict the estimation sample to schools with total enrollments close to the discontinuity as a sensitivity check.

The first stage equation takes the following form:

$$CS_{st} = \alpha_0 + \alpha_1 ACS_{st} + \alpha_2 E_{st} + \gamma X_{ist} + \tau_s + \pi_t + \epsilon_{ist} \tag{3.2}$$

where CS_{st} is actual 4^{th} grade class size in school s during school year t and ACS_{st} is the assigned class size as defined in equation (3.1). E_{st} is a third order polynomial of enrollment size and the coefficient α_2 captures any remaining relationship between total enrollment and actual class size other than the deterministic relationship captured through ACS_{st}. X_{ist} is a vector of student level background characteristics, τ_s are school fixed effects and π_t are school year fixed effects.

The statistical strength of the correlation between actual class size and the assigned class size is shown in Table 3.2. The dependent variable is actual class size. OLS regression results are shown for the full sample, the sample with valid data on school track recommendation and the discontinuity sample.

Table 3.2: First stage estimation results - Dependent variable: Actual class size

	Full Sample		TR Sample		Discontinuity Sample (+/- 5)	
	(1)	(2)	(3)	(4)	(5)	(6)
Assigned class size	0.567^{***}	0.498^{***}	0.562^{***}	0.487^{***}	0.423^{***}	0.316^{***}
	(0.020)	(0.021)	(0.024)	(0.026)	(0.027)	(0.035)
Enrollment	√	√	√	√	√	√
Gender & Nationality	√	√	√	√	√	√
Year fixed effects	√	√	√	√	√	√
School fixed effects		√		√		√
F−statistic	840.94	541.27	618.93	358.17	287.85	82.92
R−squared	0.401	0.220	0.389	0.205	0.228	0.122
Number of schools	1,156		1,101		887	
Number of classes	13,317		10,482		4,514	
Number of students	258,098		168,063		90,217	

Note: Standard errors are clustered at the school level and shown in parenthesis. TR (Track Recommendation) sample refers to observations with valid data on school track recommendation. Enrollment refers to third order polynomial of total enrollment. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*)

The table shows that assigned class size is a very strong predictor of actual class size. The correlation is positive and significant at the one percent level even after controlling for school fixed effects. The F-statistics are well above the rule of thumb of 10 which is the threshold for a weak instrument (Staiger and Stock, 1997). The assigned class size remains a strong predictor of actual class size when we limit the sample to observations close to the discontinuity and with non-missing data on teacher recommendation.

3.4.2 Effects of Class Size on Educational Outcomes

We use the following equation to estimate the effect of class size on educational outcomes.

$$Y_{ist} = \beta_0 + \beta_1 CS_{st} + \delta X_{ist} + \tau_s + \pi_t + \mu_{ist} \tag{3.3}$$

where Y_{ist} is the educational outcome of student i in school s during school year t that is either the teacher's school type recommendation or the actual school type attended in 5^{th} grade. CS_{st} is the actual class size in 4^{th} grade which is instrumented using the assigned class size from equation (3.1). X_{ist} is a vector of controls such as students' gender, age and nationality. It also includes a third order polynomial of enrollment size in order to capture any remaining influence of enrollment size on educational outcome other than its deterministic relationship with the assigned class size, τ_s are school fixed effects which control for sorting of students between schools whereas π_t are school year fixed effects.

Table 3.3 presents the estimation results for the main sample and Table 3.4 shows results for the discontinuity sample. Panel A shows the estimation results from an OLS regression. Panel B reports the reduced form regression results where the educational outcome measures are regressed on assigned class size and other controls. Results from instrumental variable regressions are presented in Panel C.

Table 3.3: Main estimation results: Coefficient of class size

	Teacher school recommendation		School track choice: Full sample		School track choice: TR Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. OLS regression						
Class size	0.001	-0.000	0.002**	0.001	0.000	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Panel B. Reduced form regression						
Assigned class size	-0.001	-0.001	0.001	-0.000	-0.001	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Panel C. Instrumental variable regression						
Class size	-0.001	-0.003	0.002	-0.001	-0.001	-0.003*
	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
School fixed effects		√		√		√
Number of schools	1,101		1,156		1,101	
Number of classes	10,482		13,317		10,482	
Number of students	168,063		258,098		168,063	

Note: Standard errors are clustered at the school level and shown in parenthesis. TR (Track Recommendation) sample refers to observations with valid data on school track recommendation. All regressions include controls for enrollment, gender, nationality and school year fixed effects. Enrollment refers to third order polynomial of total enrollment. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*)

Most of the regression results show insignificant effects of class size on the probability of getting a recommendation to attend the more academic school track and on the actual choice to attend the same school type. The statistically positive correlation that we find in the OLS regression based on the full sample in column (3) becomes insignificant when we control for school fixed effects. The class-size effect is precisely estimated and the magnitude of the estimates is close to zero. The estimated class-size effects from the reduced form and instrumental variable regressions have the expected negative sign and are precisely estimated.

Interpreting the estimated class size effect from the instrumental variable regression shown on Panel C without controlling school fixed effects, we find that an increase of class size by 10 students reduces the probability of going to the academic school track by 0.1 to 0.2 percentage points with a standard error of 0.001 or 0.002. The estimated class size effect on school track choice increases to 0.3 percentage points and becomes statistically significant at the 10 percent level when we control for school fixed effects based on the sample with valid data on school track recommendation. However, the estimated class size effect becomes statistically insignificant when we use the discontinuity sample as shown in Table 3.4 below.

Table 3.4: Estimation results based on discontinuity sample (+/-5)

	Teacher school recommendation		School track choice: Full sample		School track choice: TR Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
A. OLS regression						
Class size	-0.000	0.000	0.003**	0.002***	-0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
B. Reduced form regression						
Assigned class size	-0.001	-0.002	-0.001	-0.000	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
C. Instrumental variable regression						
Class size	-0.003	-0.005	0.002	-0.001	-0.002	-0.002
	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)
School fixed effects		√		√		√
Number of schools	759		887		759	
Number of classes	3,537		4,514		3,537	
Number of students	59,060		90,217		59,060	

Note: Standard errors are clustered at the school level and shown in parenthesis. TR (Track Recommendation) sample refers to observations with valid data on school track recommendation. All regressions include controls for enrollment, gender, nationality and school year fixed effects. Enrollment refers to third order polynomial of total enrollment. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*)

The estimation results based on the discontinuity sample shown in Table 3.4 are very similar to the results based on the main sample shown in Table 3.3. We can, therefore, rule out any substantial influence of class size on the type of school that students attend or are recommended to attend in middle schools in Germany.

3.4.3 Robustness Checks and Heterogeneous Effects

The results from the main estimation show that reducing class size in elementary school does not increase the probability of attending or getting a recommendation to attend the higher and more academic school track in Germany. Smaller classes, however, could influence the probability of not attending or not getting a recommendation to attend the lower and more applied school track (*Hauptschule*). As an alternative outcome measure, we use a binary variable that takes the value one if students receive a recommendation to attend any middle school type other than the lower school track. A second alternative outcome variable is defined following a similar notion but based on the actual school choice rather than the recommendation. We also define a third outcome variable which takes the value one if students attend the support stage thereby deferring school tracking until grade seven.

Table 3.5 reports the instrumental variable estimation results for the full sample in column (1) and (2), the track recommendation sample in column (3) and (4) and the discontinuity sample in column (5) and (6). The first column under the respective sample shows estimation results without including school fixed effects whereas we include school fixed effects in the second column under the respective sample.

The estimated class-size effects from the school fixed effects regressions show that students in larger classes are more likely to get a recommendation and to actually attend the lower school track. The estimated coefficient shows that an increase of class size by 10 students lowers the probability of not attending or not getting a recommendation to attend the lower and more applied school track by 0.1 to 0.2 percentage points. Students in large classes are also more likely to postpone tracking until grade seven by attending the support stage. However, similar to the main results shown in Table 3.3, the instrumental variable estimates are not statistically different from zero.

Table 3.5: Instrumental variable estimation results: Additional educational outcomes

	Full sample		TR sample		Discontinuity sample (+/- 5)	
	(1)	(2)	(3)	(4)	(5)	(6)
A. School track recommendation: Gymnasium/Realschule vs Hauptschule						
Class size	-	-	0.001	-0.000	-	-
			(0.001)	(0.001)		
B. School track choice: Gymnasium/Realschule/Gesamtschule vs Hauptschule						
Assigned class size	0.003	-0.001	0.000	-0.002	0.004	-0.001
	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)	(0.002)
C. Postpone school tracking until grade seven						
Class size	-0.004	0.001	-0.001	0.001	-0.005	0.002
	(0.002)	(0.001)	(0.003)	(0.001)	(0.004)	(0.002)
School fixed effects		√		√		√
Number of schools	1,156		1,101		887	
Number of classes	13,317		10,482		4,514	
Number of students	258,098		168,063		90,217	

Note: Standard errors are clustered at the school level and shown in parenthesis. TR (Track Recommendation) sample refers to observations with valid data on school track recommendation. All regressions include controls for enrollment, gender, nationality and school year fixed effects. Enrollment refers to third order polynomial of total enrollment. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*)

As mentioned in Section 3.3.2, one of the unique features of the Hessian administrative data is that it contains a time-consistent student identifier which makes it possible to follow students as they increase their level of schooling. So far, we used the panel dimension of the administrative data in order to measure class size at the end of elementary school and the subsequent educational outcomes at the beginning of middle school. The administrative data allows us to follow two school cohorts from grade one until they make a transition to different types of middle schools. By restricting the sample to these two school cohorts, we are able to generate an instrument for class size based on total enrollment in grade one instead of grade four. This may be important because enrollment size in grade four could be endogenous due to, for instance, grade repetition in lower grades. In addition, schools are more likely to implement the maximum class size rule when students start schools in grade one. Hence, the validity of the instrument for actual class size is stronger when the assigned class size is generated based on total enrollment in grade one instead of grade four.

Table 3.6 shows instrumental variable estimation results. As the impact variable, we use class size measured in grade one in panel A and in grade four in panel B.

Table 3.6: Instrumental variable estimates based on total enrollment in first grade

	Teacher school recommendation		School track choice		School track choice: TR Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
A. Class size in grade one						
Class size	-0.003	-0.001	-0.001	-0.001	-0.003	-0.002
	(0.003)	(0.004)	(0.003)	(0.002)	(0.004)	(0.004)
First stage: Assigned class size	0.571^{***}	0.454^{***}	0.553^{***}	0.453^{***}	0.571^{***}	0.454^{***}
	(0.038)	(0.046)	(0.033)	(0.038)	(0.038)	(0.046)
F-statistic	252.51	96.79	304.94	139.06	252.51	96.79
B. Class size in grade four						
Class size	-0.004	-0.001	-0.001	-0.002	-0.004	-0.003
	(0.004)	(0.005)	(0.004)	(0.003)	(0.004)	(0.005)
First stage: Assigned class size	0.477^{***}	0.363^{***}	0.431^{***}	0.345^{***}	0.435^{***}	0.458^{***}
	(0.031)	(0.040)	(0.028)	(0.034)	(0.039)	(0.046)
F-statistic	284.86	82.31	280.68	99.65	284.86	82.31
School fixed effect		√		√		√
Number of school cohorts	2		2		2	
Number of schools	987		1,185		987	
Number of classes in grade 1	2,167		2,615		2,167	
Number of classes in grade 4	2,058		2,706		2,058	
Number of students in grade 1	56,803		90,841		56,803	
Number of students in grade 4	56,751		90,841		56,751	

Note: Standard errors are clustered at the school level and shown in parenthesis. TR (Track Recommendation) sample refers to observations with valid data on school track recommendation. All regressions include controls for enrollment, gender, nationality and school year fixed effects. Enrollment refers to third order polynomial of total enrollment. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*)

The coefficients of assigned class size from the first stage regression results show that the assigned class size generated based on total enrollment in grade one is a strong predictor of class size in grade one compared to class size in grade four. The F-statistics from the school fixed effects regression is also larger when the assigned class size is generated based on total enrollment in grade one compared to grade four.

The estimated class size effects has the expected negative sign implying that larger classes in elementary school are detrimental to school track outcomes in middle school. The magnitude of the estimated class size effects and their statistical insignificance are similar to the results in our main specification shown in Table 3.3 (0.1 to 0.3 percentage points reduction in the probability of going to the academic school track or getting a recommendation to go to the academic school track if class size increases by 10 students).

Previous studies find a larger benefit of smaller classes for students from a disadvantaged background where disadvantaged background is defined, for instance, based on parental education, family size, number of books at home and location of schools. Unfortunately, the Hessian administrative data does not contain information on direct measures of disadvantaged family background. In order to test for heterogeneous class-size effects, we rely on information on students' nationality, country of origin and the language used at home as proxy measures for family background. We define three variables each indicating whether the student is a non-European national, whether the student is born outside Europe, and whether the language a student uses at home is not German.

Instrumental variable estimation results are shown in Table 3.7. Panel A shows results where we interact the class size variable with being male and a non-European national. Similarly, the class size variable is interacted with being male and being born outside of Europe in Panel B whereas class size is interacted with being male and using a language other than German at home in Panel C. The educational outcome variables are defined as before.

Table 3.7: Instrumental variable stimation results - Heterogenous class-size effects

	Teacher school recommendation		School track choice: Full sample		School track choice: TR Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
A. Non-European defined based on citizenship						
Class size (CS)	-0.003	-0.003*	0.002	-0.000	-0.002	-0.003
	(0.002)	(0.002)	(0.002)	(0.001)	(0.003)	(0.002)
CS * Male	-0.000	-0.001	-0.000	-0.000	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
CS * Non-European	-0.008*	-0.005	-0.011***	-0.005*	-0.010*	-0.008*
	(0.004)	(0.004)	(0.003)	(0.003)	(0.005)	(0.004)
CS * Male * Non-European	0.011**	0.011**	0.008**	0.008**	0.014**	0.015***
	(0.005)	(0.005)	(0.004)	(0.004)	(0.006)	(0.006)
Number of students	168,063		258,098		168,063	
B. Non-European defined based on country of origin						
Class size (CS)	-0.004	-0.003	0.001	-0.001	-0.002	-0.002
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
CS * Male	0.002	0.001	0.000	0.000	0.001	0.000
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
CS * Non-European	-0.001	0.005	-0.003	0.005	-0.001	0.004
	(0.007)	(0.007)	(0.005)	(0.005)	(0.007)	(0.006)
CS * Male * Non-European	0.002	-0.006	-0.006	-0.010	-0.001	-0.008
	(0.009)	(0.009)	(0.006)	(0.006)	(0.009)	(0.009)
Number of students	97,506		153,240		97,506	
C. Non-German defined based on language use at home						
Class size (CS)	-0.003	-0.002*	0.001	-0.001	-0.001	-0.001
	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.003)
CS * Male	0.001	-0.000	0.000	-0.000	-0.000	-0.001
	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.002)
CS * Non-German	-0.003	-0.003	-0.004	0.001	-0.004	-0.003
	(0.005)	(0.004)	(0.003)	(0.003)	(0.005)	(0.004)
CS * Male * Non-German	0.004	0.005	-0.002	-0.002	-0.001	0.004
	(0.006)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)
Number of students with valid	87,949		138,367		87,949	
School fixed effects		√		√		√

Note: Standard errors are clustered at the school level and shown in parenthesis.TR (Track Recommendation) sample refers to observations with valid data on school track recommendation. All regressions include controls for enrollment and school year fixed effects. Enrollment refers to third order polynomial of total enrollment. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*)

The conclusion of the results in Table 3.7 is that there is little evidence for any heterogeneous effect of class size. We find a marginally significant (at the 10 percent level) negative effect of larger classes on getting a recommendation or actually attending the more academic school type for female non-European nationals after including school fixed effects. The estimated class size effects for females, which is obtained by adding the estimates for *Class Size (CS)* and *CS*Non-European*, from the school fixed effect regression show that an increase of class size by 10 students reduces the probability of going to the academic school track or being recommended to join the academic school track by 0.5 to 1 percentage points. The heterogeneous effect, however, is sensitive as to how we define non-Europeans. The marginally significant heterogeneous effects we find among non-European nationals disappears when we define non-Europeans based on their country of birth. The use of non-German language at home, which serves as a better family background measure that is relevant for students' educational outcomes, does not yield any heterogeneous effects.

3.5 Conclusion

A student-level administrative data from the German state of Hesse allowed us to precisely estimate the causal effect of class size in elementary schools. Exploiting the exogenous variation in class size arising from maximum class size rules, despite of using "big data" we find no significant effect of class size in primary schools on the type of middle schools students are tracked to at about age ten. The size of the point estimate is very close to zero for all measures of educational outcome considered including the probability of getting a recommendation to attend the more academic school type, the actual choice to attend the more academic school type and the probability to defer middle school tracking until grade seven. The result that smaller class size does not substantially improve educational outcomes is in line with the previous evidence for Germany.

The class-size effect remains negligible for a number of robustness checks including restricting the sample close to the discontinuity and generating alternative instruments for class size based on total enrollment in the first grade instead of fourth grade. Furthermore, we do not find heterogeneous effects of class size by students' gender, nationality, country of origin and language used at home. However, since a direct measure of family background characteristics such as parental education or income is not available in the Hessian administrative data, we cannot rule out the possibility that smaller classes in primary schools could be beneficial for students from a disadvantaged background.

The potential explanation for the insignificant effect of class size we find in this paper could be that the average size of classes in Germany is quite small and hence it might be below the threshold that is relevant for students' academic achievement. The other explanation is related to the educational outcome measures available in the administrative data. The type of middle school students attend or are recommended to attend could be imprecise measures of academic achievement compared to test scores or academic grades. This calls for further research that matches the administrative data we use in this paper with survey or other administrative data sources that contain precise measures of students' academic achievement along with richer information on family background characteristics.

3.6 References

Angrist, J. D. and Lavy, V. (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics*, 114(2):533–575.

Bartl, W. (2012). *Entdifferenzierung der regionalen Schulstruktur in Deutschland durch demografischen Wandel?* Martin-Luther-Universität Halle-Wittenberg, Institut für Soziologie.

Bonesroenning, H. (2003). Class size effects on student achievement in norway: Patterns and explanations. *Southern Economic Journal*, 69(4):952–965.

Browning, M. and Heinesen, E. (2007). Class size, teacher hours and educational attainment. *Scandinavian Journal of Economics*, 109(2):415–38.

Card, D. and Krueger, A. B. (1996). School resources and student outcomes: An overview of the literature and new evidence from north and south carolina. *Journal of Economic Perspectives*, 10(4):31–50.

Ding, W. and Lehrer, S. F. (2010). Estimating treatment effects from contaminated multiperiod education experiments: The dynamics impacts of class size reductions. *Review of Economics and Statistics*, 92 (1):31–42.

Dobbelsteen, S., Levin, J., and Oosterbeek, H. (2002). The causal effect of class size on scholastic achievement: Distinguishing the pure class size effect from the effect of changes in class composition. *Oxford Bulletin of Economic and Statistics*, 64(17):19–38.

