Seasonal Education and Population Count Puzzle in Malawi

Tom Mtenje *
Ministry of Finance, Economic Planning and Development
Malawi
Hisahiro Naito‡
Graduate School of Humanities and Social Sciences
University of Tsukuba, Japan

May 2019

*E-mail:tom.mtenje@gmail.com; Address: University of Tsukuba Tennodai 1-1-1, Tsukuba City Ibaraki Prefecture Japan. A technical officer at the Ministry of Finance and Economic Planning and Development. This research was conducted while Mtenje was affiliated with University of Tsukuba. This articled represent the author's opinion. The ministry of Economic planning and development is not responsible for any opinion expressed in this paper. Mtenje appreciates for the final support from World bank Japan scholarship.

†E-mail:naito@dpipe.tsukuba.ac.jp ;Address: University of Tsukuba Tennodai 1-1-1 Tsukuba City Ibaraki Prefecture Japan

‡I appreciate for comments from participants of Canadian Development Economics Study Groups Session during the Canadian Economic Association Conference at Ottawa 2016 and Montreal 2018. I especially appreciate comments from Hao TongTong and Hiroaki Mori.

Abstract

We find that the years of schooling (educational attainment) in Malawi varies across birth months substantially and consistently, at least over thirty years. Those who were born in the second half of each year have 1.6 years longer of schooling than those who were born in the first half of each year. The difference is substantial given that the average number of years of schooling in Malawi is six years. This pattern is persistent in different times, geographic locations, and different demographic groups.

On the other hand, we find that the number of alive population who were born in the second half of each year is almost 60 percent lower than those who were born in the first period. This suggests that those who were born in the second half were born in worse environment than those who were born in the first, but they perform better in years of schooling.

To explain this contradictory pattern of years of schooling and number of alive population across birth months, we propose a selection mechanism hypothesis that among individuals who were born or who were to be born in the second half of each year, only those who have high innate ability could survive the malnutrition during pregnancy and the periods after birth. This implies that those who were born in the second half of each year and those who are alive now have higher innate ability on average than those who were born in the first half of the year. To prove our hypothesis, we regress each persons years of schooling on his or her parents birth months controlling for each persons birth month and parents education. We show that the number of years of schooling of children whose parents were born in the second half of each year is longer than those of children whose parents were born in the first half of each year even after controlling various demographic characteristics. This result shows that individuals who were born in the last half of each year survived severe malnutrition and have higher innate ability.

In addition, our results suggest that half of each cohort who were born in the 2nd half of each year eventually die when they become adult in Malawi. Since such an unusual pattern of mortality is quite rare, it suggests that intensive research is needed to undercover the possible cause of this pattern of mortality.

1 Introduction

The years of schooling (educational attainment) can vary across birth months due to the nutritional and health condition during pregnancy and the period after birth (Kramer, 2003; Neggers and Goldenberg, 2003). In Malawi, the one of the poorest country in the world, we find that years of schooling varies across birth months substantially and persistently. Those who are born in the second half of each year have 1.6 years longer of schooling than those who are born in the first half of each year. The variation is substantial and sharp. For example, among individuals who are at least 22 years old, the difference in the number of years of schooling between those who were born in December of a particular year and those who were born after one month (those who were born in January in the next calendar year) is on average 1.6 years. This 1.6 year difference is quite substantial given that the average years of schooling in Malawi is six years (Figures 1, 2, and 3). This pattern is distinct at least for thirty years regardless of gender, urban-rural location, region, drought or non-drought districts, and religion[1]

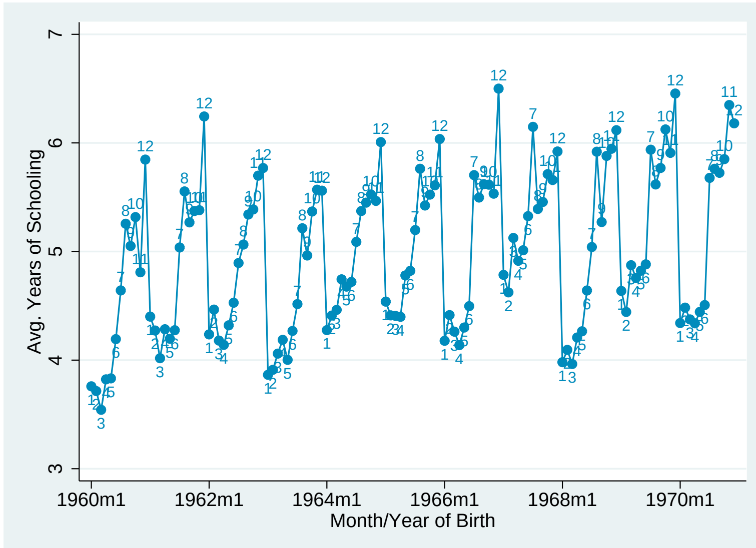
There are several possible mechanisms that could generate such sharp variation in the number of years of schooling across birth months.

As we mentioned above, one possible mechanism is the effect of nutrition or Malaria infection of pregnant mother on brain development. In developing countries, seasonal variation of food production generates seasonal nutritional intake due to the inability to smooth consumption. The resulting periodical insufficient nutrition intake affects the long term outcome of individuals such as mortality, cognitive ability and physical ability through two channels. First, malnutrition during critical stages of pregnancy increases the likelihood of prematurity and intrauterine growth retardation and such events can affect the long-term outcome (Kramer, 2003; Neggers and Goldenberg, 2003)[2]. Second, infant nutrition intake immediately after birth might be a factor to generate such a variation of years of schooling. Many studies show that the nutrition immediately after the birth affect the development in later periods(Alderman et al., 2006; Almond and

[1]We conducted an extensive literature survey examining whether there is a study that found a similar pattern of years of schooling in Malawi. To the best of our knowledge, this study is the first to find this irregular pattern in Malawi.