Dustmann, C., Puhani, P. A., and Schönberg, U. (2016). The long-term effects of early track choice. *Economic Journal*.

Dustmann, C., Rajah, N., and van Soest, A. (2003). Class size, education, and wages. *Economic Journal*, 113:F99–120.

Fredriksson, P., Oeckert, B., and Oosterbeek, H. (2013). Long-term effects of class size. *The Quarterly Journal of Economics*, 128 (1):249–285.

Fuchs, H.-W. (2009). Strukturen und strukturreform im allgemeinbildenden schulwesen der deutschen bundesländer. *Die Deutsche Schule*, 101(1):7–19.

Gary-Bobo, R. J. and Mahjoub, M. B. (2006). Estimation of class-size effects using 'miamonides' rule': The case of french junior high schools. *CEPR Discussion Paper No. 5754*.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *Economic Journal*, 113:F64–F98.

Hoxby, C. M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics*, 115:1239–85.

Krueger, A. B. (1999). Experimental estimates of educational production functions. *Quarterly Journal of Economics*, 114(2):497–532.

Krueger, A. B. (2003). Economic considerations and class size. *Economic Journal*, 113:F34–F63.

Krueger, A. B. and Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project star. *Economic Journal*, 111 (468):1–28.

Leuven, E., Oosterbeek, H., and Roenning, M. (2008). Quasi-experimental estimates of the effect of class size on achievement in norway. *Scandinavian Journal of Economics*, 110(4):663–693.

Levin, J. (2001). For whom the reductions count: A quantile regression analysis of class size and peer effects on scholastic achievement. *Empirical Economics*, 26:221–46.

Mühlenbweg, A. M. (2007). Educational effects of early or later secondary school tracking in germany. *ZEW-Centre for European Economic Research Discussion Paper 07-079.*

Mühlenbweg, A. M. and Puhani, P. A. (2010). The evolution of the school-entry age effect in a school tracking system. *Journal of human resources*, 45(2):407–438.

Piketty, T. (2004). Should we reduce class size or school segregation? theory and evidence from france. *Presentation at the Roy Seminars, Association pour le développement de la recherche en économie et en statistique (ADRES), available at: http://adres.ens.fr/IMG/pdf/22112004b.pdf, accessed on February 20, 2013.*

Puhani, P. A. (2016). Do boys benefit from male teachers in elementary school? evidence from administrative panel data. *mimeo*.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557.

Urquiola, M. (2006). Identifying class size effects in developing countries: Evidence from rural bolivia. *Review of Economics and Statistics*, 88:171–76.

Wiechmann, J. (2011). Vollständiger systemwandel in schleswig-holstein. *Zeitschrift für Erziehungswissenschaft*, 14(1):119–139.

Wößmann, L. (2005). Educational production in europe. *Economic Policy*, 20(43):445–504.

Wößmann, L. and West, M. R. (2006). Class-size effects in school systems around the world: Evidene from between-grade variation in timss. *European Economic Review*, 50(3):695–736.

3.7 Appendix

Table A3.1: Distribution of students across grades and school types

	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9
School Year 2007/08					
Higher School Type	0.44	0.42	0.42	0.40	0.37
Non-Higher School Types					
Intermediate School	0.17	0.16	0.28	0.29	0.30
Lower School	0.04	0.05	0.14	0.16	0.17
Comprehensive School	0.18	0.18	0.17	0.16	0.16
Support Stage	0.18	0.19	-	-	-
Number of Students	59,264	57,259	56,940	59,659	60,570
School Year 2011/12					
Higher School Type	0.46	0.44	0.44	0.42	0.40
Non-Higher School Types					
Intermediate School	0.16	0.16	0.26	0.28	0.29
Lower School	0.03	0.04	0.10	0.12	0.12
Comprehensive School	0.20	0.19	0.20	0.19	0.19
Support Stage	0.16	0.16	-	-	-
Number of Students	53,322	56,221	58,290	60,300	62,310

Source: Administrative Teacher and Student Data Set for the State of Hesse 2007/08-2012/13 (*Lehrer- und Schülerdatenbank, LUSD*)

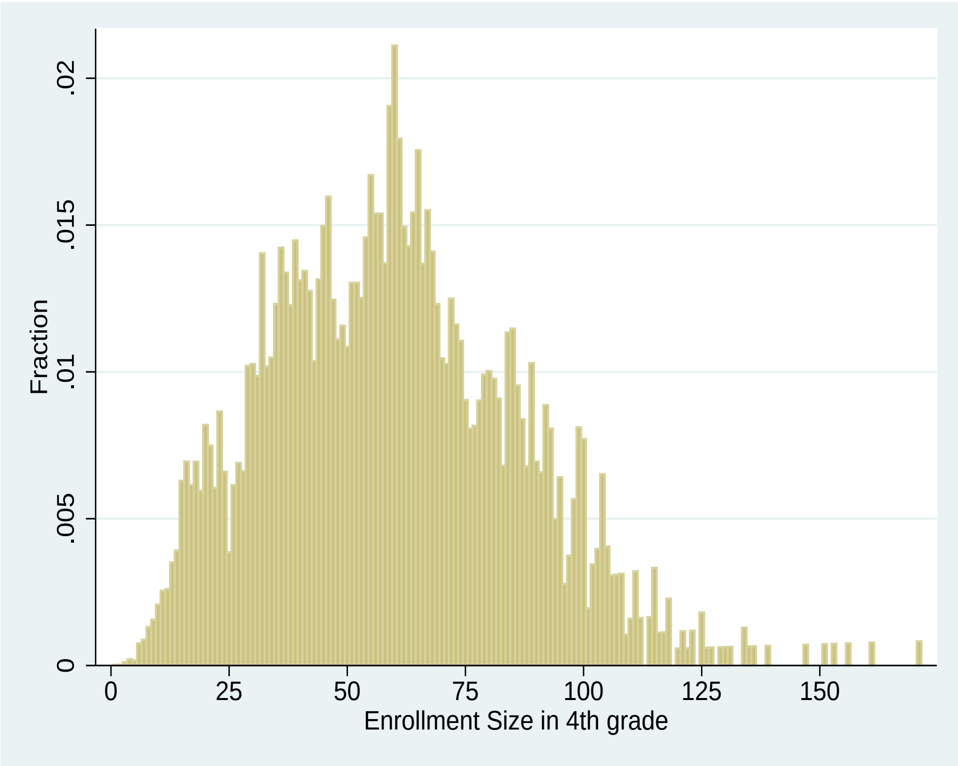


Figure A3.1: Distribution of enrollment.

Chapter 4

Mother Tongue Education, Reading Skills and Labour Market Outcomes

BETHLEHEM A. ARGAW (Leibniz Universität Hannover)

This chapter is based on the discussion paper "Argaw, Bethlehem A. (2016). Quasi-Experimental Evidence on the Effects of Mother Tongue-Based Education on Reading Skills and Early Labour Market Outcomes. *ZEW Discussion Paper No.16-016*, Mannheim" which is available at `http://ftp.zew.de/pub/zew-docs/dp/dp16016.pdf`.

The author gratefully acknowledge support from the German Ministry of Education and Research (BMBF) under grant 01PW11001 under the framework "The Economics of Science". The author also wish to thank UNU-WIDER for providing support and a stimulating environment during my PhD internship in Helsinki, Finland. While working on this paper, the author was a PhD intern at the UNU-WIDER and a researcher in the Labour Markets, Human Resources and Social Policy Department of the Centre for European Economic Research (ZEW). Free data access from the DHS program, USAID, and RTI are also greatly appreciated. The views expressed in this article are those of the author and do not necessarily reflect the views of the BMBF, UNU-WIDER, USAID, or RTI. The author would like to thank Patrick Puhani, Rachel Gisselquist, Saurabh Singhal, Friedhelm Pfeiffer, Maresa Sprietsma and Abraham Asfaw for helpful comments and discussions. An earlier version of the paper was presented at the University of Hannover, CSAE2014, SOLE2014, and SMYE2014 and feedback from conference participants is greatly appreciated. Any remaining errors are my sole responsibility.

Abstract

Prior to the introduction of mother tongue based education in 1994, the language of instruction for most subjects in Ethiopia's primary schools was the official language (Amharic) - the mother tongue of only one third of the population. This paper uses the variation in individual's exposure to the policy change across birth cohorts and mother tongues to estimate the effects of language of instruction on reading skills and early labour market outcomes. The results indicate that the reading skills of birth cohorts that gained access to mother tongue-based primary education after 1994 improved significantly by about 11 percentage points. The provision of primary education in mother tongue halved the reading skills gap between Amharic and non-Amharic mother tongue users. The improved reading skills seem to translate into gains in the labour market in terms of the skill contents of jobs held and the type of payment individuals receive for their work. An increase in school enrollment and enhanced parental educational investment at home are identified as potential channels linking mother tongue instruction and an improvement in reading skills.

JEL classification: I24, I25, I28, J24

Keywords: language of instruction, mother tongue, reading skills, labour market

4.1 Introduction

Mother tongue instruction especially during the early grades of schooling has been advocated as a silver bullet to increase educational attainment, to improve learning outcomes and to reduce educational inequality across groups (Bender et al., 2005; Smits et al., 2008). Proponents of mother tongue-based education point that school curricula are better communicated when students are taught in their mother tongue. The empirical evidence show that using mother tongue as the language of learning and teaching reduces grade repetition and dropout rates and increases school enrollment and academic achievements (Walter and Chou, 2012; Walter and Dekker, 2011; Heugh et al., 2012).

The main arguments put against mother tongue-based education is that it could reduce proficiency skills in the dominant language on the labour market thereby limiting employment and other economic opportunities in the long-term. The empirical evidence on the link between mother tongue instruction, proficiency in second language and long-term economic outcomes is rather mixed (Berman et al., 2003; Bleakley and Chin, 2004; Aldashev et al., 2009). This paper examines the relationship between mother tongue instruction, reading skills and early labour market outcomes.

It is empirically challenging to disentangle the impacts of language of instruction from other factors that influence schooling outcomes. Parents who value education could place their children in schools where education is provided in the child's mother tongue. In addition, schools that provide instructions in mother tongue may have more engaged school directors, which may affect learning outcomes independently of the language of teaching. In such cases, the positive correlation between schooling outcomes and the language used for instruction is not necessarily caused by the language of instruction per se. Instead, unmeasured and/or unobserved school characteristics and parental value for education could cause the positive correlation. One could randomly allocate students or schools into a treatment mother-tongue based group and a control group. Any statistically significant differences in the learning outcomes of the two groups could be attributed to the language of instruction. However, this kind of experiment is costly and seldom.[1]

School language policy reforms can serve as a "natural experiment" to examine the effects of language of instruction on reading skills and labour market outcomes. Ethiopia is one of the few African countries that introduced mother tongue-based primary education. Prior to 1994, the language of instruction for most subjects during primary education was the official language (Amharic), although only 30 percent of the population have Amharic as their mother tongue. As from 1994, mother tongue-based education in public schools was mandated throughout primary education. We use the quasi-exogenous variation in exposure to the language policy change in a difference-in-differences (DID) framework to identify the effects of language of instruction.

Estimation results using the 2011 Ethiopian Demographic and Health Survey show that the reform has a positive influence on reading skills. Reading skills is measured by the ability to read sentences written on a card in one of the major languages spoken in Ethiopia (Amharic, Afan Oromo, Tigrinya, Afar and Somali). Accessing primary education in mother tongue increases the probability of non-Amharic speakers' ability to read by about 11 percentage points. The DID estimates also show that the provision of mother tongue-based education has reduced the reading skills gap between Amharic and non-Amharic mother tongue users by half. Household socioeconomic status and non-random regional migration as a response to the reform do not confound the results.

The paper identifies two possible channels linking instruction based on mother tongue and an improvement in reading skills. First, school enrollment is significantly higher among birth cohorts that gained access to mother-tongue instruction after 1994. This implies that providing education in mother tongue increases the accessibility of schools. Second, parents are more likely to invest in their children's education outside schools when their children are taught in mother tongue. Parental investment takes the form

[1] Examples include small scale experiments in Cameroon (Walter and Chou, 2012), Vietnam (UNICEF, 2011) and the Philippines (Walter and Dekker, 2011).

of time investment where parents directly help children with homework and/or financial investment where parents hire a tutor to help their children with homework.

The improved reading skills and general human capital have translated into gains in the labour market. This is measured in terms of the skill contents of jobs held and the type of payment an individual receives for his/her work. Individuals whose mother tongue is different from Amharic and who gained access to mother tongue-based education after 1994 are more likely to work as skilled workers, mainly in the agriculture and service sector, as opposed to being unskilled agricultural or manual workers. They are also more likely to receive their payments in cash instead of informal means of payments such as in kind, a combination of cash and in kind or not receiving any payments for their work. This paper, hence, provides evidence against the view that mother tongue education hinders employment and other opportunities in the labour market. Rather, the findings underscore the importance of mother-tongue based education to reduce ethno-linguistic inequality with respect to human capital and labour market status.

This paper is closely related to few studies that use language policy reforms to examine the effects of language of instruction. Angrist and Lavy (1997) explore a policy change in Morocco which replaced French with Arabic as the language of instruction for post-primary education in 1984. Using the variation in exposure to the policy change across education groups, the authors find that individuals who undertook their education in Arabic have lower French writing skills and hence they experience lower wage returns to schooling. Puerto Rican public schools also introduced Spanish as the language of teaching replacing English in 1949. Contradictory to the Moroccan experience, Angrist et al. (2006) show that the use of Spanish for instruction in post-primary education has not contributed to the declining English proficiency of Puerto Ricans. Mother tongue instruction was mandated in South Africa's primary schools in 1955 through the Bantu Education Act. Exploiting the regional variation in the duration of mother tongue instruction, Eriksson (2014) finds positive effects not only on literacy skills in mother tongue, but also on English speaking skills, educational attainment and labour market outcomes.[2]

To the best of the author's knowledge, Ramachandran (2012) is the only paper that uses the 1994 language policy change in Ethiopia as a "natural experiment" to examine the effects of language of instruction. The author show that the years of education of birth cohorts affected by the language reform increased on average by one year. The author also identifies the intensive margin of education, measured by higher completion rate of grade six and above, as the mechanism at work. The effects that Ramachandran (2012) find is partly due to another reform that took place in 1994 to abolish an Amharic language based central examination that students were required to pass at the end of grade six. The main differences between Ramachandran (2012) and this paper are the following: i) we estimate the effects of language of instruction on learning and labour market outcomes ii) we isolate the effects of language of instruction from the effects of the central examination abolition by restricting the analysis to individuals whose maximum years of education is six years, since these individuals were not affected by the central examination abolition and iii) we identify parental educational investment at home as one channel linking mother tongue education and learning outcomes.

This paper's contributions to the existing literature on language of instruction are the following. First, the paper explores, for the first time, the effect of the language of instruction on reading skills within the context of the 1994 language reform in Ethiopia. It, therefore, complements the existing literature by identifying one mechanism linking language of instruction and educational attainment. Methodologically, the 1994 language reform offers an advantage since it allows us to exploit the variation across language groups over birth cohorts for identification. Previous studies that rely on variation across education groups for identification could introduce bias due to an endogenous schooling response to the policy change. Second, the paper tests, for the first time, whether or not parents engage more in their child's education outside of schools when formal education

[2] Using a fixed effect approach to control for sorting of students between and within South African schools, Taylor and Coetzee (2013) also finds an improvement in the English language skills of pupils in middle schools when taught in mother tongue in early grades of primary education.

is provided in mother tongue. Parental educational engagement is hypothesized to be one of the potential channel through which mother tongue education affects schooling outcomes (Eriksson, 2014). Finally, the paper contributes to the scarce evidence on the long-term effects of mother tongue-based education.

The rest of the paper is structured as follows. Section 4.2 provides background information on education in Ethiopia and the 1994 educational reform. In Section 4.3, the identification strategy is outlined whereas Section 4.4 describes the data. Estimation results are presented and discussed in Section 4.5 and Section 4.6 concludes.

4.2 Institutional Background

Formal education is a recent phenomenon in Ethiopia. Prior to the 1950s, education was provided through the Ethiopian Orthodox Church which predominantly uses the Amharic language and an ancient language called Ge'ez. Literacy skill and educational attainment remained quite poor despite the policies implemented in the country since then. Some of the major policies were the 1974 Education Sector Review and the National Literacy Campaign which was implemented between 1979 and 1983. With the aim of improving the "relevance, quality, accessibility and equity" of the education system, the Ethiopian People's Revolutionary Democratic Front (EPRDF) undertook a major educational reform in 1994. Through the Education and Training Policy (ETP), the organizational structure of the education system was reformed. Figure 4.1 shows the structure of the Ethiopian education system before and after 1994.

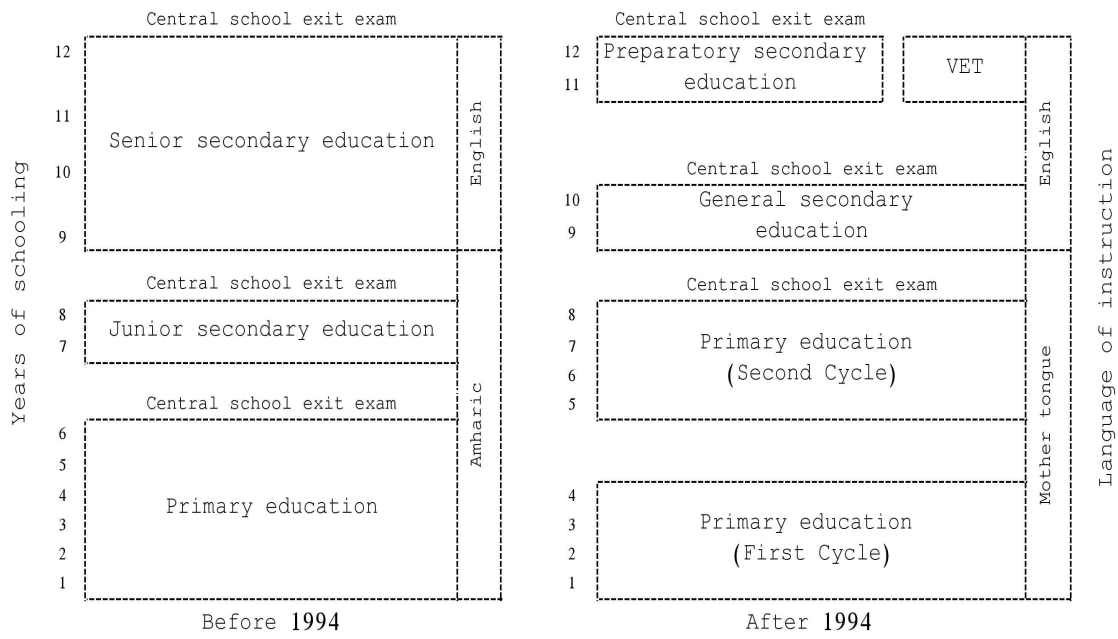


Figure 4.1: Ethiopian education system before and after 1994.

Source: Author's illustration

Prior to 1994, the education system consisted of primary education (grades 1-6), junior secondary education (grades 7-8) and senior secondary education (grades 9-12). Primary education in the current education system lasts from grade one to eight and is divided into a first cycle (grades 1-4) and a second cycle (grades 5-8). Secondary education consists of a general secondary education (grades 9-10) and a preparatory secondary education (grades 11-12). Central school exit examinations are administered at the end of each level of education. In the old system, students were required to pass national examinations at the end of grade six, eight and twelve. In the current system, national examinations are administered at the end of grade eight, ten and twelve. The national examination at the end of grade six used to be administered exclusively in Amahric language except for language subjects. This has, very likely, contributed to the poor performance of the non-Amharic mother tongue users, and hence its abolition benefits individuals whose mother tongue language is different from Amharic more than Amharic speakers.[3]

The major part of the 1994 education reform, and one that is considered a milestone in Ethiopian education, is the introduction of mother tongue-based education in primary schools. Before 1994, the language of instruction for most subjects in primary schools was Amharic.[4] The introduction of mother tongue instruction in primary schools came as a result of the 1994 Constitution of Ethiopia. The Constitution recognized the rights of nations and nationalities to learn in their language and gave equal state recognition to the more than 80 languages spoken in the country. Given the diversity of the Ethiopian population, using Amharic as the sole language of instruction in primary education is claimed to be the root cause of educational inequality across ethno-linguistic groups(MoE, 1994; Heugh et al., 2007).

Ethiopia is divided into nine ethno-linguistic based federal regions and two administrative cities. The language(s) mandated to be used as medium of instruction in each region therefore differs depending on the region's linguistic composition. Figure 4.2 shows the regions of Ethiopia along with the languages used in their respective primary schools.

[3] Students who drop out of secondary education have now the option to join vocational education and training (VET) programs. Vocational training on basic skills lasting up to a year is provided for primary education drop outs. No major changes occurred in pre-primary education. Pre-primary education is largely left for the private sector and the government plays indirect role by providing support in preparing curriculum and training of kindergarten teachers (MoE, 2001). Enrollment in pre-primary education is quite low in Ethiopia with a gross enrollment rate of less than three percent `http://data.uis.unesco.org/`.

[4] After 1994, Amharic is taught as a second language starting from grade three or four depending on the region. English was and still remains the medium of instruction in secondary and higher education and it is taught as a subject starting from grade one.

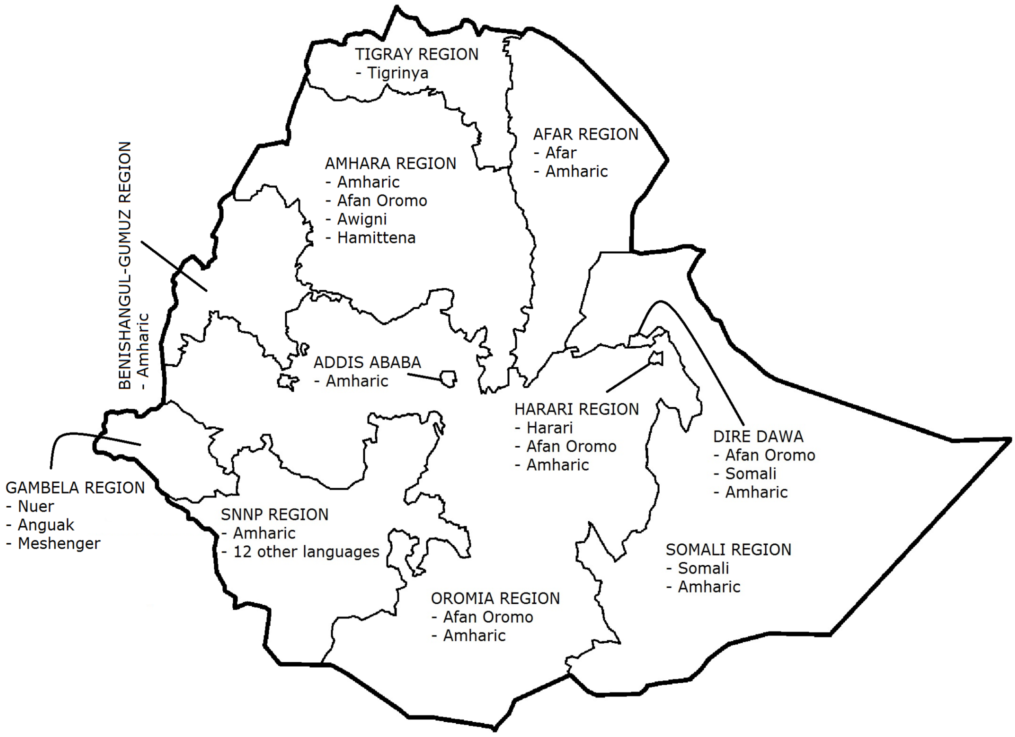


Figure 4.2: Language of instruction in primary schools across regions.

Source: Author's illustration based on Heugh et al. (2007). Map of Ethiopia with regional demarcation from commons.wikimedia.org "Ethiopia regions blank.png" accessed on 10/06/2015.

The policy reform mandated the use of mother tongue throughout primary education in all regions. However, the policy's implementation varies across regions. Some regions (e.g. Tigray and Oromia) provide mother tongue instruction from grade one to eight whereas other regions (e.g. Afar and Somali) only in the first cycle of primary education (Grade 1-4).[5] When enough speakers of a given language exist within a region, pupils are sorted into different language-specific schools. In other cases, students are sorted into different sections within schools depending on their mother tongue. The implementation of the language reform was made possible partly because the management and administration of the education system was decentralized at the regional level.

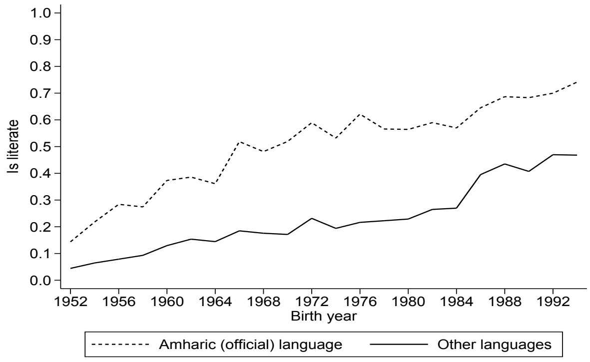
[5] For a detailed discussion on the implementation of mother tongue education and the challenges faced across regions, see Bogale (2009); Gemechu (2010); Akalu (2011); Gobana (2013).

The federal government remains in charge of setting curriculum standards and national examinations and also financing the education system (MoE, 1994).

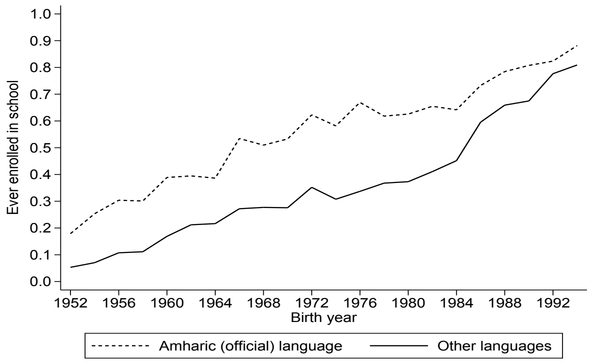
The 1994 education reform also involved changes to the curriculum content at all levels of education. In primary education, more focus is now given to science, mathematics and languages. In addition, the first cycle of primary education is provided in a self-contained unit where one teacher teaches all subjects. Textbooks are also made to reflect conditions of the specific regions. In general secondary education, new subjects such as civic education were introduced and the curriculum in preparatory secondary education is replaced with subjects which used to be provided in the first year of higher education in the old system. Furthermore, more schools were constructed especially in rural areas, fees in primary and general secondary schools were abolished whereas cost-sharing was introduced in upper secondary education and higher education (MoE, 1994, 2001). Since none of these reforms are conditional on the language spoken by individuals, they are very much likely to affect Amharic and non-Amharic mother tongue speakers equally.

Figure 4.3 shows the trend in various measures of educational outcomes across age groups. Separate trends are shown for individuals who use Amharic as their mother tongue and those whose mother tongue is different from Amharic.[6] Panel (a) shows the average reading literacy of the population across birth cohorts measured using an objective assessment of individuals' ability to read simple sentences written in one of the major languages. The share of birth cohorts that have completed at least one year of education and at least four years of education are displayed in Panel (b) and (c) respectively.

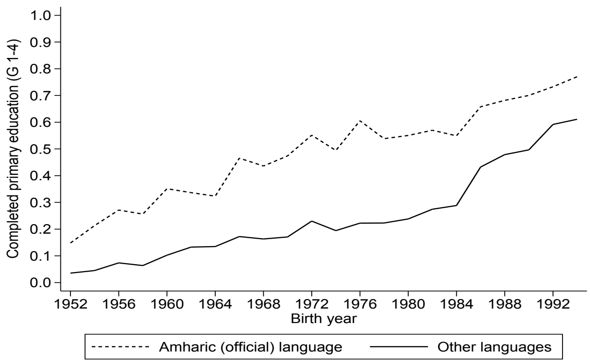
[6] Amharic mother tongue users make up about 31% of the population. Afan Oromo is the largest non-Amharic language spoken in Ethiopia with a share of 33% whereas Tigrinya and Somali language users make up 6% and 5% of the population respectively.



(a) Is literate



(b) Ever enrolled in school



(c) Completed primary education (Grade 4)

Figure 4.3: Trend in educational outcomes by birth cohorts and languages.

Source: Author's illustration based on data from the Ethiopian Demographic and Health Surveys of 2000, 2005 and 2011. The sample consists of men and women aged 15-49.

The figures show that educational outcomes in Ethiopia have improved over birth cohorts. Much of the improvements until the birth cohorts of the early 1970s occurred among individuals who use Amharic as their mother tongue. The pattern has reversed since the birth cohorts of the mid 1980s (i.e., eight years old or younger at the time of the reform) in favor of individuals who use languages other than Amharic as their mother tongue. How much the 1994 language of instruction reform contributed to the improvement in educational outcomes is the main empirical question addressed in the next sections.

4.3 Empirical Strategy

This paper uses the 1994 language reform in Ethiopia as a "natural experiment" in order to examine the effects of language of instruction on learning and labour market outcomes. The quasi-exogenous variation in exposure to the language policy change comes from differences among individuals in their year of birth and mother tongue. The first factor that determines individuals' exposure to the policy change is their year of birth. Individuals who are of primary or pre-primary school age in 1994 are potentially affected by the reform. Individuals who are above primary school age are not affected by the reform since they are, very likely, outside primary education at the time of the reform. The school entry age in Ethiopia is seven and the primary school leaving age is fourteen. Following this official school age classification, we define the treated birth cohort as those aged 0 to 14 in 1994 whereas those aged 15 to 25 in 1994 form the control birth cohort.

Late school enrollment and grade repetition rates are high in Ethiopia, as in most other Sub-Saharan African countries. For instance, the average age at enrollment ranges between seven and eleven (Nega, 2012). This makes it difficult to precisely determine the treatment status of individuals in the absence of data on age at school entry and grade repetition. Those aged 8-14 might have attended some or all grades of primary education in their mother tongue depending on their age at school entry and grade progression.

In the same way, individuals aged 15-19 at the time of the reform could have partial or no exposure to the reform. In a further specification, we use a more disaggregated age group in order to check the sensitivity of the results to the potential bias that may occur due to late school enrollment and grade repetition.

The second factor that determines individuals' exposure to the policy change is their mother tongue. Individuals whose mother-tongue is a language other than Amharic gained access to primary education in their mother tongue only after 1994. Individuals whose mother-tongue is Amharic, however, could access primary education in Amharic even before 1994. Therefore, the effects of the language change on reading skills and labour market outcomes are obtained by comparing the outcomes over birth cohorts and among individuals whose mother-tongue is Amharic and a language different from Amharic.[7].

Formally, the difference-in-differences estimate can be obtained from the following equation:

$$Y_{ilkr} = \beta_0 + \beta_1 TL_{lr} + \beta_2 TC_k + \beta_3 TL_{lr} * TC_k + \alpha_j + \delta_r + \gamma X_i + \epsilon_{ilkr} \tag{4.1}$$

where Y_{ilkr} is a measure of reading skills and labour market outcomes of individual i with mother tongue l born in year k and living in region r. TL_{lr} is a binary variable that takes the value one if the individual's mother-tongue is a language other than Amharic and is introduced as medium of instruction in primary education in 1994 in his/her region, and zero if the individual's mother-tongue is Amharic and is one of the medium of instruction in primary education in the region. TC_j is a binary variable that takes one if the individual belongs to the treated birth cohort (i.e., aged 0-14 in 1994) and zero if the individual belongs to the control birth cohort (i.e., aged 15-25 in 1994). α_k and δ_r control for cohort and region specific educational and economic developments. X_i

[7] Even though more than 80 languages are spoken in Ethiopia, the languages that are considered in this study, other than Amharic, are Afan Oromo and Tigrinya. This is because, as discussed in Section 4.2, the language policy change was implemented throughout primary education for these languages. In addition, the language of respondents in the Ethiopian Demographic and Health Survey can only be identified for these languages. This point is discussed in detail when describing the data in Section 4.4

is a vector of individual covariates including gender, an indicator for residing in a rural area as opposed to an urban area, an indicator for being a non-Christian where non-Christian religion includes Islam and traditional religions. Standard errors are clustered at the region-birth year level.

β_3 is the coefficient of interest. It is the difference-in-differences estimate which gives the effects of mother tongue instruction on reading skills and labour market outcomes. Individuals' actual exposure to the policy change, i.e, whether or not they received primary education in mother tongue, is not directly observed in the data used for the empirical analysis. Hence, the estimate from the difference-in-differences approach gives the intent to treat (ITT) effect of the reform. The empirical analysis focuses on comparing the reading skills and labour market outcomes of individuals whose mother tongue is Amharic, Afan Oromo or Tigrinya. Since the language policy reform was implemented throughout primary education for these languages, β_3 represents the effects of accessing mother tongue education for eight years in primary schools. When estimating equation 4.1, the estimation sample is restricted to individuals who attended formal education for atleast one year.

4.4 The Data

The empirical analysis relies on two data sources. The first data source is the Ethiopian Demographic and Heath Survey (EDHS) which is a national representative repeated cross sectional survey of around 20,000 households and their members.[8] Among others, it provides data on readings skills, employment, health, fertility and various socio-economic characteristics of men and women aged 15-49. The second data source is the 2010 Early Grade Reading Assessment (EGRA), which is a regionally representative survey conducted to assess the reading ability of pupils in primary education. It contains a wealth of information on pupils' reading achievement in grade two and three and various measures of school and family background characteristics including the language used in

[8] The data can be accessed at `http://dhsprogram.com/`.

school and at home.[9] The effects of language of instruction on reading skills and labour market outcomes are estimated using the EDHS undertaken in 2011. The EGRA, on the other hand, is used to analyze the mechanisms linking mother tongue-based education and reading skills.

The main advantage of the EDHS over other data sources, such as the Population and Housing Census and Labour Force Surveys, is that it provides an objective assessment of reading skills. As to the author's knowledge, EDHS is the only data source which provides an objective assessment of the reading skills of adults in Ethiopia. EDHS respondents, whose highest educational attainment is primary education, are asked to read simple sentences written on a card in one of the five major languages used in Ethiopia. These languages are Amharic, Afan Oromo, Tigrinya, Afar and Somali. The available information does not allow us to precisely determine whether reading skills are assessed in Amharic or mother tongue for individuals whose mother tongue is different from Amharic. Those who went to primary schools before 1994 are likely to read the cards in Amharic irrespective of their mother tongue since the language of instruction during primary education was Amharic. On the other hand, individuals who attended primary school after the reform are likely to read the cards in their mother tongue. The implication is that the difference-in-differences estimates from equation 4.1 are interpreted as the influence of instruction in mother tongue on general reading skills. The reading assessment results are available in three categories indicating whether the individual can read the whole sentence written on the card or he/she can read part of the sentence or he/she cannot read at all. For ease of interpreting the estimation results, the reading skill outcome is binary coded taking the value one if the individual is able to read the whole sentence and zero if the individual can read part of the sentence or none at all.

Early labour market outcomes are measured using information on employment status, occupation and earnings type. An individual is considered employed if he/she has worked during the last seven days preceding the interview. Individuals who have a job but were

[9] For detail information on EGRA, see Piper (2010).

absent from work for temporary reasons such as illness, maternity leave, etc. are also considered employed. The second variable categorizes the occupation of the respondents based on the occupation's presumed skill content. The variable takes the value (1) if respondents hold unskilled jobs (e.g. laborers in mining and manufacturing, subsistence agriculture workers and sales and service workers in elementary occupations), (2) if respondents hold skilled agriculture/manual jobs (e.g. handicraft, market-oriented skilled agricultural jobs), (3) if respondents hold skilled sales jobs (e.g. models sales and demonstrators) and (4) if respondents hold professional jobs (e.g. professional, managerial, technical and clerical jobs).[10] Finally, we use the information on the type of earning an individual receives for his/her work as an indicator for employment in the formal (non-traditional) sector. The variable takes one if an individual receives cash payments and zero if the individual is paid in cash and in kind, in kind only or not paid at all.

Mother tongue is one of the vital information required to identify individuals' exposure to the language policy change. Respondents' mother tongue is inferred from the language used during the interview process. Respondents were asked their language at the beginning of the interview so that the interviewer can decide which questionnaire to use and whether or not a translator is required for the interview. Since the EDHS questionnaires were available in three languages, namely Amharic, Afan Oromo and Tigrinya, the language of the respondents is available for only these three languages. The rest of the respondents are put in the category "Other languages". Descriptive analysis from the 2007 Census show that about 70% of the Ethiopian population uses one of these three languages as mother tongue. Individuals whose mother tongue is neither Amharic, Afan Oromo nor Tigrinya are dropped from the analysis.

Information on region of residence during primary education is required in order to precisely determine whether individuals can access primary education in their mother tongue. For instance, a Tigrinya language speaker has access to primary education in

[10] DHS also provides a similar classification of occupations.

his/her mother tongue only in Tigray region. As Figure 4.2 shows, most of regions use Amharic and one or more local languages for medium of instruction in primary education. The 2011 EDHS provides information on the current region of residence and not on the region of residence at the time of primary education. The 2007 Census shows that only 17% of the population lives in a region different from the region of birth. Given the low regional mobility in Ethiopia, using the current region of residence to determine treatment status could be feasible. Nevertheless, we use the information on the duration in the current place of residence available in the 2005 wave of EDHS to check the sensitivity of the results.