[2]There are numerous studies on this issue. For example, see Rayco-Solon et al. (2005), Ceesay et al. (1997),Moore et al. (2004) and Verhoeff et al. (2001)

Figure 1: Average Years of Schooling for 1960-1970 Cohorts



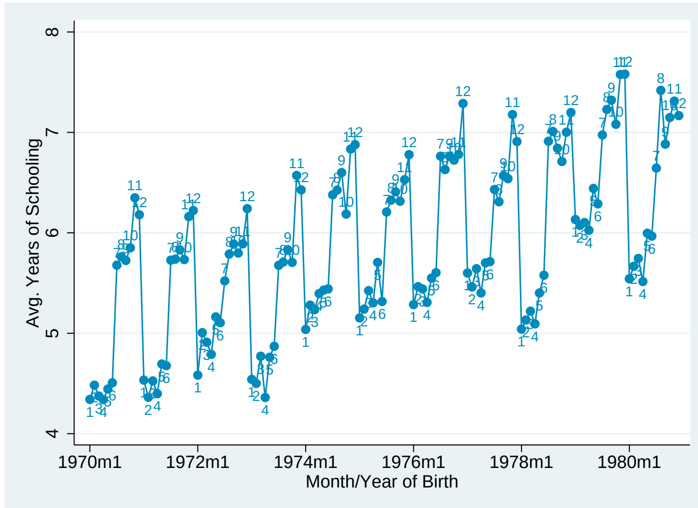
Notes: The source is the MPHC 2008. n=97,392. The years of schooling excludes years in pre-school. For all figures below, years in pre-school is excluded for calculating the years of schooling unless it clarifies.

Currie, 2011; Glewwe et al., 2001). Similarly, the Malaria infection during pregnancy could induce lower birth-weight for new born child, which will affect the development in later period (Guyatt and Snow, 2004).

In our data set on Malawi, the birth weight of newborn babies across birth months is flat and does not show a consistent pattern. Since birth weight should be highly correlated with maternal nutrition or malaria infection during pregnancy, the effect of malnutrition or malaria infection during pregnancy on the brain development is not likely to be a factor to generate the variation of years of schooling across birth months. In addition, the monthly availability of food and malaria infection data are not consistent with the patten of years of schooling.

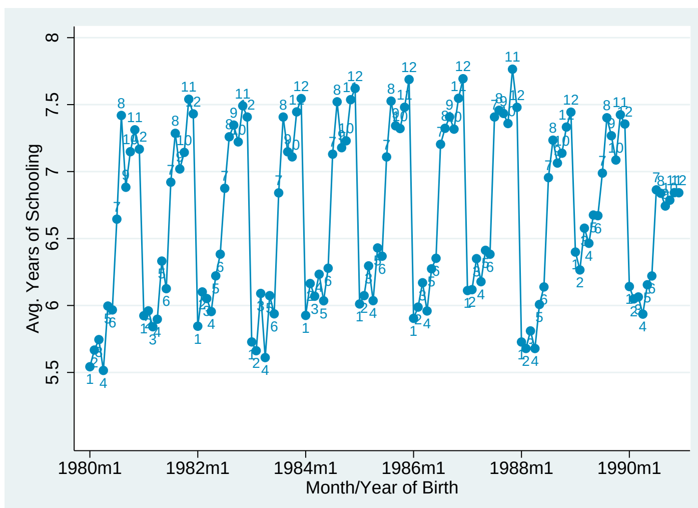
Furthermore, we find a sharp evidence against the hypothesis the malnutrition during pregnancy or malaria infection directly affect the brain development and cause the observed pattern of years of schooling in later periods. More specifically, we find that the number of individuals who were born in the second half of each year (cohort whose years of schooling are longer) and who were alive in the 2008 census is almost 50 percent lower than the number of individuals who were born in the first half and who

Figure 2: Average Years of Schooling for 1970-1980 Cohorts



Notes: The source is the MPHC 2008. n=181,844 .

Figure 3: Average Years of Schooling for 1980-1990 Cohorts



Notes: The source is the MPHC 2008. n=264,957.

are alive in the same data set. If the nutritional or health environment is good to those who were born in the second half of each year, we should observe more population who were born in the second half of each year. However, the pattern of the number of population across birth month is not consistent with this hypothesis.

One might argue that academic rules might generate such an pattern. Angrist and

Krueger (1991) shows that years of schooling varies substantially depending on the quarter of birth. If some students are allowed to exit from school early depending on the real age, not school age, it can be possible that years of schooling is associated with birth month. In addition, students who were born just before the starting month of the academic schedule might perform less successfully than those who were born just after the starting month of the academic schedule especially in the early stage of schooling. If the performance of early stage of schooling affects the outcome of schooling in later year, it is possible that the birth month can affect the years of schooling. Interestingly, in Malawi the government changed the start month of academic schedule in two times, in 1994 and 2009. However, as we demonstrate in section 2.1, the pattern of the variation of years of schooling across birth month did not change even after the change of a starting month of academic schedule. This suggests that the cause of the variation of years of schooling is not likely due to the academic schedule.

To explain these discrepancies among the variation in the number of years of schooling, birth weight across birth months and the variation of the number of population across birth months, we propose the hypothesis that a selection mechanism generates the variation in the number of years of schooling across birth months.

Recently, in literature on economics and demography, the selection mechanism has gained attention, and several cases relevant to the selection mechanism were found. In African countries, the average height is negatively correlated with income per capita while in middle-high income countries, the average height is positively correlated with income per capita. Deaton (2007) and Bozzoli et al. (2009) argue that this is because of the selection effect. In extremely low income countries, when income goes up, individuals who would otherwise die in early childhood or in utero start to survive. As a result, the average height of persons of such countries starts to decrease. During the Great Chinese Famine in which three million people died because of hunger, the height of those who experienced the famine (treated cohort) is as tall as those who did not experience the famine (untreated cohort). Gørgens et al. (2012) shows that the mechanism in which the height of the treated-cohort is almost as tall as the untreated cohort is through the selection effect. In the Great Chinese Famine, only those who had inherently high ability could survive. As a result, the observed height of the treated cohort is as tall as the untreated cohort.

Economists and demographers started to be interested in the selection mechanism because of its implication for policy and research design. If the selection mechanism exists, it implies that a change of outcome because of policy intervention can even go in the opposite direction. For example, consider the effect of policy intervention to increase the nutrition intake of pregnant women in extremely low income countries such as countries in Sub-Saharan Africa. The presence of the selection effect implies that when the government starts such policy intervention, the observed outcome might not be improved because of the selection effect. However, it does not mean that it does not improve the outcome. Second, for designing research, the presence of the selection mechanism implies that the randomness of policy intervention is not sufficient. The researcher needs to control for the effect of the selection.

In our hypothesis, children born during the second half of each year experience malnutrition during pregnancy or in the period after birth. Only those who have good innate ability can survive. As a result, those who are born in the second half of each year demonstrate a higher level of education.