In addition to estimating the effects of language of instruction on reading skills and labour market outcomes, the empirical analysis aims to identify some of the channels at work. The first potential channel linking mother tongue-based education and an improvement in outcomes is an increase in school enrollment. Providing education in mother tongue could make school more accessible, for instance, by increasing the willingness of parents to send their children to schools. To check whether school enrollment has increased among birth cohorts who gained access to mother tongue education after 1994, we define a binary outcome variable that takes one if an individual has spent at least a year in school and zero if the individual never went to school.

An increase in parental educational investment at home is the second potential channel that links mother tongue-based education and improvement in reading skills. When the language used at home and in school is the same and textbooks are provided in languages that the parents could also understand better, they could be more involved in their child's education. In the 2010 EGRA, pupils in grade two and three are asked whether or not and from whom they receive help with homework. Based on this information, we measure parental educational investment from a time investment perspective (whether or not the parents help with homework) and a financial perspective (whether or not a tutor is hired to help with homework). Descriptive statistics are shown on Appendix Table A4.1.

4.5 Results and Discussions

4.5.1 Mother Tongue Education and Reading Skills

Table 4.1 shows estimation results for the effects of mother tongue-based education on reading skills. Simple correlations are shown in column (1) whereas basic background controls and fixed effects for region of residence and year of birth are included in column (2). Separate estimates by gender are shown in column (3) and (4).

Table 4.1: Effects of mother tongue-based primary education on literacy skills

			Female	Male
	(1)	(2)	(3)	(4)
1 if Treated language	-0.333***	-0.215***	-0.300***	-0.166***
	(0.031)	(0.026)	(0.036)	(0.032)
1 if Aged 0-14 in 1994	-0.009	0.055	0.268***	-0.070**
	(0.020)	(0.041)	(0.083)	(0.034)
Difference-in-Differences	0.131***	0.112***	0.103***	0.155***
	(0.036)	(0.024)	(0.031)	(0.030)
1 if Female		-0.124***		
		(0.011)		
1 if Rural residence		-0.225***	-0.255***	-0.210***
		(0.012)	(0.018)	(0.016)
1 if Not a Christian		-0.067***	-0.106***	-0.040***
		(0.011)	(0.018)	(0.013)
Constant	0.856***	1.049***	0.875***	1.029***
	(0.017)	(0.041)	(0.085)	(0.038)
Region fixed effects		√	√	√
Birth cohort fixed effects		√	√	√
R-squared	0.075	0.162	0.202	0.147
Sample	9,028	9,028	4,449	4,579

Note: Estimation: OLS. Dependent variable: A binary indicator for the ability to read. The treated languages are Afan Oromo and Tigrinya and control language is Amharic. The interaction between "Treated Language" and "Aged 0-14 in 1994" gives the difference-in-difference estimate. Standard errors, clustered at the region-birth year level, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011.

Individual whose mother tongue is different from Amharic have lower reading skills compared to Amharic mother tongue users. The gap is twice as large among females than males. The difference-in-differences estimate that gives the effect of the reform on reading skills is shown by the interaction term between being aged 0-14 in 1994 and having a mother tongue other than Amharic. Accessing education in mother tongue during primary education increases the probability of being able to read by 11 percentage points.[11] As the result of the reform, the literacy gap between Amharic speakers and individuals whose mother tongue is different from Amharic halved from 22 to 11 percentage points[12]. This implies that the language of learning and teaching in Ethiopia's primary schools is one of the key determinants of reading ability and providing primary education in mother tongue reduces reading skills gap across language groups.[13]

Providing education in pupils' mother tongue is hypothesized to improve the learning outcomes of girls more than boys. Girls in most developing countries are involved in housework compared to boys who spend more time outside the home environment. This lowers girls exposure to the dominant language spoken in the country. The results in column (3) and (4) of Table 4.1 show that the penalty on reading skills of having a mother tongue other than Amharic is 13 percentage points higher for females than males. Due to the huge push made to reduce gender inequality in educational outcomes, the reading skills of females in general has improved over birth cohorts more than males. However, females have not benefited from the language reform more than males in terms gains in reading skills. The DID estimates seem larger for males than females but the difference across gender is not statistically significant. When comparing the reading skills gap among language groups between males and females, the estimates show that the reform almost eliminated the reading skills gap between Amharic and non-Amharic speaking

[11] We get similar results when we use logistic regression to take into account the binary nature of the outcome variables. The estimation results from the logistic regression are shown in Appendix Table A4.2.

[12] In this analysis, individuals whose mother tongue is different from Amharic are those individuals whose mother tongue is either Oromiffa or Tigrigna. These group comprises about 65 percent of the Ethiopian population whose mother tongue is different from Amharic.

[13] When we include the interaction between birth year and region in the regression equation to control for differential cohort-specific trends in educational expansion across regions, the DID estimate becomes 0.144 (standard error: 0.035). The household level clustered standard errors lowers to 0.022. The standard error of the DID estimate increases to 0.030 when clustering both at region and birth year level to account for within-region and within-year serial correlation (Bertrand et al., 2004; Cameron and Miller, 2015).

males from 16.6 percentage points to 1.1 (=-0.166 + 0.155) percentage points. The reading skills gap between Amharic and non-Amharic speaking females reduced only by one third from 30 to 20 percentage points as a result of the reform.

Figure 4.4 shows estimation results disaggregated by age groups. For the sake of clarity, only the difference-in-differences estimates are reported.

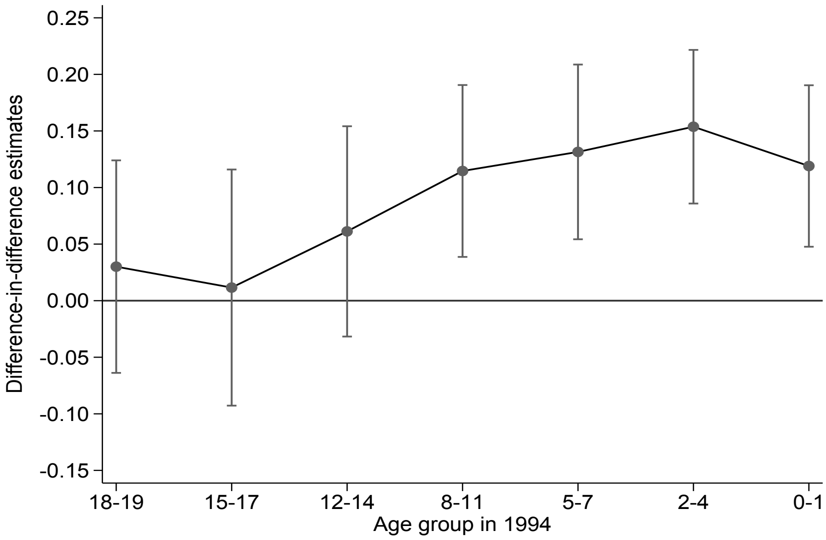


Figure 4.4: Effect of mother tongue instruction on reading skills across age groups.

Source: Author's illustration based on EDHS 2011

As expected, the difference-in-differences estimates increase when individuals' age at the time of the reform declines. This implies that the effect of the reform on reading skills is larger among individuals who potentially have full exposure to mother tongue-based primary education.

4.5.2 Robustness and Sensitivity Checks

Table 4.2 show results for various robustness checks. The first robustness check deals with similarity across languages. 65% of individuals in the treated language group are Afan Oromo mother tongue speakers. The Afan Oromo language differs substantially from Amharic. Afan Oromo language belongs to the Cushitic language family and uses the latin script. Amharic, on the other hand, belongs to the Semitic language family and uses an ancient writing system called Ge'ez (Ethiopic). The rest of the individuals in the treated language group use the Tigrinya language, a language that belongs to the same family and uses the same script as Amharic. In order to check whether or not Afan Oromo speakers benefited more than Tigrinya speakers, the treated language is restricted to Afan Oromo speakers in Column (1) and Tigrinya speakers in column (2) of Table 4.2. As expected, the substantial share of improvement in reading skills occurred among Afan Oromo speakers.[14]

Table 4.2: Robustness check using various specifications

				Background characteristics		
	Afan Oromo	Tigrinya	Literacy Program	Ethnicity	HH Size & wealth	HH FE
	(1)	(2)	(3)	(4)	(5)	(6)
1 if Treated Language	-0.241***	0.085*	-0.212***	-0.219***	-0.211***	-0.238***
	(0.027)	(0.043)	(0.026)	(0.027)	(0.026)	(0.051)
1 if Aged 0-14 in 1994	0.036	0.029	0.057	0.054	0.056	-0.083
	(0.044)	(0.094)	(0.041)	(0.041)	(0.042)	(0.087)
Difference-in-Differences	0.131***	0.064*	0.110***	0.113***	0.116***	0.159***
	(0.027)	(0.036)	(0.024)	(0.024)	(0.025)	(0.040)
1 if Female	-0.130***	-0.076***	-0.124***	-0.113***	-0.126***	-0.137***
	(0.012)	(0.011)	(0.011)	(0.024)	(0.012)	(0.016)
1 if Rural residence	-0.218***	-0.192***	-0.222***	-0.224***	-0.070***	-0.220***
	(0.014)	(0.015)	(0.013)	(0.012)	(0.017)	(0.032)
1 if Not a Christian	-0.070***	-0.013	-0.062***	-0.070***	-0.071***	-0.101***
	(0.011)	(0.012)	(0.011)	(0.011)	(0.010)	(0.028)
R-squared	0.181	0.089	0.163	0.162	0.182	0.145
Sample	7,732	6,552	9,021	9,028	9,028	4,579

Note: Estimation: OLS. Dependent variable: A binary indicator for the ability to read. The interaction between "Treated Language" and "Aged 0-14 in 1994" gives the difference-in-differences estimate. All regressions include controls for region of residence and birth year fixed effects. Additional control variables included are an indicator for attending literacy program outside formal education in Column (3), ethnicity in Column (4), household wealth and size in Column (5) and household fixed effects in Column (6). Standard errors, clustered at the region-birth year level, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011.

[14] The differences across the two language groups are statistically different from each other at 13 significance level with an F statistic of 2.31.

The second robustness check deals with educational programs provided outside formal education that may confound the difference-in-differences estimates. The state and non-governmental organizations including churches or mosques provide basic literacy programs outside the formal education in order to improve literacy skills in Ethiopia. About 17% of individuals in the sample have attended literacy program outside of the formal primary education. Any similar change of the language used for instruction in such literacy programs could confound the difference-in-differences estimates. To account for this, in column (3), we included a binary indicator for attending a literacy program outside of formal primary education and the difference-in-differences estimate remain unchanged.[15]

Family background characteristics such as parental educational and occupational status could influence both schooling choices as well as educational outcomes. Regression results in column (4)-(6) take into account differences in socio-economic background characteristics using various approaches. In column (4), we control for ethnicity. In column (5), we include controls for household wealth index, which is constructed based on the household's ownership of assets and access to sanitation and other basic facilities, and household size whereas column (6) exploits the within household variation by including a household fixed effect. The end results show that systematic differences in socio-economic background characteristics barely changes the difference-in-differences estimate.

The main identifying assumption required for a difference-in-differences approach to give a consistent estimate is the parallel trend assumption. For the parallel trend assumption to be fulfilled, reading skills across the treated and the control language groups should follow a similar trend across birth cohorts in the absence of the language policy change.

[15] Reading skills are negatively correlated with attending literacy programs outside formal education (-0.036 with standard error of 0.014). This is far from being a causal relation because of negative selection bias (i.e., individuals who attended literacy programs likely have lower unobeservable attributes than individuals who never attended literacy programs). Restricting the sample to individuals who never attended literacy program outside of formal education gives a similar result. The estimated effect of the reform is 0.123 with standard error of 0.028.

The parallel trend assumption cannot be tested directly. One way to check whether or not the parallel trend assumption is fulfilled is to see if there are any systematic differences in reading skills across language groups in the pre-reform period. This can be done by creating a placebo or psuedo treatment. Any systematic reading skills differences across the language groups in the pre-reform period would question the parallel trend assumption.

As a placebo treatment, individuals aged 26-35 in 1994 are assumed to gain access to mother-tongue based primary education. The reading skills of the placebo-treated birth cohort is compared with individuals aged 15-25 in 1994. The difference-in-differences estimation results are shown in Column (1) of Table 4.3. The estimate on the interaction term indicates that there is no systematic difference in reading skills across language groups in the pre-reform period.

Table 4.3: Accounting for possible threats to identification

	Placebo treatment	Schooled for less than seven years	Regional mobility
	(1)	(2)	(3)
1 if Treated language	-0.196***	-0.145***	-0.175***
	(0.031)	(0.047)	(0.035)
1 if Treated birth cohort	0.017	0.073	0.015
	(0.059)	(0.078)	(0.062)
Difference-in-Differences	0.016	0.085**	0.090***
	(0.037)	(0.042)	(0.031)
1 if Female	-0.171***	-0.178***	-0.118***
	(0.018)	(0.020)	(0.015)
1 if Rural residence	-0.287***	-0.097***	-0.257***
	(0.025)	(0.025)	(0.021)
1 if Not a Christian	-0.098***	-0.004	-0.063***
	(0.022)	(0.019)	(0.020)
Constant	1.112***	0.809***	1.093***
	(0.049)	(0.077)	(0.057)
R-squared	0.263	0.108	0.210
Sample	2,746	3,833	3,454

Note: Estimation: OLS. Dependent variable: A binary indicator for the ability to read. The treated languages are Afan Oromo and Tigrinya and control language is Amharic. The interaction between "Treated Language" and "Aged 0-14 in 1994" gives the difference-in-difference estimate. Estimation sample restricted to individuals who were schooled for maximum of six years in Column (2) and individuals who lived in the same area since birth in Column (3). Standard errors, clustered at the region-birth year level, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011 in column (1) and (2) and Ethiopian Demographic and Health Survey (EDHS) 2005 in column (3).

The 1994 education reform in Ethiopia is a package of reforms as discussed in Section 4.2. Any other policy change that influences the educational outcomes of individuals in the treated and the control language group differently would confound the difference-in-differences estimates. One such reform is the elimination of the central school exit examination that students were required to pass at the end of grade six. The central examination abolition increases the educational attainment (and possibly reading skills) of everyone irrespective of mother tongue. Since the examination used to be administered in Amharic, its abolition automatically increases the educational attainment of non-Amharic mother tongue users more than their Amharic counterparts. This makes it difficult to disentangle the effects of the two reforms on years of completed education.[16]

To identify the sole effect of the language of instruction reform on reading skills, we restrict the estimation sample to individuals whose educational attainment is not affected by the national examination abolition. These are individuals whose maximum years of education is six years and they make up about 55 percent of the total estimation sample. Results are displayed in column (2) of Table 4.3. The difference-in-differences estimate reduces in size but does not become insignificant both economically and statistically. The estimate is lower partly because controlling for educational attainment partials out the effect of the language reform on reading skills.

Another factor that could threaten identification is non-random regional mobility in response to the policy change. Non-random regional mobility could occur if, for instance, parents who value education move to a different region so that their children could access education in mother tongue while at the same time putting extra investment in their children's education. Ideally, one would need data on the region that the individuals undertook primary education. Due to lack of data in the EDHS 2011, the current region of residence is assumed to be the same region that individuals finished primary education. This is not a strong assumption given that about 70% of individuals live in the same area

[16] This issue is not discussed in Ramachandran (2012). In fact, the 12 percent increase in percentage of individuals completing 6 years or more of schooling that the author finds could be partly due to the central examination abolition.

where they were born according to data from the previous wave of the EDHS. We use the regional information available in the EDHS 2005 to check the sensitivity of the results. Column (3) of Table 4.3 show results using EDHS 2005 after restricting the estimation sample to individuals who have been living in the same region since birth. The DID estimate show that the reading skills of those affected by the reform has increased by about 9 percentage points which is very close to the 11 percentage points found in the main specification implying that regional mobility does not alter the main finding.

4.5.3 Channels of Influence

Improving the accessibility of schools could be one of the mechanisms through which mother tongue instruction improves reading skills. We estimate equation 4.1 using school enrollment as the outcome variable to check whether there is a positive relationship between access to mother tongue-based education and school enrollment. School enrollment is a binary variable which is one if individuals ever attended formal education and zero otherwise. Table 4.4 shows the results.

Table 4.4: Mother tongue education and school enrollment

			Female	Male
	(1)	(2)	(3)	(4)
1 if Treated language	-0.198***	-0.164***	-0.283***	-0.032
	(0.049)	(0.022)	(0.028)	(0.031)
1 if Aged 0-14 in 1994	0.134**	0.324***	0.428***	0.208***
	(0.057)	(0.047)	(0.060)	(0.060)
Difference-in-Differences	0.100***	0.094***	0.149***	0.052*
	(0.061)	(0.022)	(0.028)	(0.029)
1 if Female		-0.201***		
		(0.011)		
1 if Rural residence		-0.312***	-0.326***	-0.293***
		(0.012)	(0.015)	(0.016)
1 if Not a Christian		-0.071***	-0.101***	-0.033**
		(0.011)	(0.014)	(0.014)
Constant	0.599***	0.950***	0.798***	0.870***
	(0.049)	(0.044)	(0.052)	(0.059)
Region and birth cohort fixed effects		√	√	√
R-squared	0.050	0.313	0.355	0.223
Sample	14,206	14,206	7,976	6,230

Note: Estimation: OLS. Dependent variable: A binary indicator for ever been enrolled in school. The treated languages are Afan Oromo and Tigrinya and control language is Amharic. The interaction between "Treated Language" and "Aged 0-14 in 1994" gives the difference-in-difference estimate. Standard errors, clustered at the region-birth year level, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011.

School enrollment is on average lower among non-Amharic mother tongue users compared to Amharic mother tongue users. The negative correlation between non-Amharic mother tongue users and school enrollment is observed only among females. The enrollment rate of the young birth cohorts (those aged 0-14 in 1994) has improved substantially compare to the old birth cohorts. The improvement in enrollment among females is twice as large as for males. The difference-in-differences estimates show that school enrollment among birth cohorts that gained access to mother tongue-based education after 1994 increased by about 10 percentage points. In terms of school enrollment, the language reform was thus of greatest benefit for non-Amharic speaking women.

A further positive side effect of the reform could be that parents enhance their investment at home when the language used at home and in school is the same and textbooks are provided in languages that parents could understand better. The EGRA allows measuring parental educational investment at home with respect to helping with homework from two perspectives: time investment and financial investment. We test for positive correlation between the language of instruction in schools and parental time and financial investment at home.

The estimation results are shown in Table 4.5. The dependent variable in columns (1) & (2) is one if the child receives help with homework from somebody in the household and zero if nobody helps the child with homework. The dependent variable in columns (3) & (4) is one if the child's father helps with homework and zero if otherwise. The dependent variable in columns (5) & (6) takes the value one if a tutor is hired to help the child with homework and zero if otherwise. The main explanatory variable of interest is whether a child goes to a school where instruction is provided in his/her mother tongue. Simple correlations are shown in the first column for each outcome variable whereas child level, family background characteristics and school fixed effects are included in the rest of the columns.

Table 4.5: Mother tongue education and parental educational investment at home

	Who helps with home work?					
	Somebody		Father		Tutor	
	(1)	(2)	(3)	(4)	(5)	(6)
1 if Child is in MT school	0.131***	0.045**	0.043***	0.040**	0.012**	0.015***
	(0.036)	(0.022)	(0.015)	(0.017)	(0.005)	(0.004)
Child's age		-0.019***		-0.007***		-0.000
		(0.003)		(0.002)		(0.001)
1 if Child is female		-0.016		-0.005		-0.007*
		(0.010)		(0.008)		(0.004)
1 if Child is in grade 3		0.020*		0.025***		-0.005
		(0.011)		(0.008)		(0.006)
1 if Child attended KG		0.023*		0.017		-0.006
		(0.014)		(0.012)		(0.005)
1 if Father is literate		0.076***		0.110***		-0.014**
		(0.015)		(0.013)		(0.006)
1 if Mother is literate		0.047***		0.007		-0.005
		(0.014)		(0.011)		(0.004)
Socio-economic status score		0.067***		0.019**		0.010**
		(0.008)		(0.009)		(0.004)
Constant	0.585***	0.739***	0.096***	0.043	0.006**	0.031
	(0.038)	(0.048)	(0.014)	(0.030)	(0.003)	(0.020)
School fixed effects		√		√		√
Mean [sd] of dependent variable	0.700 [0.458]		0.134 [0.341]		0.016 [0.127]	
R-squared	0.009	0.037	0.002	0.031	0.001	0.007
Sample	7,915	7,913	7,915	7,913	7,915	7,913

Note: Estimation: OLS. Standard errors, clustered at the school level, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels. MT stands for mother tongue.
Source: Early Grade Reading Assessment (EGRA) 2010.

The results show that parental educational investment at home is positively associated with the use of mother tongue in primary schools. Parents increase their educational investment at home either directly by providing support to their child with homework or indirectly by hiring a tutor to provide help with homework. The probability that a child receives help with homework increases by about 1.5 to 4.5 percentage points if the child uses the same language at home and in school.

4.5.4 Impacts on Early Labour Market Outcomes

Human capital is one of the main determinants of labour market outcomes. The improved reading skills that resulted from the introduction of mother tongue-based primary education could ultimately lead to better labour market outcomes. This section examines the labour market returns to mother tongue-based education by estimating the reduced-form effects of the reform on early labour market outcomes.[17] Labour market outcomes are measured using three variables: employment status, occupation and earning type. The variable indicating employment status takes the value one if an individual is employed and zero if otherwise. An individual is considered employed if he/she has worked during the last seven days preceding the interview. Individuals who have a job but were absent from work for temporary reasons such as illness, maternity leave, etc. are also considered employed.

The second variable categorizes the occupation of the respondents based on the occupation's presumed skill content. The variable takes the value (1) if respondents hold unskilled jobs (e.g. laborers in mining and manufacturing, subsistence agriculture workers and sales and service workers in elementary occupations), (2) if respondents hold skilled agriculture/manual jobs (e.g. handicraft, market-oriented skilled agricultural jobs), (3) if respondents hold skilled sales jobs (e.g. models sales and demonstrators) and (4) if respondents hold professional jobs (e.g. professional, managerial, technical and clerical jobs). In the base category are individuals who hold unskilled manual/agriculture/service jobs. The third labour market outcome is based on the information on the type of earning an individual receives for his/her work which indicates for employment in the formal (non-traditional) sector. The variable takes one if an individual receives cash payments and zero if the individual is paid in cash and in kind, in kind only or not paid at all. Results are shown in Table 4.6.

[17] It is called "early" because the birth cohort that gained access to primary education in mother tongue is between the age of 16 and 30 at the time where labour market outcome measures are collected. The average age of the treated birth cohort is about 22 and the average individual in the control birth cohort is about 36 years old.

Table 4.6: Labour market outcomes: Employment status, occupation and earnings

	Employed		Skilled worker - Manual/Agriculture		Skilled worker - Sales		Skilled worker - Professional/technical/ managerial		Receives earnings in cash	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1 if Treated language	-0.074***	-0.082***	0.358***	0.369***	0.565***	0.622***	0.499***	0.563**	-0.142***	-0.128***
	(0.023)	(0.023)	(0.060)	(0.063)	(0.119)	(0.135)	(0.111)	(0.152)	(0.026)	(0.025)
1 if Aged 0-14 in 1994	-0.303***	-0.304***	0.329***	0.337***	0.315***	0.326***	0.090***	0.252**	-0.130***	-0.112***
	(0.056)	(0.056)	(0.106)	(0.115)	(0.082)	(0.088)	(0.053)	(0.173)	(0.046)	(0.043)
Difference-in-Differences	0.034	0.041*	1.687***	1.651***	1.732***	1.642**	1.418	1.318	0.072***	0.065***
	(0.022)	(0.022)	(0.268)	(0.263)	(0.334)	(0.324)	(0.305)	(0.327)	(0.022)	(0.021)
1 if Female	-0.288***	-0.289***	0.588***	0.616***	3.994***	4.178***	1.619***	2.459***	0.034**	0.040***
	(0.010)	(0.010)	(0.073)	(0.075)	(0.406)	(0.421)	(0.156)	(0.286)	(0.014)	(0.014)
1 if Rural residence	0.094***	0.081***	0.208***	0.240***	0.288***	0.328***	0145***	0.446***	-0.444***	-0.382***
	(0.015)	(0.015)	(0.023)	(0.026)	(0.036)	(0.044)	(0.020)	(0.066)	(0.017)	(0.017)
1 if Not a Christian	-0.044***	-0.048***	1.168	1.246**	2.001***	2.121***	0.785**	1.317***	-0.019	0.002
	(0.012)	(0.012)	(0.122)	(0.132)	(0.229)	(0.247)	(0.088)	(0.198)	(0.013)	(0.014)
1 if Secondary education		-0.066***		1.480***		1.714***		9.434***		0.095***
		(0.015)		(0.138)		(0.181)		(1.505)		(0.017)
1 if Higher education		-0.005***		2.078***		1.714***		111.515***		0.237***
		(0.020)		(0.257)		(0.181)		(21.423)		(0.022)
Constant	1.127***	1.148***	1.658***	1.376	0.430***	0.323***	0.403**	0.026***	0.699***	0.613***
	(0.042)	(0.045)	(0.469)	(0.421)	(0.120)	(0.098)	(0.163)	(0.013)	(0.046)	(0.044)
Region fixed effects	√	√	√	√	√	√	√	√	√	√
Birth cohort fixed effects	√	√	√	√	√	√	√	√	√	√
Model	OLS		Multinomial logit						OLS	
Sample	9,028		6,767						6,767	

Note: Estimation: OLS in Columns (1-2) and Columns (9-10). Multinomial logit in Columns (3-8). The base categories are "not employed" in Column (1-2), "unskilled manual/agriculture/service worker" in Columns (3-8) and "receiving payment in kind/in cash and in kind/nothing" in Columns (9-10). Relative risk ratios are reported in the multinomial logit regression. The treated languages are Afan Oromo and Tigrinya and control language is Amharic. The interaction between "Treated Language" and "Aged 0-14 in 1994" gives the difference-in-difference estimate. Standard errors, clustered at the region-birth year level, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011.

The outcome measure in column (1)-(2) is whether the respondent is employed or not. Individuals whose mother tongue is different from Amharic have a lower probability to become employed compared to Amharic mother tongue users. The young birth cohorts also have a lower probability to be employed mainly because they have less work experience compared to the older birth cohorts. The interaction term shows that there is a very small positive correlation between the reform and the probability to be employed. The finding that the reform has a negligible influence on the probability of finding a job allows us to look at the quality of job's held by restricting the estimation sample to individuals who are employed, without compromising the sample composition.

Based on the sample of employed workers, Column (3)-(8) show results from a multinomial regression where occupations are categorized into four groups based on their skill content (unskilled jobs, skilled manual or agricultural jobs, skilled sales jobs and professional jobs). Results show that individuals whose mother tongue is different from Amharic and the youth are less likely to hold skilled jobs in general. However, the difference-in-differences estimate show that the skill content of jobs held by those affected by the reform has improved significantly. In addition, Column (5) show that non-Amharic mother tongue speakers affected by the reform are more likely to receive their earnings in cash as opposed to informal means of payments such as cash and in kind, in kind only or receiving any payments for their work. A robustness check on Appendix Table A4.3 shows stronger effects on the labour market outcomes of individuals whose mother tongue is Afan Oromo. This result complements the finding in Table 4.2 that the reform improved the reading skills of Afan Oromo speakers substantially.

The main arguments put against mother tongue-based education is that it could limit employment and other economic opportunities in the labour market. It does so by reducing proficiency skill in the language that dominates the labour market. The findings of this paper are against this criticism. The use of mother tongues as medium of instruction in Ethiopia's primary schools has led to improved labour market outcomes in terms of the skill contents of jobs held and the type of payment an individual receives for his/her work. The positive returns on the labour market occurred because the improved reading skills provided better employment opportunities in the formal labour market or it led to better educational attainments and general human capital development which are in turn positively rewarded in the labour market.

4.6 Conclusion

The change in the language of instruction in Ethiopia's primary schools provides a "natural experiment" to examine the relationship between language of instruction, literacy skill and labour market outcomes. The difference-in-differences estimates show that birth cohorts that gained access to mother tongue-based primary education after 1994 have significantly better reading skills as adults. The reform was thus effective in improving long-term literacy skills. The results are found to be robust to various sensitivity checks such as controlling for family background characteristics and regional mobility among others, and a placebo-treatment test.

Heterogeneous analysis reveals that the reform has not improved the reading skills of females more than males. This is in contrast to the belief that mother tongue education benefits girls more than boys since girls are involved in housework and hence have lower exposure to the dominant language spoken in the country. A substantial increase in school enrollment, especially among females, and in parental educational investment outside schools are identified as some of the channels of influence. Moreover, the improved human capital as a result of being taught in mother tongue do translate into gains in the labour market. Non-Amharic mother tongue speakers affected by the reform are more likely to work as skilled workers and to receive their earnings in cash as opposed to informal means of payments.

The results have important implication for the status of educational and economic inequality across groups in Ethiopia as well as for the potential role that mother-tongue education policy plays in reducing inequality across groups in other countries. The findings show that changing the language of learning and teaching from Amharic to mother tongue-based has reduced inequality across ethno-linguistic groups in Ethiopia. Furthermore, parents are engaged more in their child's education outside school when education is provided in mother tongue thereby potentially improving inter-generational mobility. Mother tongue-based education policy, therefore, could play a key role in reducing inequality across groups divided based on language in other countries. Language related inequality could exist as a result of colonization in African countries and migration in Western countries.

Further research is required in the following area. The increased enrollment rate among non-Amharic speaking females has not translated into an improvement in their reading skills compared to their Amharic speaking as well as their male counterparts. Further investigation is required to identify the reasons why the gap in reading skills remain high between Amharic speaking and non-Amharic speaking females. Furthermore, the consequence of providing formal education in mother tongue on regional mobility and the efficient allocation of skill in labour markets across regions is open for further research.

4.7 References

Akalu, T. (2011). Assessment of Mother Tongue Education Policy in Relation to its Implementation in Gambella Adminstrative Town: The Case of Agnuak Language. Master's thesis, Addis Ababa University.

Aldashev, A., Gernandt, J., and Thomsen, S. L. (2009). Language Usage, Participation, Employment and Earnings: Evidence for Foreigners in West Germany with Multiple Sources of Selection. *Labour Economics*, 16(3):330–341.

Angrist, J., Chin, A., and Godoy, R. (2006). Is Spanish-only Schschool Responsible for the Puerto Rican Language Gap? *Journal of Development Economics*, 85(1):105–128.

Angrist, J. D. and Lavy, V. (1997). The Effect of a Change in Language of Instruction on the Returns to Schooling in Morocco. *Journal of Labor Economics*, pages S48–S76.

Bender, P., Dutcher, N., Klaus, D., Shore, J., and Tesar, C. (2005). In Their Own Language: Education for All. Technical report, The World Bank.

Berman, E., Lang, K., and Siniver, E. (2003). Language-Skill Complementarity: Returns to Immigrant Language Acquisition. *Labour Economics*, 10(3):265–290.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics*, 119(1):249–275.

Bleakley, H. and Chin, A. (2004). Language Skills and Earnings: Evidence from Childhood Immigrants. *Review of Economics and Statistics*, 86(2):481–496.

Bogale, B. (2009). Language Determination in Ethiopia: What Medium of Instruction? In *Proceedings of the 16th International Conference of Ethiopian Studies*, pages 2–7, Trondheim.

Cameron, A. C. and Miller, D. L. (2015). A Practitioner's Guide to Cluster-Robust Inference. *Journal of Human Resources*, 50(2):317–372.

Eriksson, K. (2014). Does the Language of Instruction in Primary School Affect Later Labour Market Outcomes? Evidence from South Africa. *Economic History of Developing Regions*, 29(2):311–335.

Gemechu, D. T. (2010). *The Implementation of a Multilingual Education Policy in Ethiopia: The Case of Afaan Oromoo in Primary Schools of Oromia Regional State.* PhD thesis, University of Jyväskylä.

Gobana, J. A. (2013). *Challenges of Mother Tongue Education in Primary Schools: The Case of Afan Oromo in the East Hararge Zone, Oromia Regional State, Ethiopia.* PhD thesis, University of South Africa.

Heugh, K., Benson, C., Bogale, B., and Yohannes, M. A. G. (2007). Final Report Study on Medium of Instruction in Primary Schools in Ethiopia. Technical report, Ministry of Education.

Heugh, K., Benson, C., Yohannes, M. A. G., and Bogale, B. (2012). *Implications for Multilingual Education: Student Achievement in Different Models of Education in Ethiopia*, chapter Multilingual Education and Substainable Diversity Work from Periphery to Center. Routledge.

MoE (1994). Education and Training Policy. Technical report, Ministry of Education.

MoE (2001). The Education and Training Policy and Its Implementation. Technical report, Ministry of Education.

Nega, A. (2012). Improving Education Quality, Equity and Acess in Ethiopia: Findings from the Young Lives School Component. Technical report, Young Lives.

Piper, B. (2010). Ethiopia Early Grade Reading Assessment Data Analysis Report: Language and Early Learning. Technical report, RTI International.

Ramachandran, R. (2012). Language Use in Education and Primary Schooling Attainment: Evidence from a Natural Experiment in Ethiopia. *Documents de Treball de l'IEB 2012/34.*

Smits, J., Huisman, J., and Kruijff, K. (2008). Home Language and Education in the Developing World. Background Paper for the EFA Global Monitoring Report Overcoming Inequality: Why Governance Matters. Technical report, UNESCO.

Taylor, S. and Coetzee, M. (2013). Estimating the Impact of Language of Instruction in South African Primary Schools: A Fixed Effects Approach. *Stellenbosch Economic Working Papers 21/13*.

UNICEF (2011). Action Research on Mother Tongue-Based Bilingual Education: Achieving Quality, Equitable Education. Technical report, UNICEF.

Walter, S. and Chou, K. G. (2012). The Kom Experimental Mother Tongue Education Project Report for 2012. Technical report, SIL International.

Walter, S. L. and Dekker, D. E. (2011). Mother Tongue Instruction in Lubuagan: A Case Study from the Philippines. *International Review of Education*, 57(5-6):667–683.

4.8 Appendix

Table A4.1: Summary statistics

	Overall	Amharic language		Non-Amharic language	
		Pre-1994	Post-1994	Pre-1994	Post-1994
	(1)	(2)	(3)	(4)	(5)
1 if able to read	0.733	0.856	0.847	0.523	0.645
	(0.443)	(0.351)	(0.360)	(0.500)	(0.479)
1 if employed	0.736	0.849	0.696	0.909	0.739
	(0.441)	(0.359)	(0.460)	(0.288)	(0.440)
Occupational status					
1 if unskilled worker	0.565	0.345	0.485	0.675	0.691
	(0.496)	(0.476)	(0.500)	(0.469)	(0.462)
1 if skilled manual/agricultural worker	0.159	0.244	0.175	0.120	0.111
	(0.366)	(0.429)	(0.380)	(0.325)	(0.314)
1 if skilled sales worker	0.131	0.166	0.123	0.119	0.122
	(0.337)	(0.372)	(0.329)	(0.325)	(0.328)
1 if professional worker	0.145	0.246	0.218	0.087	0.076
	(0.352)	(0.431)	(0.413)	(0.281)	(0.265)
1 if receives earnings in cash only	0.506	0.764	0.638	0.332	0.312
	(0.500)	(0.425)	(0.481)	(0.471)	(0.464)
1 if female	0.492	0.481	0.545	0.316	0.474
	(0.470)	(0.500)	(0.498)	(0.465)	(0.499)
1 if rural resident	0.470	0.206	0.303	0.727	0.738
	(0.499)	(0.404)	(0.459)	(0.446)	(0.440)
1 if not a Christian	0.301	0.184	0.210	0.457	0.432
	(0.459)	(0.387)	(0.407)	(0.499)	(0.495)
Household Wealth Index					
1 if poorest quantile	0.072	0.033	0.050	0.109	0.109
	(0.259)	(0.179)	(0.218)	(0.311)	(0.312)
1 if poorer quantile	0.097	0.044	0.064	0.136	0.154
	(0.103)	(0.205)	(0.245)	(0.343)	(0.361)
1 if middle quantile	0.103	0.036	0.071	0.155	0.158
	(0.304)	(0.187)	(0.258)	(0.362)	(0.365)
1 if richer quantile	0.151	0.066	0.089	0.255	0.244
	(0.359)	(0.249)	(0.284)	(0.436)	(0.430)
1 if richest quantile	0.576	0.820	0.726	0.345	0.335
	(0.494)	(0.384)	(0.446)	(0.476)	(0.472)
Household size	4.887	4.430	4.631	5.645	5.220
	(2.150)	(2.019)	(2.169)	(1.853)	(2.150)
Observations	9,028	1,236	4,020	792	2,980
Employed sample	6,767	1,049	2,797	720	2,221

Note: Standard deviation are in parentheses.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011.

Table A4.2: Results from a logistic regresssion

			Female	Male
	(1)	(2)	(3)	(4)
1 if Treated language	-0.291***	-0.166***	-0.223***	-0.127***
	(0.028)	(0.023)	(0.030)	(0.031)
1 if Aged 0-14 in 1994	-0.012	0.056	0.248***	-0.076*
	(0.027)	(0.040)	(0.064)	(0.040)
Difference-in-Differences	0.099***	0.082***	0.069**	0.119***
	(0.034)	(0.023)	(0.029)	(0.031)
1 if Female		-0.126***		
		(0.012)		
1 if Rural residence		-0.204***	-0.220***	-0.208***
		(0.012)	(0.016)	(0.018)
1 if Not a Christian		-0.062***	-0.093***	-0.039***
		(0.009)	(0.014)	(0.012)
Constant	1.782***	2.931***	1.720***	3.062***
	(0.141)	(0.260)	(0.436)	(0.288)
Region fixed effects		√	√	√
Birth cohort fixed effects		√	√	√
Pseudo R-squared	0.065	0.151	0.174	0.152
Sample	9,028	9,028	4,449	4,579

Note: Estimation: Logit model. Marginal effects are reported. Dependent variable: A binary indicator for the ability to read. The treated languages are Afan Oromo and Tigrinya and control language is Amharic. The interaction between "Treated Language" and "Aged 0-14 in 1994" gives the difference-in-difference estimate. Standard errors, clustered at the region-birth year level, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels. *Source*: The Ethiopian Demographic and Health Survey (EDHS) 2011.