To show empirically that the selection mechanism works to explain the variation in the number of years of schooling across birth months in Malawi, we follow an empirical approach that was first used by Gørgens et al. (2012). They argue that if an individual who had good innate ability survived the Great Chinese Famine more successfully, then the children of those individuals who survived should be taller than children of other individuals because they tend to have good innate ability. He shows that, in the regression analysis, children are taller if their parents experienced the Great Chinese Famine even after controlling for parents education and childrens household characteristics. In our study, by applying the same logic, we show that children of parents who experienced malnutrition have longer years of schooling.

The reader might think that such a huge variation in the population across birth month must be caused by an error in the data collection process of the census. To check such a possibility, we examine not only the census but also the Demographic Health and Integrated Household Surveys in Malawi. All the data sets exhibit the same pattern.

In addition, the reader might argue that such a huge variation in population across birth month is unrealistic and is caused by the illiteracy of parents. For example, some parents might not be able to read and count. As a result, they might simply report that their children are born in January or February. Because children whose parents are illiterate tend to have lower levels of years of schooling, those who were reported to be born in the first half of each year might have lower years of schooling.

To show that this explanation does not apply to Malawi, we show that the difference in the numbers between those who are born in the first half and second half of year is not initially big and even has an opposite pattern. However, as time passes, the difference of the number of alive population between the first half and the second half becomes larger.

Our study has several important policy implications. To the best of our knowledge, this is the first study which has found that the variation in the number of years of schooling and the alive population across birth month varies substantially and is persistent over at least thirty years in Malawi. We conducted an extensive literature survey. However, the fact that the number of live population born in December is 60 percent lower than the number of live population born in January despite the fact that conception occurs evenly across months seems to be unknown even to researchers. This implies that there is substantial child and infant mortality of those born in the second half of the year. Serious discussion on what policy should be implemented is needed.

Our study contributes to the existing literature in several ways. First, in the previous census before 2008 in Malawi, birth month information was not collected. Thus, our finding is the result of the availability of birth month information in Malawis most recent census. By looking at several other survey data sets, we confirm this observation. We also establish that this variation holds regardless of gender, region, location, and religion. Second, most importantly, we show that other mechanisms are not likely to explain the pattern of variation in schooling and alive population across birth month and that the selection mechanism is Malawi. The fact that the selection mechanism works in a dimension other than height is important. In the literature, researchers are interested in height and income not because height itself is important but because height is a good indicator of individual health. We present that the selection mechanism existing in the variation in the number of years of schooling has an important policy implication as we discuss in Section 5.

The remainder of this paper is organized as follows. In Section 2.1, we look at the variation in the number of years of schooling across birth months and its robustness. We show that the pattern holds regardless of gender, urban-rural location, north-south regions, drought-non-drought districts, or religions (muslim or non-muslim). Then, we explore whether the compulsory educational law causes this variation. We also examine whether family characteristics can explain this variation. In Section 2.2, we look at the pattern of hunger and food prices across months using household expenditure surveys and the food price index in Malawi. In section 2.3, we propose our hypothesis. In section 2.4, we show our regression results. Section 3 provides a summary of our analysis and its implications.

2 Institutional background and data set

2.1 Malawi's Economic and Educational Situation

Malawi has the lowest income per capita in the world with the GDP per capita being only USD 320 as of 2013. The fact that Malawis income per capita is the lowest in the world suggests that the selection mechanism discussed in Deaton (2007), Bozzoli et al. (2009) and Gørgens et al. (2012) is more likely to exist because the selection mechanism tends to appear when malnutrition is quite severe. The population of Malawi is 13 million as of 2008. Fifty percent of the population is considered poor (World Bank, 2014). Education outcomes in Malawi are also very weak. Although gross primary enrolment is very good at 115 percent (Ministry of Education, Science and Technology, 2008), high repetition and dropout rates result in only 35 percent of pupils completing primary education and 14 percent completing secondary education(Brossard, 2010).

2.2 Data sets

We use the Integrated Public Use Microdata Series version of Malawi Population and Housing Census (IPUMS-MPHC) in 2008(Minnesota Population Center, 2013). IPUMS-MPHC 2008 is a 10 percent sample of the original Malawian Census in 2008. It collects the basic demographic characteristics such as birth year, birth month, gender, years of schooling, current school attendance, place of residence, dwelling, and family composition. In IPUMS-MPHC 2008, data on 1,343,078 individuals are available. Because of its size and its sampling structure, IPUMS-MPHC 2008 is the main source of our analysis.

The second data set is the Demographic Health Survey (DHS) 2000, 2004, and 2009. The DHS collects basic demographic and health data for a nationally representative sample of all households in Malawi. The DHS collects information on age, gender, residence, years of schooling, school attendance, and other demographic characteristics for each household member. In sampled households, all women aged 15-49 years are individually interviewed. However, among a third of these households, men who are aged 15-54 years are individually interviewed. From individual women interviewed, the DHS collects information on all interviewees children including month and year of birth and mortality status. It also includes a marker for the childs parents record if they live in the same household. For children born to the female respondent in the five years prior to the date of interview, the DHS collects birth weight data. The DHS also collects fertility data for each female respondent for the five years prior to the date of interview. This data consists of every pregnancy in the last five years, the term of the pregnancy and whether the pregnancy ended in a termination or a live birth. Children are matched with their parents from this data.