Table A4.3: Labour market outcomes: Afan Oromo as the treated language

	Employed		Skilled worker - Manual/Agriculture		Skilled worker - Sales		Skilled worker - Professional/technical/ managerial		Receives earnings in cash	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1 if Treated language	-0.075***	-0.082***	0.307***	0.308***	0.593**	0.634*	0.540**	0.574	-0.111***	-0.104***
	(0.027)	(0.028)	(0.059)	(0.060)	(0.149)	(0.167)	(0.152)	(0.207)	(0.030)	(0.025)
1 if Aged 0-14 in 1994	-0.285***	-0.287***	0.182***	0.177***	0.332***	0.331***	0.066***	0.167**	-0.104**	-0.095*
	(0.068)	(0.069)	(0.062)	(0.061)	(0.125)	(0.127)	(0.052)	(0.150)	(0.049)	(0.049)
Difference-in-Differences	0.027	0.031	1.915***	1.952***	1.881**	1.853**	1.449	1.465	0.059**	0.065**
	(0.029)	(0.029)	(0.387)	(0.389)	(0.500)	(0.499)	(0.427)	(0.543)	(0.026)	(0.025)
1 if Female	-0.319***	-0.322***	0.791	0.795	5.600***	5.663***	1.855***	2.910***	0.097**	0.102***
	(0.013)	(0.014)	(0.129)	(0.129)	(0.703)	(0.712)	(0.266)	(0.473)	(0.017)	(0.016)
1 if Rural residence	0.059***	0.039	0.231***	0.249***	0.322***	0.340***	0.123***	0.432***	-0.479***	-0.422***
	(0.019)	(0.019)	(0.033)	(0.037)	(0.058)	(0.065)	(0.019)	(0.083)	(0.022)	(0.022)
1 if Not a Christian	-0.026*	-0.033***	1.273**	1.327**	1.774***	1.830***	0.764*	1.244	-0.010	0.006
	(0.015)	(0.015)	(0.154)	(0.162)	(0.252)	(0.269)	(0.108)	(0.252)	(0.016)	(0.016)
1 if Secondary education		-0.069***		1.249*		1.402**		11.347***		0.061***
		(0.021)		(0.169)		(0.220)		(2.908)		(0.023)
1 if Higher education		-0.047		1.647***		1.357		156.545***		0.242***
		(0.030)		(0.303)		(0.313)		(42.050)		(0.028)
Constant	1.050***	1.078***	1.013	0.949	0.258***	0.238***	0.519**	0.029***	0.675***	0.605***
	(0.046)	(0.045)	(0.247)	(0.243)	(0.095)	(0.092)	(0.137)	(0.010)	(0.046)	(0.046)
Region fixed effects	√	√	√	√	√	√	√	√	√	√
Birth cohort fixed effects	√	√	√	√	√	√	√	√	√	√
Model	OLS		Multinomial logit						OLS	
Sample	5,043		3,712						3,712	

Note: Estimation: OLS in Columns (1-2) and Columns (9-10). Multinomial logit in Columns (3-8). The base categories are "not employed" in Column (1-2), "unskilled manual/agriculture/service worker" in Columns (3-8) and "receiving payment in kind/in cash and in kind/nothing" in Columns (9-10). Relative risk ratios are reported in the multinomial logit regression. The treated language is Afan Oromo and control language is Amharic. The interaction between "Treated Language" and "Aged 0-14 in 1994" gives the difference-in-difference estimate. Standard errors, clustered at the region-birth year level, are reported in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: The Ethiopian Demographic and Health Survey (EDHS) 2011.

Chapter 5

Risk Attitudes, Job Mobility and Subsequent Wage Growth

BETHLEHEM A. ARGAW (Leibniz Universität Hannover)

MICHAEL F. MAIER (ZEW, Mannheim)

OLGA J. SKRIABIKOVA (Provincie Limburg)

Bethlehem A. Argaw and Michael F. Maier gratefully acknowledge financial support from the German Ministry of Education and Research (BMBF) under grant 01PW11001 under the funding framework “The Economics of Science”. The views expressed in this article are those of the authors and do not necessarily reflect the views of the BMBF. While working on this paper, Bethlehem Argaw was a researcher in the Labour Markets, Human Resources and Social Policy Department of the Centre for European Economic Research (ZEW). For helpful discussions, the authors would like to thank Thomas Dohmen, Bernd Fitzenberger, Susanne Steffes and participants at the Labour Markets, Human Resources and Social Policy department’s seminar at the ZEW, AiO seminar at the School of Business and Economics (GSBE)-Maastricht University, GlaD PhD workshop at the University of Goettingen and participants at the SES, EALE, ESPE, EEA and the Annual Conference of the German Economic Association. We are grateful to Carla Krolage and Frederik Weber for excellent research assistance. Any remaining shortcomings are our sole responsibility.

Abstract

We investigate the relationship between willingness to take risks and job mobility during the early career. Changing jobs is a risky decision since it involves incurring substantial costs without entirely foreseeing future benefits. We incorporate a risk preference parameter that influences job change behavior in on-the-job search model accounting for non wage job characteristics. Using data from the German Socio-Economic Panel (SOEP), we show that risk-averse individuals experience fewer voluntary job changes compared to more risk-tolerant individuals. In addition, since risk-averse individuals demand higher compensation for the risk associated with changing jobs, their job changes are on average associated with relatively higher wage gains. These empirical findings enhance our understanding of job mobility decisions during the early career in that risk attitudes play a crucial role in explaining job changing behavior and may lead to heterogeneous patterns of subsequent wage growth during the early career.

JEL classification: D81, J31, J62

Keywords: risk attitudes, early-career job mobility, wage growth

5.1 Introduction

The majority of lifetime job changes occur during the first ten years of labour market experience (Topel and Ward, 1992; Dustmann and Pereira, 2008). The importance of job mobility for individuals' success on the labour market in general and for wage growth in particular is a widely discussed topic in theoretical and empirical studies in labour economics (e.g. Topel and Ward, 1992; Light and McGarry, 1998; Hunt, 2001; le Grand and Taehlin, 2002; Neumark, 2002; Fuller, 2008). Job changes within the first years on the labour market are especially important since decisions that are made in this period can strongly influence labour market prospects of the whole career (Topel and Ward, 1992; von Wachter and Bender, 2006; Möller and Umkehrer, 2014).

We argue that willingness to take risks is a key factor that affects job changing behaviour of labour market entrants, who subsequently experience heterogeneous patterns of wage growth. Individual attitudes towards risk has received little attention in theoretical models of job search (for instance in search and matching theories) since these models assume risk neutrality and specify the wage of the current and new job as the main determinants for job changes. However, a growing literature has demonstrated the importance of job characteristics other than the wage for on-the-job search and mobility decisions (e.g. Bonhomme and Jolivet, 2009; Sullivan and To, 2014; Bonhomme et al., 2015). Unlike wages, the non-wage characteristics of jobs such as the possibility of promotion, actual working hours, the flexibility of working time and other working conditions cannot be fully specified ex-ante in working contracts. Instead, they are revealed after spending some time on the job.

We argue that individual attitudes towards risk is one of the crucial determinants of job mobility since changing a job is a risky decision. To change jobs, workers need to incur several costs while future benefits are uncertain at the moment the decision is taken. Costs associated with job changes include the loss of fringe benefits, costs related to moving and adjusting to a new environment (van den Berg, 1992). This implies that risk-averse individuals will make fewer job changes compare to individuals who are more risk-tolerant. Additionally, since risk-averse individuals demand higher compensation for the risk associated with changing a job compared to more risk-tolerant individuals, the job changes of risk-averse individuals will be accompanied by larger wage increases relative to risk-tolerant individuals.

We test these predictions empirically using data from the German Socio-Economic Panel survey (SOEP). The SOEP is one of the longest panel surveys and it contains a direct measure of individual risk attitudes. We first relate risk attitudes of labour market participants to the number of job changes they make during the early years of their career. Second, we examine whether there is a difference in the relationship between job change and wage change depending on individuals risk attitudes. Our first result is that individuals who are less willing to take risks make fewer job changes compared to individuals who are more tolerant to risk. We then document a significant difference in the wage growth associated with the job changes of risk-averse individuals compared to more risk-tolerant individuals. We find that the wage gain associated with the job changes of risk-averse individuals is, on average, higher compared to the wage gains of individuals who are risk-tolerant.

Our results shed some light on the contradicting findings in the literature studying the effects of job mobility on wage growth. Different wage growth patterns associated with job changes made by risk-averse and risk-tolerant individuals that we empirically document in this paper enhance our understanding of the relationship between job mobility and wage growth. Our empirical findings have important implications when examining differences in labour market outcomes across individuals who presumably differ in their attitudes towards risk. For instance, the literature shows that heterogeneity in willingness to take risks exists between men and women (Dohmen et al., 2011) as well as between native and migrants (Bonin et al., 2009). Given our finding that risk attitudes is a crucial behavioural trait that influences individuals job mobility decisions, the existing wage gap across groups can partially be explained by differences in job mobility decisions driven by the variation in risk attitudes across groups.

Our paper is closely related to studies that examine the influence of risk attitudes on various labour market outcomes. Empirical studies in the economics and sociological literature show that individuals who are more willing to take risks are more likely to invest in human capital (Shaw, 1996; Brown and Taylor, 2005; Budria et al., 2012), become private sector employees (Fuchs-Schündeln and Schündeln, 2005; Pfeifer, 2010) or self-employed (Caliendo et al., 2009; Fossen, 2011; Skriabikova et al., 2014). Furthermore, risk-tolerant individuals sort into occupations with high earnings variance (Bonin et al., 2007; Skriabikova et al., 2013) or high mortality risk (DeLeire and Levy, 2004). To the best of our knowledge, we are the first to empirically document the relationship between risk attitudes and job mobility.

This paper's contributions are three fold. First, we empirically examine the relationship between risk attitudes and job changes. As to our knowledge, how variation in risk attitudes lead to different job mobility patterns among labour market entrants has not been studied so far. Second, we analyze the heterogeneity of the relationship between job mobility and wage growth with respect to risk attitudes. By doing so, we are able to shed some light on the contradicting results found in the literature that links job mobility and wage growth. Finally, since we show that job mobility is one channel that links risk attitudes and wage growth, we are able to find a bridge between the literature that focuses on the relationship between risk attitudes and wages on the one hand, and the relationship between job mobility and wage growth on the other.

The paper proceeds as follows. In the next section we summarize previous literature. In section 5.2 we sketch our conceptual framework to explain why risk attitudes should affect job changes and how the relationship between job changes and associated wage growth depends on the willingness to take risks. In section 5.3, the data and estimation sample are described. Section 5.4 presents our main findings and tests for various sensitivity checks. The last section concludes.

5.2 Conceptual Framework

In the following, we present a theoretical model that formalises the relationship between risk attitudes and job mobility. In the canonical on-the-job search model (Burdett, 1978; Mortensen, 1986), a worker changes his/her job by comparing the current wage and the offered wage at the new job. Hwang et al. (1998) highlights the importance of job characteristics other than the wage for on-the-job search and introduced nonwage components in the on-the-job search model. Non-wage characteristics of jobs include the flexibility of working, actual working hours, the working environment and working conditions. Subsequent empirical analysis confirm the importance of nonwage job characteristics for the decision to change jobs (e.g. Bonhomme and Jolivet, 2009; Sullivan and To, 2014; Bonhomme et al., 2015).

We extend the recent on-the-job search model by Sullivan and To (2014) and incorporate individual risk attitudes into the model to take into account a worker's uncertainty about the nonwage characteristics of jobs when making a decision to change jobs. Our model additionally reflects that workers might value job characteristics heterogeneously (Bhaskar and To, 1999). The discounted expected value of lifetime utility for a worker in the current job is given as

$$V_e(U_0) = U + q[V_u - V_e(U_0)] + \delta[\lambda E \max\{V_e(U_1), V_e(U_0)\}] \tag{5.1}$$

where $V_e(U_0)$ is the lifetime utility of the current job, $V_e(U_1)$ is the lifetime utility of the new job, V_u is the expected utility of unemployment, q is the job destruction rate, λ is the new job offers arrive rate and δ is the discount rate. New job offers arrive at a rate of λ as a random draw (w, ω) from the distribution $F(w, \omega)$ which is ex-ante unknown to the worker. Workers decide whether to change their job or not based on the expected value of the better option $E \max\{V_e(U_1), V_e(U_0)\}$ evaluating the discounted expected value of lifetime utility of the current job $V_e(U_0)$ and of the job offer $V_e(U_1)$.

Unlike the canonical on-the-job search model, the utility from employment (U) is determined by the wage (w) and the utility from nonwage job characteristics (ω) such that

$$U = w + \omega \tag{5.2}$$

Individuals change their job if the utility level of the offered job is greater than the utility level of the current job ($U_1 > U_0$) as $V_e(U)$ is increasing in U. In this case, the offer is exceeding the reservation utility U_r such that

$$V_e(U_1) \geq V_e(U_r) \tag{5.3}$$

We extend Sullivan and To (2014)'s model by introducing a parameter that represents differences in individuals' degree of risk aversion, which we term as π. Let's assume that an individual with degree of risk aversion π_1 is more risk-averse than an individual with the degree of risk aversion π_2. That is,

$$\pi_1 > \pi_2, \quad \pi_1, \pi_2 \in \Pi \tag{5.4}$$

Thus, the expected utility of a new job is evaluated differently by the individuals according to their risk aversion. We assume that the reservation utility (U_r) is augmented by a risk premium (P), which depends on individuals' degree of risk aversion (π).

$$V_e(U_r) \geq V_e(U) + P(\pi) \tag{5.5}$$

As a result, the discounted value of employment of an employed person given by equation (5.1) changes to

$$V_e(U_0) = U + q[V_u - V_e(U_0)] + \delta[\lambda E \max\{V_e(U_1) - P(\pi), V_e(U_0)\}] \tag{5.6}$$

Individuals change their job only if the expected utility of the job offer exceeds the expected utility of the current job by the risk premium $P(\pi)$. The exit rate out of the current job is then defined by

$$\theta(U_0, U_1, \pi) = \lambda 1[V_e(U_1) > V_e(U_0) + P(\pi)] \tag{5.7}$$

Hence, the main hypothesis based on our theoretical model is that more risk-averse individuals change their jobs less often compared to individuals who are more risk-tolerant. Our exploration of the relationship between risk attitudes and job mobility also has important implications on the relationship between job mobility and wage growth. There is less consensus both in theoretical models and empirical studies on the question whether or not job-to-job transitions yield positive wage growth. On the one hand, on-the-job search theory predicts that individuals change their job only when a wage offer exceeds the wage of the current job, implying that job-to-job transitions lead to positive wage growth (Burdett, 1978; Jovanovic, 1979a). On the other hand, matching models state that the quality of a job match is only revealed after the individual spends some time on the job and hence job-to-job transitions do not necessarily lead to positive wage growth (Jovanovic, 1979b).

The introduction of risk attitudes into these models provides several predictions on the relationship between job mobility and wage growth. First, if we assume that a job's uncertain utility is only related to the nonwage job characteristics ω, the contracted wage at the new job includes a compensation for these uncertainties. The realized wage growth accompanied with each voluntary job change should therefore be higher for more risk-averse individuals, as they require higher compensation when choosing a risky alternative compared to individuals who are more risk-tolerant. On the other hand, as risk-averse individuals are less likely to change their job ($\theta(U_0, U_1, \pi_1) < \theta(U_0, U_1, \pi_2)$), the probability of choosing jobs where the match quality reveals to be worse or better than the current job is lower. A lower frequency of job changes $\theta_1(U_0, U_1, \pi_1)$ can lead to positive wage growth since it lowers the probability of choosing a job where the match quality reveals to be worse than the current job while leading to positive returns from accumulated firm-specific capital. On the contrary, lower frequency of job changes often provides fewer opportunities to learn about one's own ability and preferences and hence the opportunity to improve job match quality (Farber and Gibbons, 1996; Antonovics and Golan, 2012; Papageorgiou, 2014). Hence, our model's prediction that the frequency of voluntary job changes varies by individual risk attitudes provides ambiguous predictions on the relationship between job mobility and overall wage growth, i.e., the total wage growth during certain years in the labour market and hence, remains an empirical question that we investigate in this paper.

5.3 Data and Variables

The empirical analysis is based on data from the German Socio-Economic Panel Survey (SOEP). The SOEP is a representative household panel survey conducted yearly since 1984 in Western Germany and since 1991 in Eastern Germany (see Wagner et al. (2007) for the details). Our research question pertains to early-career job mobility. There is no widely accepted number of years in the labour market that captures 'the early career'. For instance, Topel and Ward (1992) and Manning and Swaffield (2008) define the early career as the first ten years after labour market entry whereas other studies consider the first five years (Neumark, 2002), seven years (Johnson, 1978) or a combination of different years (Light and McGarry, 1998). We use the first seven years after labour market entry as 'the early career' in order to capture sufficient job mobility patterns and to maintain an adequate number of observations for the analysis.[1]

[1] Following labour market entrants over a longer period of time has the advantage of capturing their lifetime job mobility patterns. However, this comes at a cost of fewer number of observations as individuals drop out of the survey for various reasons.

Our sample consists of individuals who are observed entering the labour market in the SOEP since 1992. The main outcome variables in our empirical analysis are the number of job changes and wage levels. These variables are measured in the SOEP only for individuals who are employed at the time of the interview. Following this, we use two samples of labour market entrants depending on their labour market status at the time of the interview. Our main sample consists of individuals who are employed at each interview during their first seven years in the labour market. Alternatively, we use a less restrictive and, thus, a more heterogeneous sample of labour market entrants who are not necessary employed at the time of the interview.

We further limit both samples to those individuals who started as full-time or regular part-time employees in order to exclude student jobs and other irregular employments. Individuals who were below 18 or above 32 years of age at the time of entry as well as those who entered the labour market as self-employed are excluded from the analysis. Table 5.1 describes the step-by-step sample selection procedure and the sample of labour market entrants available for the analysis. Our preferred estimation sample is the more restrictive sample since it represents a more homogeneous sample of labour market entrants with strong labour market attachment. This, however, comes at the cost of a smaller size (280 observations) compared to the non-restrictive sample which has 1370 observations albeit missing yearly information on job changes and wage levels.

Table 5.1: Sample selection procedure

Respondents	Sample selection steps
4077	Labor market entrants in SOEP sample starting from 1992
−1428	not employed as full-time or regular part-time at first job
2649	
−95	not between 18 and 32 years of age at first job
2554	
−52	self-employed at first job
2502	
−907	missing data on risk attitudes
1595	
−225	missing data on control variables
1370	**Nonrestrictive sample**
−1090	not employed consecutively during the first seven years after labor market entry
280	**Restrictive sample**

Note: Step by step deletion of respondents which do not fit the sample definition or for which essential information is missing.
Source: German Socioeconomic Panel Survey (SOEP) 1992-2013.

The first variable of interest is the total number of job changes that individuals experience during the first seven years in the labour market. The variable is measured using the survey question on whether respondents have started a new job since the previous interview. Our focus is on voluntary job changes, i.e. job changes that are initiated by the worker. We focus on voluntary job changes since there is no clear theoretical prediction on the relationship between individual risk attitudes and involuntary job changes and the subsequent wage growth. We distinguish voluntary from involuntary job changes using the unemployment duration workers experience between jobs as in Perez and Sanz (2005) and Pavlopoulos et al. (2014). We consider a job change as voluntary if workers experience at most three months of unemployment after leaving a job.

An alternative approach to distinguish voluntary from involuntary job change is based on the reasons provided by respondents for changing a job (Hunt, 2001; Fuller, 2008). Due to substantial missing information, we are not able to fully exploit this information on job changes. We test the robustness of the results by defining a job change as voluntary if i) a worker experiences at most three months of unemployment during a job change, and ii) a worker changed the job for reasons other than termination by employer and company closure. Of those job changers who stated the reason for a job change in our estimation sample, about 18 percent specify "terminated by employer" and "company closed down" as the reasons for a job change. This alternative definition of voluntary job change allows us to exclude job changes that, most likely, occurred involuntarily.

Our measure of job change comprises voluntary job changes within and across firms, industries and/or occupations. The risk associated with changing a job across firms, industries and/or occupations is likely to be greater than job changes within the same firm, industry and/or occupation. Survey respondents are asked whether they change jobs within the same firm or to a new firm. Of those job changers with non-missing information, about 63 percent of the job changes are to a new firm and about 10 percent of the job changes took place within the same firm. Given the sample size, it is not possible to separately analyse job changes within and across firms, industries and occupations. We instead control for industry and occupation fixed effects to capture industry and occupation level differences in job mobility and wages.

Our main explanatory variable of interest is individuals' willingness to take risks. During certain survey years starting in 2004, respondents are asked to provide their attitude towards taking risk in different domains such as occupations and health. We use individuals' willingness to take risk in occupational matters. Individuals were asked about their willingness to take risk in occupational matters during the survey years 2004 and 2009. The variable is measured on a scale from zero to ten, in which higher values reflect greater willingness to take risks. The exact wording of the question, translated from German, reads "People can behave differently in different situations. How would you rate your willingness to take risks in the following areas. How is it in your occupation? Please give me a number from 0 to 10, where the value 0 means: "Risk-averse" and the value 10 means: "Fully prepared to take risks". You can use the values in between to make your estimate." Dohmen et al. (2011) show that this question is significantly related to paid lottery choices, and it explains behaviour in a range of important domains of real life decisions. For ease of interpretation, we define a binary risk attitudes measure by grouping individuals with risk attitudes below the median as risk-averse and those with risk attitudes above the median as more risk-tolerant.

A recent literature strand addresses the stability of individuals' risk preferences over time using measures based on subjective assessments and (hypothetical and incentivised) lottery choices (e.g. Harrison et al. (2005); Sahm (2007); Andersen et al. (2008); Baucells and Villasís (2009); Reynaud and Couture (2012)). These studies do not find evidence for systematic changes of a person's risk preferences, with the exception of age. Sahm (2007), for example, reports a general increase in risk aversion with age, but finds that risk preferences are rank-order stable. Dohmen et al. (2012b) also provide evidence that risk attitudes measured by the survey question that we use in this study, are rather stable. In our restrictive sample of labour market entrants, the mean and median differences between reported risk attitudes in 2004 and 2009 is 0.60 and 1, respectively. When we define risk attitudes as a binary indicator for being risk-averse, the mean and median difference between 2004 and 2009 are zero. This implies that individual attitude towards risk is rather stable over the period of time relevant for our study.

Descriptive statistics are shown in appendix Table A5.1 for the restrictive sample in column (1) and the nonrestrictive sample in column (2). About 60% of labour market entrants changed their job voluntarily at least once during the first seven years in the labour market. Figure 5.1 shows the average number of voluntary job changes experienced by risk-averse and by risk-tolerant individuals. The group of risk-averse individuals on average experience about 0.69 voluntary job changes during the first seven years on the labour market whereas more risk-tolerant individuals make one voluntary job change. As years of experience in the labour market increases, the gap in the number of job changes experienced by risk-averse and more risk-tolerant individuals widens.

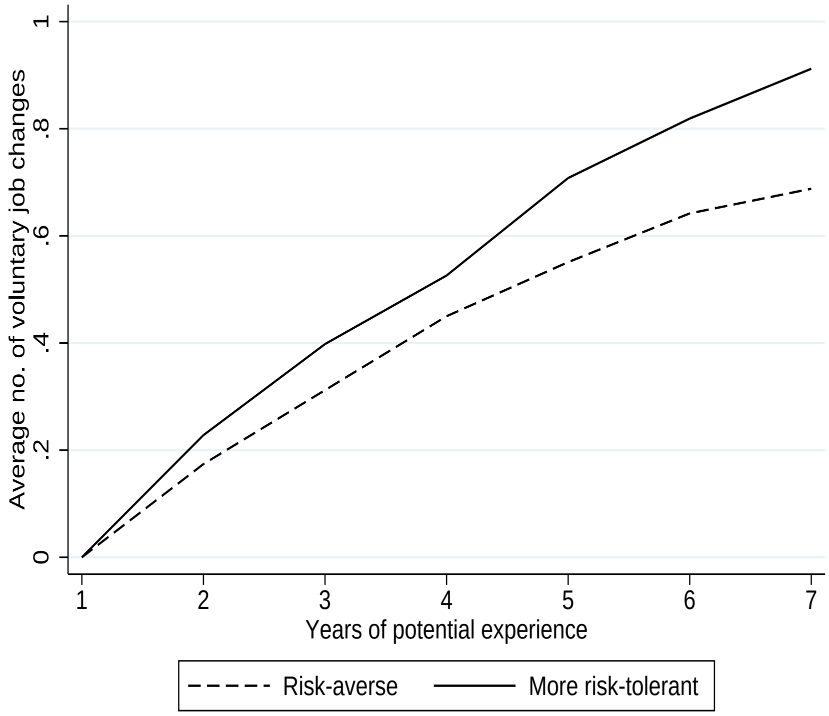


Figure 5.1: Average number of voluntary job changes during the early career by level of risk aversion.

The average number of job changes that workers experience in Germany is considerably lower compared to the United States and the United Kingdom. Without differentiating voluntary from involuntary job changes, Dustmann and Pereira (2008) show that workers in Germany hold on average about 2.7 jobs during the first ten years in the labour market. Workers in the United States and in the United Kingdom hold about 6.96 and 4 jobs respectively during the same period. The low job mobility in Germany can be attributed to the higher costs of firing and the substantial screening at job entry.

5.4 Estimation Method and Results

5.4.1 The Influence of Risk Attitudes on Voluntary Job Changes

We use the following specification to estimate the relationship between individual risk attitudes and job mobility during the early career.

$$TJC_i = \alpha_0 + \alpha_1 RA_i + \delta X_i + \mu_i \tag{5.8}$$

where TJC_i is the total number of voluntary job changes during the first seven years in the labour market. RA_i is a binary measure of individual risk attitudes which takes the value one if individual i's risk attitude is below the median of the distribution. X_i includes other control variables. μ_i is an error term. We use an OLS regression to estimate equation (5.8). α_1 gives a consistent estimate of the effect of risk attitudes on the number of voluntary job changes under the assumption that individual risk attitudes is uncorrelated with μ_i. We control for various factors and undertake different sensitivity checks, which we explain in detail below, to support causal interpretation of the coefficient.

Table 5.2 shows the estimation results. In column (1), the simple correlation between risk attitudes and the number of job changes is displayed. The coefficient on risk attitudes is negative and shows that risk-average individuals make fewer job changes during their first seven years in the labour market compared to more risk-tolerant individuals. The estimate remains significant when we control for basic demographic and socio-economic characteristics (gender, age, indicators for residing in West Germany (in 1989) and for being a German national) and year dummies in column (2).

Table 5.2: The influence of risk attitude on total number of job changes

	(1)	(2)	(3)	(4)	(5)	(6)
1 if Risk-averse	-0.224**	-0.246**	-0.252**	-0.262**	-0.276**	-0.340***
	(0.099)	(0.100)	(0.102)	(0.102)	(0.109)	(0.127)
1 if Male		-0.098	-0.089	-0.065	-0.068	-0.040
		(0.101)	(0.111)	(0.112)	(0.125)	(0.147)
1 if Originates from West Germany		0.055	0.021	0.042	-0.029	-0.050
		(0.126)	(0.123)	(0.126)	(0.126)	(0.140)
1 if German national		0.213	0.134	0.154	0.074	0.187
		(0.139)	(0.172)	(0.184)	(0.189)	(0.259)
1 if Low or no education degree		0.016	-0.002	-0.075	-0.013	-0.056
		(0.214)	(0.231)	(0.231)	(0.223)	(0.263)
1 if Tertiary education degree		0.249	0.226	0.246	0.090	0.135
		(0.154)	(0.169)	(0.176)	(0.214)	(0.233)
Age at first job		-0.024	-0.027	-0.027	-0.028	-0.038
		(0.022)	(0.024)	(0.024)	(0.024)	(0.030)
Wage at first job			-0.253*	-0.243*	-0.244*	-0.267*
			(0.141)	(0.136)	(0.134)	(0.151)
Job satisfaction at first job			-0.031	-0.029	-0.032	-0.038
			(0.025)	(0.026)	(0.026)	(0.029)
1 if Public sector at first job			-0.016	0.036	-0.038	-0.065
			(0.126)	(0.144)	(0.150)	(0.177)
1 if Permanent contract at first job			-0.066	-0.110	-0.117	-0.139
			(0.113)	(0.118)	(0.121)	(0.133)
1 if Part-time employment at first job			0.004	-0.069	-0.074	-0.092
			(0.163)	(0.156)	(0.156)	(0.161)
Tenure at first job			-0.052	-0.044	-0.047	-0.061
			(0.044)	(0.045)	(0.045)	(0.055)
Wage variance in occ. at first job			-0.311	-0.643	-0.034	0.242
			(0.757)	(0.197)	(1.034)	(1.206)
Constant	0.912***	1.265***	3.657***	3.872***	3.893***	3.291**
	(0.067)	(0.485)	(1.239)	(1.200)	(1.342)	(1.342)
Year dummies		√	√	√	√	√
Parental background		√	√	√	√	√
Industry dummies and firm size				√	√	√
Occupation dummies					√	√
Econometric model	OLS	OLS	OLS	OLS	OLS	Poisson
(Pseudo) R-squared	0.017	0.041	0.113	0.141	0.182	0.070
Sample	280	280	280	280	280	280

Note: Dependent variable: Total number of job changes during the first seven years on the labour market. Heteroscedasticity-consistent standard errors are in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: German Socioeconomic Panel Survey (SOEP) 1992-2013.

Column (3) takes into account additional confounding factors. Individuals who experience a good match at the first job, in terms of productivity and preferences, are less likely to change jobs compared to individuals who experience a bad match. To rule out the possibility that job match quality could differ by individual attitudes towards risk, we control for wage and job satisfaction at the labour market entry job. In addition, we also control for the type of contract (temporary or permanent), the type of employment (part-time or full-time), previous experience with the firm and the size of the firm at the labour market entry job.

Previous studies show that individuals with higher risk tolerance invest more in human capital(Brown and Taylor, 2005; Budria et al., 2012), sort into occupation with high earning variance (Bonin et al., 2007; Skriabikova et al., 2013) whereas risk-averse individuals sort into public sector employment (Pfeifer, 2010). We capture these differences in initial sorting factors by controlling for individuals' highest education degree, whether one is employed in the public or private sector, occupation-specific earnings variance at 2-digit level and occupation and industry fixed effects at 1-digit level. Furthermore, if individuals from rich families are more risk-tolerant and at the same time change jobs more often, family background characteristics could lead to a positive correlation between risk attitudes and job mobility.[2] We use paternal education and parents occupation to control for such an intergenerational link.

As the estimation results in column (3) show the coefficient on risk attitudes remains negative when we account for additional confounding factors. The result is in line with our theoretical prediction supporting the proposition that risk-averse individuals experience fewer number of voluntary job changes compared to more risk-tolerant individuals. Adding fixed effects at for the industry and occupation at the labour market entry job in columns (4) and (5) does not affect the estimation results. This implies that risk attitudes matters for voluntary job changes within as well as across industries and occupations.

Interpreting the magnitude of the effect of risk attitudes from column (5), being risk-averse in occupational matters reduces the average number of job changes experienced

[2] For an empirical evidence on the intergenerational transmission of risk attitudes, see Dohmen et al. (2012a).

during the first seven years in the labour market by 0.276, which is one third of a standard deviation. We additionally estimate a poisson regression to take into account the discrete nature of the dependent variable. The result, shown in column (6), gives us an estimate larger than the OLS estimate both in magnitude and statistical significance. Given the considerably low number of job changes that workers experience in Germany, we also estimated an ordered logit model using a categorical job change variable for experiencing no job change, one job change, two or more job changes. The effect of individual risk attitudes on job changes remains statistically significant. The estimated odds ratio is 0.476 (standard error: 0.133).

It is noteworthy to interpret the relationship between job mobility and some of the control variables. We find some evidence supporting Jovanovic (1979b)'s model of job matching, which states that workers change jobs when the value of an outside offer is higher than the value of the current productivity match. Across all specifications, the wage level of the first job is significantly negatively correlated with the number of job changes. Furthermore, higher level of job satisfaction at first job is associated with experiencing fewer number of job changes during the early career.

We also find that university educated workers make more job changes compared to workers with vocational education or less skilled workers. This is because higher education provides a broader set of skills, which increases the range of job opportunities whereas vocational education provides better match quality at labour market entry thereby reducing the probability of changing a job. As expected, individuals employed in the public sector as well as those who started their career with a permanent contract are less likely to experience frequent job changes. Moreover, individuals with work experience at the firm prior to their first job, e.g. as an apprentices, are less likely to change jobs. However, the coefficients are not statistically significant.

Table 5.3 provides estimation results for three robustness checks. In column (1) and (2), we use the unrestricted sample which contains all labour market entrants irrespective of their employment status at the time of the interview. In column (3) and (4), we show estimation results where we use the subjective information available in the SOEP to define voluntary job change. In column (5) and (6), we restrict the estimation sample to labour market entrants who started their career in the year 2000 and on wards instead of 1992 to reduce the possibility of reverse causality.

Table 5.3: Robustness checks

	Non restrictive sample		Alternative voluntary job change definition		Year of labor market entry since 2000	
	(1)	(2)	(3)	(4)	(5)	(6)
1 if Risk-averse	-0.103***	-0.087**	-0.277**	-0.381***	-0.302**	-0.366**
	(0.040)	(0.036)	(0.095)	(0.102)	(0.127)	(0.156)
1 if Male		0.073*		-0.053		-0.172
		(0.044)		(0.128)		(0.152)
1 if from West Germany		0.009		-0.014		-0.052
		(0.042)		(0.132)		(0.122)
1 if German national		0.156**		-0.034		0.122
		(0.066)		(0.185)		(0.281)
1 if Low or no education degree		0.001		0.003		0.260
		(0.052)		(0.235)		(0.308)
1 if Tertiary education degree		0.117*		-0.024		0.015
		(0.065)		(0.189)		(0.272)
Age at first job		-0.017**		-0.015		0.006
		(0.008)		(0.022)		(0.034)
Wage at first job		-0.123***		-0.254*		-0.125
		(0.035)		(0.136)		(0.174)
Job satisfaction at first job		-0.019**		-0.048*		-0.059*
		(0.009)		(0.025)		(0.034)
1 if Public sector at first job		-0.075		-0.025		0.026
		(0.055)		(0.142)		(0.171)
1 if Permanent contr. at 1st job		-0.037		-0.123		-0.262
		(0.040)		(0.114)		(0.159)
1 if Part-time empl. at 1st job		-0.114**		-0.090		-0.098
		(0.049)		(0.152)		(0.208)
Tenure at first job		-0.037***		-0.046		-0.013
		(0.014)		(0.043)		(0.051)
Wage variance in occ. at 1t job		0.758**		0.250		-0.820
		(0.312)		(1.094)		(1.510)
Constant	0.538***	2.049***	0.847***	3.503**	0.912***	3.024
	(0.028)	(0.351)	(0.068)	(1.423)	(0.086)	(1.875)
Controls		√		√		√
R-squared	0.005	0.242	0.029	0.235	0.031	0.313
Sample	1370	1370	257	257	166	166

Note: Estimation: OLS. Dependent variable: Total number of job changes during the first seven years on the labour market. Controls include year dummies, parental background, firm size, industry and occupation fixed effects. Heteroscedasticity-consistent standard errors are in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: German Socioeconomic Panel Survey (SOEP) 1992-2013.

Over all, the estimation results show that the effect of risk attitudes remains negative and statistically significant. The estimated influence of risk attitudes on number of job changes is smaller in magnitude when we use the unrestricted sample of labour market entrants. Given the larger sample size, most of the explanatory variables which were statistically insignificant in Table 5.2 become significant. Defining voluntary job changes based on the reasons given by the respondents instead of the unemployment duration they experience between jobs has negligible impact on the estimated influence of individual attitudes towards risk on the number of voluntary job changes. When we further restrict the total number of voluntary job changes to those that occurred only across firms, the estimated influence of risk attitudes becomes -0.314 (standard error: 0.097).

Though the literature and our descriptive analysis show that individuals attitudes towards risk are stable, we are able to check for a possible reverse causality running from job mobility to risk attitude by restricting the estimation sample to labour market entrants who started their career in the year 2000 and on wards instead of 1992. The survey question on risk attitudes in the SOEP was asked for the first time in 2004. By starting the year of labour market entry in 2000 (and not in 1992), we only consider job changes that took place after (close to) the year when individuals attitudes towards risk are measured. The results, shown in column (5) and (6), show that reverse causality does not influence the central message of our empirical analysis.

5.4.2 Risk Attitudes, Job Changes and Wage Growth

The conceptual framework discussed in section 5.2 predicts that the voluntary job changes of risk-averse individuals is, on average, accompanied by higher wage gains compared to the voluntary job change of more risk-tolerant individuals. We use the following wage growth equation to test this prediction.

$$lnW_{it} = \beta_0 + \beta_1 JC_{it} + \theta Z_{it} + \lambda Q_i + c_i + \epsilon_{it} \tag{5.9}$$

where lnW_{it} is the log of real hourly wage at time t; JC_i is an indicator variable for making a voluntary job change. Z_{it} includes time-variant individual and firm level characteristics such as work experience, tenure, firm size, industry and occupation dummies. Q_i includes time-invariant demographic characteristics, measures of job match quality in the first job and initial sorting factors.