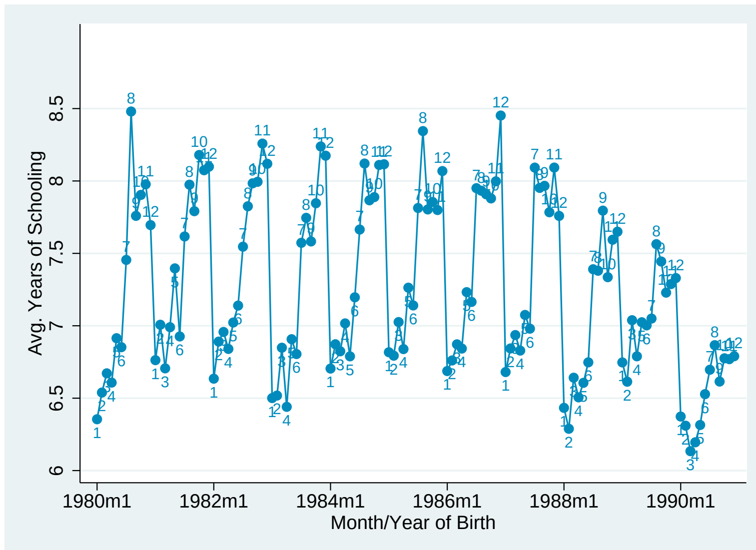
3 Analysis

3.1 Variation of Years of Schooling across Birth Months

Figure 1,2 and 3 show the average years of schooling in each birth month and year in the last thirty years in IPUMS-MPHC 2008. Those who were born in the second half of each year have longer yeas of education than those who were born in the fist half of each year. Those three figures show that the variation of years of schooling across birth months is distinct and consistent at least over thirty years. The difference in years of schooling between those two groups is approximately 1.7 years. This is quite substantial considering that the mean of the average years of schooling is only about six years in Malawi. Figure 4 show years of schooling for male individuals and female individuals across birth month and year. Figure 4 shows that both male and female individuals show the same seasonal variation of the years of schooling. Figure 5 show the seasonal variation of years of schooling for urban residents and rural residents.

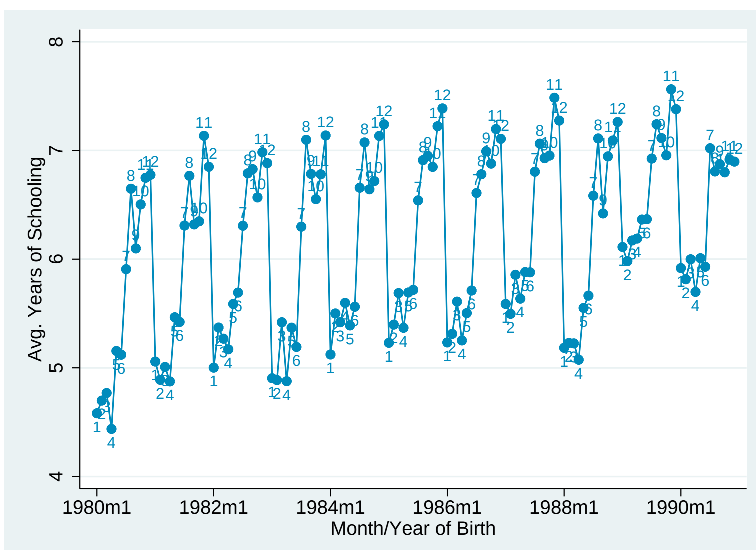
Again, both urban residents and rural residents show the same seasonal variation of the years of schooling.

Figure 4: Average Years of Schooling for Male 1980-1990 Cohorts



Notes: The source is the MPHC 2008. n=123,773.

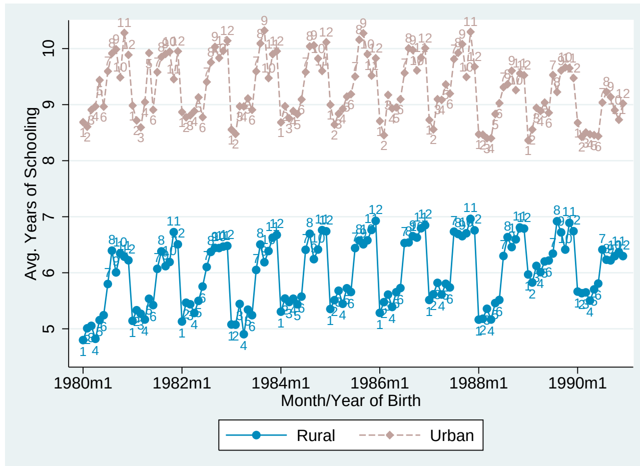
Figure 5: Average Years of Schooling for Female 1980-1990 Cohorts



Notes: The source is the MPHC 2008. n=141,184 .

In Malawi, the country can be divided roughly into three regions. Also, we can

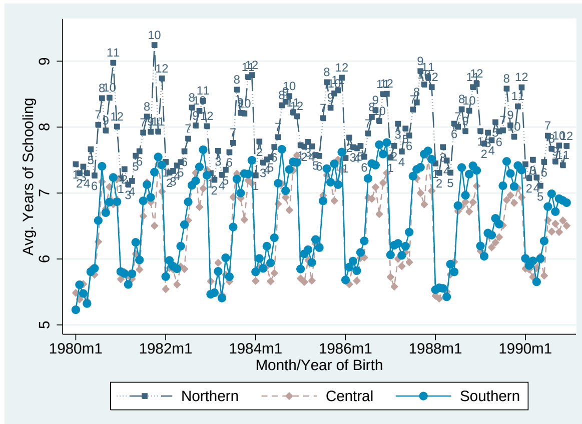
Figure 6: Average Years of Schooling for Urban-Rural 1980-1990 Cohorts



Notes: The source is the MPHC 2008. n=264,957

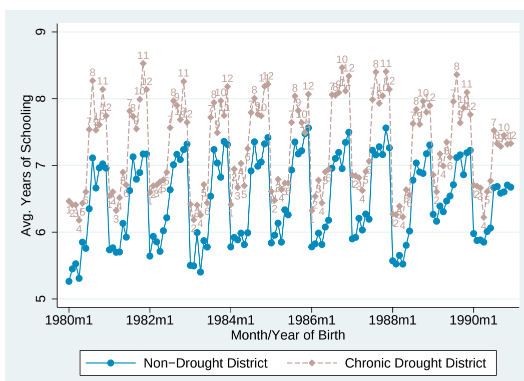
categorize all districts into non-drought district and drought district. One might wonder some region-district specific factors such as weather or drought cause the seasonal variation of years of schooling in Malawi. Figure 7 shows seasonal variation of years of schooling for resident of northern area, central area and southern area. Although the average years of schooling are different in three regions, all three regions exhibits the same patten of the variation of years of schooling across birth months. Figure 8 shows the variation of years of schooling across birth month for drought and non-drought districts. Figure 8 shows that drought and non-drought districts show the same pattern of years of schooling across birth months.

Figure 7: Average Years of Schooling for Northern, Central and Southern Region 1980-1990 Cohorts



Notes: The source is the MPHC 2008. n=264,957.