We estimate equation (5.9) for a sample of labour market entrants who experienced at least one voluntary job change during the first seven years in the labour market. We restrict the estimation sample to movers in order to get a homogeneous group of individuals with comparable wage growth as in Bono and Vuri (2011). Individuals who have never made a job change could be substantially different from job changers and hence their wage growth might not be comparable. We estimate the wage growth equation using a fixed effect estimator. The fixed effect estimator controls for any time-invariant unobservable characteristics which may bias the estimated effect of voluntary job change on wage growth.

We estimate equation (5.9) separately for risk-averse and more risk-tolerant individuals. This is important because risk attitudes have been shown to influence other factors that are associated with wage growth. For instance, Shaw (1996) provides both theoretical and empirical evidence showing that more risk-tolerant individuals invest more on human capital which in turn leads to higher wage growth. Budria et al. (2012) replicate Shaw (1996) using data from Germany, Italy, Spain and additional observations from the US. The authors find mixed results. For Germany, they find that more risk-tolerant individuals obtain higher wage growth by obtaining higher work experience.[3]

Table 5.4 shows the estimation results based on the restrictive estimation sample whereas Table 5.5 shows the estimation results based on the less respective estimation sample. Column (1) and (2) of both tables show the estimation results for a pooled sample of risk-averse and risk-tolerant individuals. We then estimate equation (5.9) separately for

[3] We check the sensitivity of the results by interacting risk attitudes and job change in the wage equation instead of estimating a separate wage equation for risk-averse and more risk-tolerant individuals.

risk-averse and risk-tolerant individuals in columns (3) and (4) and in columns (5) and (6), respectively.

Table 5.4: Risk attitudes, job changes and wage growth

	All		Risk-averse		Risk-tolerant	
	(1)	(2)	(3)	(4)	(5)	(6)
1 if Voluntary job change	0.053**	0.053***	0.060	0.060*	0.049*	0.048*
	(0.021)	(0.020)	(0.037)	(0.033)	(0.026)	(0.025)
Experience		0.090***		0.092***		0.088***
		(0.011)		(0.019)		(0.014)
Experience square		-0.003**		-0.001		-0.003**
		(0.001)		(0.002)		(0.002)
Tenure		0.004		-0.007		0.006
		(0.006)		(0.012)		(0.007)
Occupation dummies		√		√		√
Industry dummies		√		√		√
Within R-squared	0.006	0.312	0.007	0.400	0.006	0.281
Overall R-squared	0.002	0.103	0.005	0.134	0.001	0.091
Observations (person)	165		58		107	
Observations (person*year)	1155		406		749	

Note: Estimation: Fixed Effect Model. Dependent variable: Monthly real wage. Estimation sample contains job changers only. Heteroscedasticity-consistent standard errors are in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: German Socioeconomic Panel Survey (SOEP) 1992-2013.

The results based on the restrictive estimation sample shows that voluntary job change is associated with a 5.3 percent increase in wages. The influence of voluntary job change on wage growth is about 6 percent for risk-averse individuals and 4.8 percent for more risk-tolerant individuals. The estimated coefficients for risk-averse and more risk-tolerant individuals, however, are not statistically different from each other.

When we use the less restrictive estimation sample, the influence of voluntary job change on wage growth becomes statistically and economically different between risk-averse and risk-tolerant individuals. A voluntary job is associated with a 13 percent increase in wages whereas the job changes of risk-tolerant individuals is associated with a 6 percent increase in wages. Similar to Budria et al. (2012), we find a marginally significant difference (at the 10 percent level) in the return to years of experience between risk-averse and more risk-tolerant individuals. The descriptive statistics shown in Appendix Table A5.1 indicate that there are no substantial differences in the characteristics of job changers between the restrictive and nonrestrictive estimation samples. Hence, we can conclude that the different results we find across the two samples is mainly driven by the sample size rather than the sample composition. The main message of our analysis remains the same when we use the interaction of risk attitudes and job change in the wage equation instead of estimating a separate wage equation for risk-averse and more risk-tolerant individuals (See Appendix Table A5.2).

Table 5.5: Risk attitude, job change and wage growth on nonrestrictive sample

	All		Risk-averse		Risk-tolerant	
	(1)	(2)	(3)	(4)	(5)	(6)
1 if Voluntary job change	0.109***	0.086***	0.143***	0.131***	0.088***	0.061***
	(0.017)	(0.017)	(0.028)	(0.029)	(0.021)	(0.022)
Experience		0.087***		0.066***		0.096***
		(0.011)		(0.018)		(0.013)
Experience square		-0.002*		0.001		-0.004**
		(0.002)		(0.002)		(0.002)
Tenure		0.018***		0.023**		0.016**
		(0.006)		(0.010)		(0.007)
Occupation dummies		√		√		√
Industry dummies		√		√		√
Within R-squared	0.020	0.228	0.003	0.255	0.013	0.225
Overall R-squared	0.001	0.199	0.002	0.201	0.001	0.174
Observations (person)	504		199		305	
Observations (person*year)	2608		1007		1601	

Note: Estimation: Fixed Effect Model. Dependent variable: Monthly real wage. Estimation sample contains job changers only. Heteroscedasticity-consistent standard errors are in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: German Socioeconomic Panel Survey (SOEP) 1992-2013.

Finally, we shed some light on the question whether the job mobility difference between risk-averse and more risk-tolerant individuals during the early career results in different overall wage growth, where we define overall wage growth as the absolute difference between the wage in the first job and the wage in the job at the seventh year of work experience. We do so by estimating a model where overall wage growth is regressed on number of voluntary job changes, a dummy variable for being risk-averse (compared to being risk-tolerant) and their interaction.The specification also includes the control variables specified in equation (5.8). We then calculate and compare the average predicted overall wage growth between risk-averse and risk-tolerant individuals.

The average predicted wage growth in the restrictive sample is about 39 percent for a risk-averse individual and about 37 percent for a more risk-tolerant individual. This implies an overall wage growth difference of about 0.023 percent with a standard error of 0.039, which is statistically and economically insignificant. The average predicted overall wage growth between risk-averse and more risk-tolerant individuals is also very similar in the nonrestrictive sample. The average predicted wage growth in the nonrestrictive sample is about 30.6 percent for a risk-averse individual and about 31.7 percent for a more risk-tolerant individual. The implication is that though risk-tolerant individuals obtain wage gains by improving their job match quality through frequent job changes, risk-averse individuals earn similar wage benefits through their higher returns to job mobility.

5.5 Conclusion

Based on a conceptual framework, we find empirical support for the theoretical prediction that individuals attitudes towards risks is an important factor that influences job mobility decisions thereby generating heterogeneous wage growth patterns. Job change is a risky decision as it entails significant costs while benefits cannot be entirely foreseen. We use a sample of labour market participants from the German Socio-economic panel to examine the job mobility and wage growth patterns of risk-averse and more risk-tolerant individuals.

We find that risk-averse individuals change their jobs less often during the early career. To our knowledge, we are the first to document a robust correlation between willingness to take risks and the number of voluntary job changes. Furthermore, the wage gain associated with job changes made by risk-averse individuals is, on average, higher than the wage gains of individuals who are more risk-tolerant. The different wage growth patterns associated with job changes made by risk-averse and risk-tolerant individuals that we empirically document in this paper enhance our understanding of the relationship between job mobility and wage growth.

Our empirical findings have important implications when examining differences in labour market outcomes across individuals who presumably differ in their attitudes towards risk. For instance, the literature shows that heterogeneity in willingness to take risks exists between men and women (Dohmen et al., 2011) as well as between native and migrants (Bonin et al., 2009). Given our finding that risk attitudes is a crucial behavioural trait that influences individuals job mobility decisions, the existing wage gap across groups can partially be explained by differences in job mobility decisions driven by the variation in risk attitudes across groups.

The low job mobility rate among labour market participants in Germany puts restriction on how further one can explore the relationship between risk attitudes, job mobility and wage growth. A further research would compare and extend the analysis using data from other countries which have higher job mobility rate such as the United States and United Kingdom. A larger sample size also allows one to tackle the potential endogeneity of job changes in the wage growth equation by employing estimation techniques such as instrumental variables. Further research is also required to empirically investigate the importance of individual risk attitudes to on-the-job search. Since not all on-the-job searchers results in actual job mobility, job search intensity may also be related to risk attitudes as the costs and benefits of on-the-job search are uncertain for individuals in a similar to actual job changes.

5.6 References

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2008). Lost in state space: are preferences stable? *International Economic Review*, 49(3):1091–1112.

Antonovics, K. and Golan, L. (2012). Experimentation and job choice. *Journal of Labor Economics*, 30(2):333–366.

Baucells, M. and Villasís, A. (2009). Stability of risk preferences and the reflection effect of prospect theory. *Theory and Decision*, 68(1-2):193–211.

Bhaskar, V. and To, T. (1999). Minimum wages for Ronald McDonald monopsonies: A theory of monopsonistic competition. *The Economic Journal*, 109:190–203.

Bonhomme, S. and Jolivet, G. (2009). The pervasive absence of compensating differentials. *Journal of Applied Econometrics*, 24(5):763–795.

Bonhomme, S., Jolivet, G., and Leuven, E. (2015). School characteristics and teacher turnover: Assessing the role of preferences and opportunities. *The Economic Journal*, forthcoming.

Bonin, H., Constant, A. F., Tatsiramos, K., and Zimmermann, K. F. (2009). Native-migrant differences in risk attitudes. *Applied Economics Letters*, 16 (15):1581–1586.

Bonin, H., Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2007). Cross-sectional earnings risk and occupational sorting: The role of risk attitudes. *Labour Economics*, 14:926–937.

Bono, E. D. and Vuri, D. (2011). Job mobility and the gender wage gap in italy. *Labour Economics*, 18:130–142.

Brown, S. and Taylor, K. (2005). Wage growth, human capital and financial investment. *The Manchester School*, 73(6):686–708.

Budria, S., Diaz-Serrano, L., Ferrer-i-Carbonell, A., and Hartog, J. (2012). Risk attitude and wage growth: Replicating Shaw (1996). *Empirical Economics*, 44 (2):981–1004.

Burdett, K. (1978). A Theory of Employee Job Search and Quit Rates. *The American Economic Review*, 68(1):212–220.

Caliendo, M., Fossen, F. M., and Kritikos, A. S. (2009). Risk Attitudes of Nascent Entrepreneurs - New Evidence from an Experimentally Validated Survey. *Small Business Economics*, 32(2):153–167.

DeLeire, T. and Levy, H. (2004). Worker Sorting and the Risk of Death on the Job. *Journal of Labor Economics*, 22(4):925–954.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2012a). The intergenerational transmission of risk and trust attitudes. *Review of Economic Studies*, 79(2):645–677.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences. *Journal of the European Economic Association*, 9(3):522–550.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2012b). The Measurement and Stability of Risk Preferences. Maastricht University, mimeo.

Dustmann, C. and Pereira, S. C. (2008). Wage growth and job mobility in the United Kingdom and Germany. *Industrial and Labor Relations Review*, 61(3):374–393.

Farber, H. S. and Gibbons, R. (1996). Learning and wage dynamics. *The Quarterly Journal of Economics*, 111(4):1007–1047.

Fossen, F. M. (2011). The Private Equity Premium Puzzle Revisited: New Evidence on the Role of Heterogeneous Risk Attitudes. *Economica*, 78:656–675.

Fuchs-Schündeln, N. and Schündeln, M. (2005). Precautionary savings and Self-Selection: Evidence from the German Reunification Experiment. *The Quarterly Journal of Economics*, 120(3):1085–1120.

Fuller, S. (2008). Job Mobility and Wage Trajectories for Men and Women in the United States. *American Sociological Review*, 73:158–183.

Harrison, G. W., Johnson, E., McInnes, M. M., and Rutström, E. E. (2005). Temporal stability of estimates of risk aversion. *Applied Financial Economics Letters*, 1(1):31–35.

Hunt, J. (2001). Post-unification wage growth in East Germany. *The Review of Economics and Statistics*, pages 190–195.

Hwang, H., Mortensen, D. T., and Reed, W. R. (1998). Hedonic wages and labor market search. *Journal of Labor Economics*, 16(4):815–847.

Johnson, W. R. (1978). A Theory of Job Shopping. *Quarterly Journal of Economics*, 92(2):261–277.

Jovanovic, B. (1979a). Firm-specific Capital and Turnover. *Journal of Political Economy*, 87(6):1246–1260.

Jovanovic, B. (1979b). Job matching and the theory of turnover. *Journal of Political Economy*, 87(5):972–990.

le Grand, C. and Taehlin, M. (2002). Job mobility and earning growth. *European Sociological Review*, 18 (4):381–400.

Light, A. and McGarry, K. (1998). Job change patterns and the wages of young men. *The Review of Economics and Statistics*, 80(2):276–286.

Manning, A. and Swaffield, J. (2008). The gender gap in early-career wage growth. *The Economic Journal*, 118(530):983–1024.

Möller, J. and Umkehrer, M. (2014). Are there long-term earnings scars from youth unemployment in Germany? *ZEW Discussion Paper 2014-089*.

Mortensen, D. (1986). Job search and labor market analysis. In *Handbook of Labor Economics*, volume 2 of *849-919*. North-Holland, Amsterdam.

Neumark, D. (2002). Youth labor markets in the united states: Shopping around vs. staying put. *The Review of Economics and Statistics*, 84(3):462–482.

Papageorgiou, T. (2014). Learning your comparative advantage. *The Review of Economic Studies*, 81(3):1263–1295.

Pavlopoulos, D., Fouarge, D., Muffels, R., and Vermunt, J. (2014). Who benefits from a job change: The dwarfs or the giants? *European Societies*, 16(2):299–319.

Perez, J. I. G. and Sanz, Y. R. (2005). Wage Changes Through Job Mobility in Europe: A Multinomial Endogenous Switching Approach. *Labour Economics*, 12:531–555.

Pfeifer, C. (2010). Risk aversion and sorting into public sector employment. *German Economic Review*, 12(1):85–99.

Reynaud, A. and Couture, S. (2012). Stability of risk preference measures: results from a field experiment on french farmers. *Theory and decision*, 73(2):203–221.

Sahm, C. (2007). How Much Does Risk Tolerance Change? Federal Reserve Board, Washington, D.C.: Finance and Economics Discussion Series.

Shaw, K. L. (1996). An empirical analysis of risk aversion and income growth. *Journal of Labor Economics*, 14(4):626–653.

Skriabikova, O., Dohmen, T., and Kriechel, B. (2013). Risk attitudes and occupational choice: Evidence from an era of evolving heterogeneity in occupation-specific earnings risk. *mimeo*.

Skriabikova, O., Dohmen, T., and Kriechel, B. (2014). New evidence on the relationship between risk attitudes and self-employment. *Labour Economics*, 30:176–184.

Sullivan, P. and To, T. (2014). Search and nonwage job characteristics. *Journal of Human Resources*, 49(2):472–507.

Topel, R. H. and Ward, M. P. (1992). Job mobility and the careers of young men. *The Quarterly Journal of Economics*, 107(2):439–479.

van den Berg, G. (1992). A structural dynamic analysis of job turnover and the costs associated with moving to another job. *The Economic Journal*, 102:1116–1133.

von Wachter, T. and Bender, S. (2006). In the right place at the wrong time: The role of firms and luck in young workers' careers. *American Economic Review*, 96(5):1679–1705.

Wagner, G. G., Frick, J. R., and Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. *Schmollers Jahrbuch*, 127 (1):139–169.

5.7 Appendix

Table A5.1: Descriptive statistics

	Restrictive sample				Nonrestrictive sample			
	(1)				(2)			
	Mean	SD	Min	Max	Mean	SD	Min	Max
1 if atleast one job change	0.59	0.49	0	1	0.37	0.48	0	1
Number of job changes	0.83	0.84	0	4	0.49	0.75	0	4
Number of job changes (among job changers)	1.40	0.62	1	4	1.34	0.61	1	4
Risk attitude	4.85	2.42	0	10	4.52	2.43	0	10
1 if Risk-averse	0.39	0.49	0	1	0.43	0.50	0	1
1 if Male	0.51	0.50	0	1	0.47	0.50	0	1
1 if Originates from West Germany	0.78	0.42	0	1	0.73	0.44	0	1
1 if German national	0.89	0.32	0	1	0.89	0.31	0	1
Degree (Intermediate degree)								
1 if Low degree or missing	0.09	0.28	0	1	0.17	0.38	0	1
1 if Tertiary degree	0.30	0.46	0	1	0.22	0.41	0	1
Age at first job	23.58	3.15	19	32	23.31	3.05	18	32
Job satisfaction at first job	7.42	1.94	1	10	7.24	1.98	0	10
1 if Public sector at firt job	0.27	0.44	0	1	0.24	0.43	0	1
1 if Permanent contract at first job	0.59	0.49	0	1	0.50	0.50	0	1
1 if Part-time emp. at first job	0.09	0.29	0	1	0.17	0.38	0	1
Tenure at first job	1.58	1.48	0.0	6.6	1.37	1.36	0.0	6.60
Wage variation in occ. at first job	0.51	0.09	0.2	0.7	0.54	0.09	0.19	0.84
Log of monthly wage at first job	7.22	0.44	5.63	8.22	7.00	0.61	4.54	8.40
Observations	280				1370			

Notes: Sample: Labour market entrants (aged 18-32 at the time of entry) during the seven years after labor-market entry. SD stands for standard deviation.
Source: German Socioeconomic Panel Survey (SOEP) 1992-2013.

Table A5.2: Risk attitudes, job change and wage growth: Interaction

	Restrictive Sample		Nonrestrictive Sample	
	(1)	(2)	(3)	(4)
1 if Voluntary job change	0.049*	0.045*	0.088***	0.065***
	(0.026)	(0.024)	(0.021)	(0.021)
1 if Voluntary job change * 1 if Risk-averse	0.011	0.022	0.055	0.050*
	(0.045)	(0.038)	(0.034)	(0.03)
Experience		0.090***		0.087***
		(0.011)		(0.010)
Experience square		-0.003**		-0.020*
		(0.001)		(0.01)
Tenure		0.004		0.018***
		(0.006)		(0.005)
Occupation dummies		√		√
Industry dummies		√		√
Within R-squared	0.006	0.311	0.021	0.230
Overall R-squared	0.002	0.103	0.001	0.196
Observations (person)	165		504	
Observations (person*year)	1155		2608	

Note: Estimation: Fixed Effect Model. Dependent variable: Monthly real wage. Estimation sample contains job changers only. Heteroscedasticity-consistent standard errors are in parentheses. ***, **, * denote significance at the 0.01, 0.05, and 0.10 levels.
Source: German Socioeconomic Panel Survey (SOEP) 1992-2013.