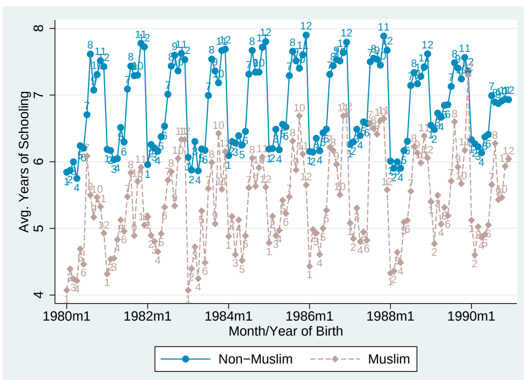
Figure 8: Average Years of Schooling for Drought and Non-drought Districts 1980-1990 Cohorts



Notes: The source is the MPHC 2008. n=264,957. Drought prone districts are Zomba, Chiradzulu, Blantyre, Mwanza, Phalombe, Chikwawa, Nsanje, Balaka, Neno.

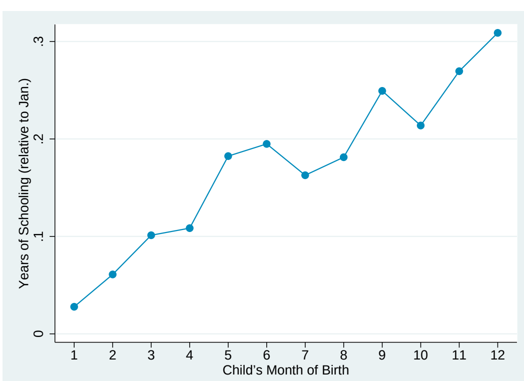
In Malawi, 15 percent of the population is Muslim.[3] As Almond and Mazumder. (2011) show, the practice of Ramadan can affect the nutritional intake of pregnant mothers and the development of cognitive ability of children. Thus, it is possible that the presence of muslim population generates the variation of years of schooling across birth months. Figure 9 shows the seasonal variation of years of schooling of muslim and non-muslim. The Figure 9 shows that the seasonal variation of years of schooling becomes more distinct among non-muslim individuals. This suggest that the seasonal variation of years of schooling is not likely to come from practising ramadan.

Figure 9: Average Years of Schooling for 1980-1990 Muslim and Non-muslim Cohorts



Notes: The source is the MPHC 2008. n=264,957.

Figure 10: Years of Schooling over Birth Months of Children aged 6-18



Notes: The sample is restricted to individuals aged from age 6-18 living with parents.

[3] In 10 percent census, the percentage of muslim is 14.57%

In Figure 10, we show the years of schooling across birth month of children aged 6-18. The Figure shows the same pattern although the difference is relatively small due to the fact that the difference of years of schooling cannot become big for small grade children.

3.2 Institutional Causes and the Effect of Parents

Data Mishandling

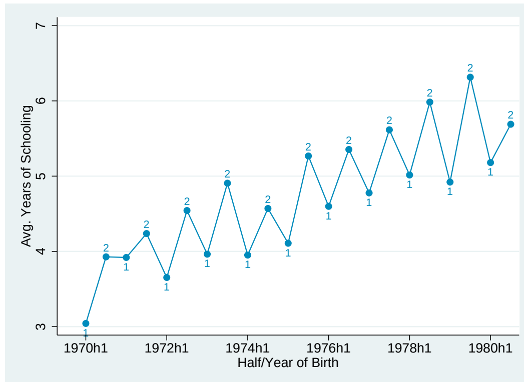
In developing countries, the data collection is not so accurate as in developed countries. One might argue that the seasonal variation of years of schooling MPHC is due to errors to organize answers in the census data set at the government agency of Malawi. To check such possibilities, we examine the DHS data sets. The Figure 11, 12 and 13 show that the seasonal variation of years of schooling in DHS 2000, 2004, and 2010. Since the sample size of DHS data set is only one fortieth to one tenth of the sample size of the census data set, we aggregate years of schooling for those who were born in the first and the last half of each year. Figure 11, 12 and 13 show that those data sets exhibit the same seasonal pattern of years of schooling across birth months as the pattern in the census data set, except 1975 cohort in DHS 2004. On the other hand, in DHS 2000 and DHS 2010, 1975 cohort shows the same seasonal pattern of years of schooling as the pattern in the census. Thus, we can reasonably conclude that the irregular pattern in 1975 in DHS 2004 is due to the relatively small sample size of DHS 2004 compared with the sample size of DHS 2000 and DHS 2010.

The fact that DHS data sets show the same seasonal variation of years of schooling across birth months as the seasonal variation in the census data set suggests that it is not likely to be caused by an error in computer program to scan the answer sheet of the census or mishandling of the data set by the data collection agency.

Compulsory Education Law

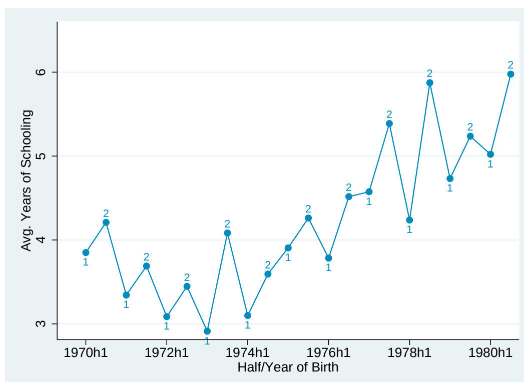
In one of the most cited papers regarding the effect of years of schooling on earnings, Angrist and Krueger (1991) argue that in the United States, the compulsory educational law induces the variation of years of schooling across birth months. This is because if a person reaches a certain age, they are exempted from the compulsory ed-

Figure 11: Average Years of Schooling for 1970-1980 Cohorts in DHS 2000



Notes: The source is the DHS 2000. n=6,134.

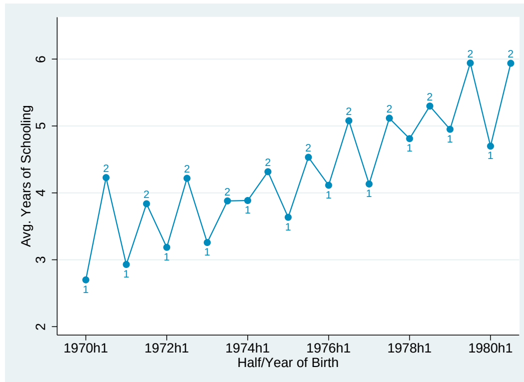
Figure 12: Average Years of Schooling for 1970-1980 Cohorts in DHS 2004



Notes: The source is the DHS 2004. n=3,675.

ucational law. Readers might think the same mechanism apply to the case of Malawi. However, this is not likely to be the case for three reasons. First, the difference of years of schooling between those who were born in the first half and the second half is more than one year. If the variation of years of schooling across birth months is caused by the compulsory education law, it cannot be more than one year. Second, even though compulsory schooling is stipulated in the constitution, this policy is non-binding largely due to supply-side constraints. There are simply not enough schools to accommodate all school age children were the policy to be enforced. Even if parents do not send their children to school, they will not be penalized. Third, interestingly, the academic calendar has changed in 1994. Before 1994, the school calendar started

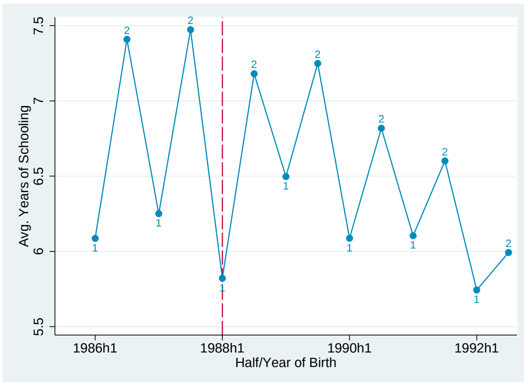
Figure 13: Average Years of Schooling for 1970-1980 Cohorts in DHS 2010



Notes: The source is the DHS 2010. n=5,369.

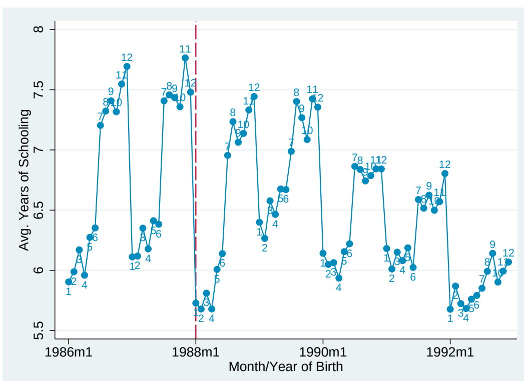
from September. But after 1994, the school calendar started from January. Figure 17 shows the years of schooling of cohorts who were born around 1988. Those are cohort who are close to grade 1 in 1994 when the academic calender changed. Figure 17 shows that the variation of years of schooling across birth month does not change for cohorts who are grade 1 before 1994 and cohorts who are grade 1 after 1994. Figure **??** shows the years of schooling of cohorts who were born around 1982. Those are cohorts who were about grade 6 in 1994 when the academic calender was changed. Figure 17 shows that the variation of the years of schooling does not change between cohorts who were grade 6 before 1994 and cohorts who were grade 6 after 1994. As those two figure shows, although the academic calendar has changed, the seasonal variation of years of schooling has not changed. This suggests that the compulsory education law is not likely to be the source of seasonal variation of years of schooling across birth months.

Figure 14: Average Years of Schooling around 1988 Cohorts (1986-1994 Cohorts)



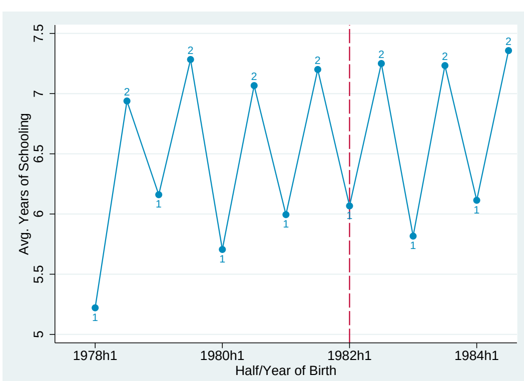
Notes: The source is the MPHC 2008. n=173,780. Cohort 1988 was supposed to be grade 1 in 1994 when the academic calendar was changed.

Figure 15: Average Years of Schooling around 1988 Cohorts(disaggregated)



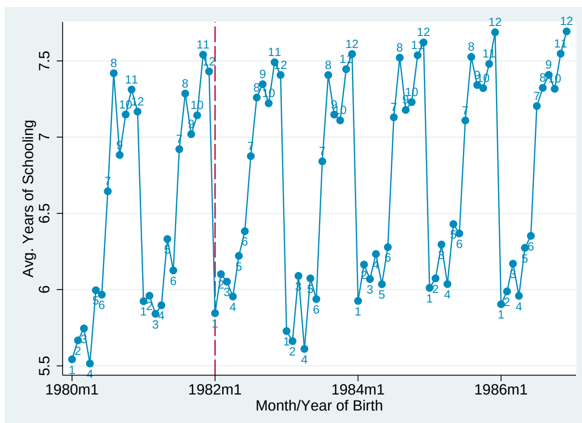
Notes: The source is the MPHC 2008. n=124,314. 1988 Cohort is supposed to be grade 1 in 1994 when the academic calendar was changed.

Figure 16: Average Years of Schooling of around 1982 Cohorts (1978-1984 cohorts)



Notes: The source is the MPHC 2008. n=173,780. Cohort born in 1982 supposed to be grade 6 in 1994 when the academic calendar was changed.

Figure 17: Average Years of Schooling around 1982 Cohorts(disaggregated)



Notes: The source is the MPHC 2008. n=124,314. 1982 Cohort is supposed to be grade 6 in 1994 when the academic calendar was changed.

3.3 Non-Selection Mechanism

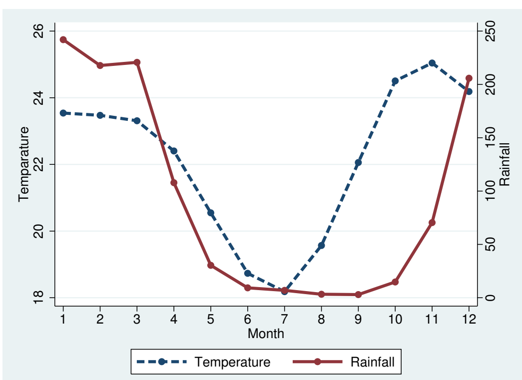
Malaria Infection, Nutrition during Pregnancy and Birth Weight

The fact that the compulsory education law cannot explain a systematic variation of years of schooling across birth months indicates there must be other channels.

One possible channel is the infection to Malaria. Malaria is endemic throughout Malawi and is a leading cause of morbidity and mortality in pregnant women(Ministry of Health of Malawi, 2011). Malaria infection during pregnancy has adverse effects including stillbirth, miscarriage, maternal anaemia and low birth weight(World Health Organization, 2008).

The anopheles mosquito is the primary malaria vector. Vector abundance and transmission follow seasonal rainfall and temperature patterns. Temperature and rainfall patterns in Malawi follow a distinct U-shape pattern (Figure 18). The months from May to August are the coldest months and May to October are the driest ones. The rainy season runs from November to April. October to March are the hottest months.

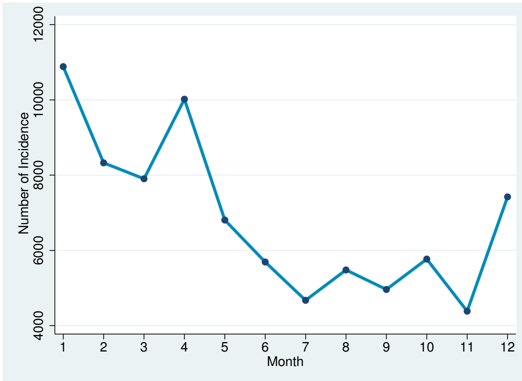
Figure 18: Average Monthly Temperature and Rainfall in Malawi (1901-2018)



Notes: The source of the data is CRU TS 4.03, which is provided by the Climate Research Unit of the University of East Anglia (2019).

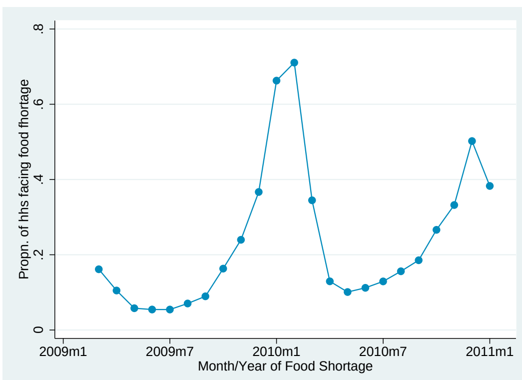
The variation of the malaria incidence is seasonal and consistent with rainfall and temperature patterns (Figure 19). The infections is particularly high during the rainy season, from November through April (Mathanga et al. 2012). The incidence peaks in January. Note that those who conceived at mid January will have an expected birth in the first week of October. However, those who are born in the October have a *higher* years of schooling than those who were born in the first half of each year. Thus, the pattern of malaria incidence looks inconsistent with pattern of years of schooling.

Figure 19: Monthly Malaria Incidence in Malawi (2012-2015)



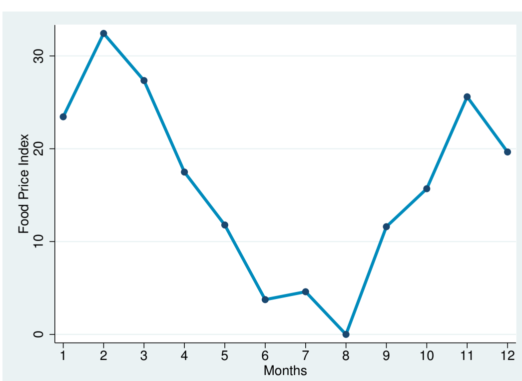
Notes: The source of the data is CRU TS 4.03, which is provided by the Climate Research Unit of the University of East Anglia (2019).

Figure 20: The Experience of Hunger over Months



Notes: The vertical axis measures the percentage of the household who experience the food shortage in a particular month. Source: Third Integrated Household Survey Malawi 2010-11.

Figure 21: Seasonality of General Food Price in Malawi (2007-2018)



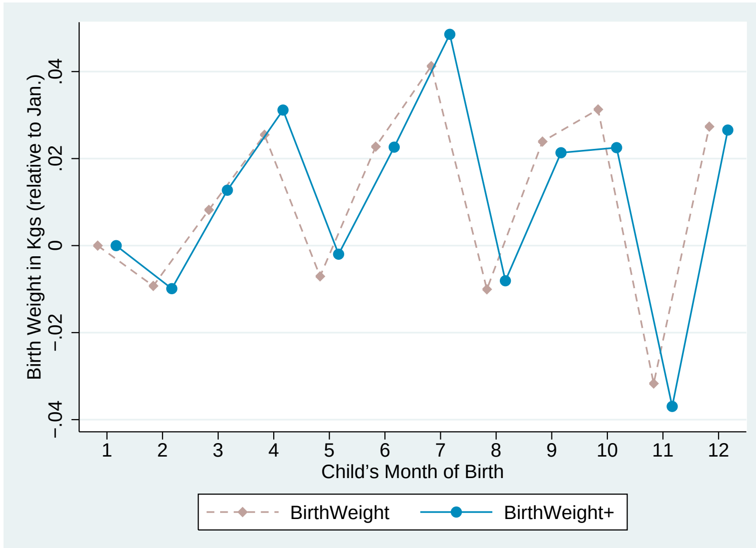
Note: Source is General Food Price Index in Malawi (FAO (2019)).

Another possible channel is the seasonal variation of nutrition intake during pregnant mothers. The literature shows that insufficient nutrition intake during pregnancy generate new born baby with lower birth weight (Kramer, 2003; Neggers and Goldenberg, 2003). A baby with lower birth weight might have less cognitive development and less years of schooling.

With respect to food security, the months from November to February are the lean months, the green harvest is available in February and March and the main harvest period runs from April through July where food is abundant (Fews Net, 2014). The 2010-2011 Malawi Integrated Household Survey asks households whether they experienced any food shortage in the past month prior to the interview date. Figure 20 plots the percentage of households which experience hunger in each month. It shows that the percentage of the household who experience the huger peak in January. Figure 21 shows the average monthly food price from 2007-2018. It shows that the food price peaks in February. Thus, both the survey data and food price data shows that January and February is the months where nutrition is least available. On the other hand, those who conceived in the January and February will have an expected birth on October and November. This is inconsistent with the observed pattern of years of schooling that those who were born in the second half of each year have longer years of schooling.

To support our argument further, we also examine the birth weight of new born children over birth months. If the infection to Malaria or malnutrition during pregnancy is the cause of generating the variation of years of schooling across birth months, then it should be reflected on birth weight across birth months. More specifically, we should see lower birth weight for babies who were born in the first half of any year and higher birth weight for those who were born in the last half of any year. In Figure 22, we plot the birth weight across birth months. In Figure 22, we do not find a pattern in birth weight that is consistent with the variation of years of schooling. In the regression, the only the coefficient in November is significant. However, the size of the coefficient is very small and it is opposite sign. This result suggests that malaria infection and insufficient nutrition intake are not likely to be the cause of the variation of years of schooling in Malawi.

Figure 22: Birth Weight over Birth Months



Notes: The graph shows the coefficient of birth month dummies in two different regressions. BirthWeight includes only birth month dummies in the regression. BirthWeight+ controls gender, mother's education, age, location in the regression. n=18,533. The source is the DHS 2000, 2004, 2010.

3.4 Selection Mechanism Hypothesis

Hypothesis

In the above sub-sections, we have explored several mechanisms that can explain the variation of years of schooling across birth months: compulsory educational law, nutrition during pregnancy, malaria infection, nutrition after birth. However, we find that none is not consistent with the variation of years of schooling across birth months.

To explain the variation of years of schooling across birth months, we now hypothesize that the selection mechanism exists and it generates the variation of years of schooling across birth months in Malawi.

When a egg is conceived during October and March, this conceived egg, embryo and fetus will experience severe conditions such as mal-nutrition and malaria infection. Thus, a conceived egg, embryo or fetus can be damaged during this period and a damaged cell can cause pregnancy termination and early death after birth. For such a damage, only those with higher innate ability can survive. Higher innate ability offset the damage caused by malnutrition and malaria infection. Individuals that survive hardship during pregnancy are positively selected. They have better genetic traits. They are stronger and have a higher innate ability which would explain the positive association between years of schooling and the negative environment such as constrained maternal nutrition intake and malaria infection and years of schooling.

Empirical Strategy

To prove that the selection mechanism is working to explain the variation of years of schooling in Malawi, we provide several evidences.

First, we show that the number of the population who were born in the second half and who are alive now, which show a longer year of schooling, is 50 % lower than the population who were born in the first half of each year.

For the second evidence, we examine the correlation of birth month of parents and children. If a parent who was born in the second half of each year has a higher innate ability that a parent who was born in the first half of each year, then children of a parent who were born in the second half of each year must have a higher innate ability on average. Such children can survive ill environment more easily than children who have lower innate ability. Note that children who were supposed to be born in the second half of each year experience more severe environment during the pregnancy. But children with higher innate ability can survie ill environment more than children with less innate ability. This implies that when a mother was born in the second half of each year, then a conceived egg, embryo and fetus of such a mother has a higher probability of surviving ill condition. Thus, the ratio of those who are born in the second half of each year from such a mother become higher than from other mother who were born in the first half of each year.

For the third evidence, we use the empirical strategy that is used by Gørgens et al. (2012). In the case of the Great Chinese Famine, Gørgens et al. (2012) regressed the height of each individual on individual demographic characteristics and individual parent's treatment status dummy where parent's treatment status dummy indicates whether parent experienced the Great Chinese Famine while he or she is mother's uterus. The idea of the Gørgens et al. (2012) is that if parent survives the Great Chinese Famine, then Parent must have good inherent characteristics (height). As a result, holding the environment condition constant, their children must be taller than the children of parents who did not experience the Great Chinese Famine. In our direct evidence of the existence of the selection mechanism, we use a similar idea. We speculate that if the selection mechanism is working to explain the variation of years of schooling in Malawi, the individual who were born in the latter six months of each year must have high innate ability. This implies that, holding other conditions constant, children of those individual must have high innate ability than other children who do not have parent who were born in the latter six months of each year. Thus, as the direct evidence, we run the following regression:

$$E_{ni} = \beta_f BM^f_{tni} + \beta_m BM^m_{tni} + \gamma BM_{ni} + \alpha X_{tni} + \varepsilon_{tni} \tag{1}$$

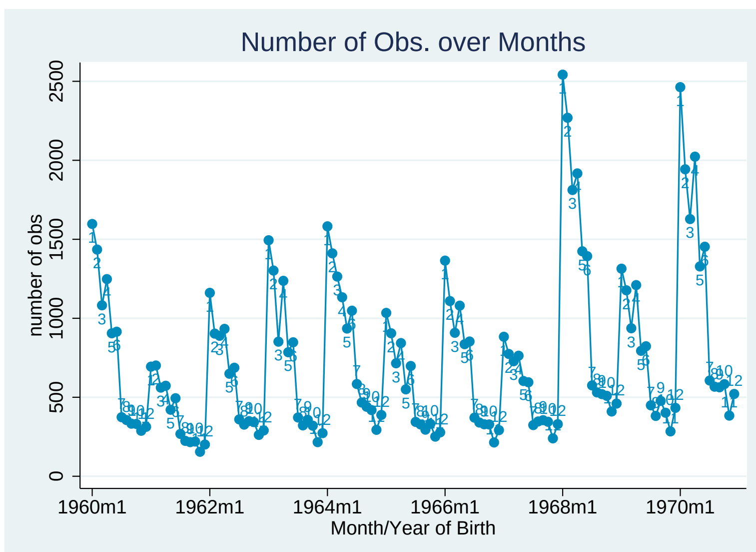
where i is the index of individual,,n the index of birth month of individual i. E_{ni} is years of schooling of individual i born in month n. BM^f_{ni} is a vector of dummy variable indicating birth month of father of individual i who were born in month n. BM^m_{ni} is a vector of dummy variable indicating birth month of mother of individual i who were born in and month n. BM_{ni} is the vector of birth month dummy of individual i. X_{ni} is the vector of demographic characteristics of individual i which include parents' education, grade for age of individual i. ϵ_{ni} is the error term. The coefficients of our interest are β_f ,β_m and γ. β_f and β_m show that how parents' birth month affects children's year of schooling even after controlling the education of parents. If parent's birth in the second half of each year is correlated with children's years of schooling positively even after controlling parents' education and occupation, it indicates that the selection mechanism is working. By comparing γ with or without parents' birth month, we can infer to what extent the selection mechanism explain the variation of years of schooling across birth months.

Results

Figures 23–26 show the number of observation across birth months in the census 2008 dataset. It is one of the clearest evidence that the selection is happening in the data set. Figures 23–26 show that the number of individuals who were born in the second half of each year (cohort whose years of schooling are longer) and who are alive is almost 50 percent lower than the number of individuals who were born in the first half and who are alive. Figure 27– shows a similar patten in the DHS dataset and the Integrated Household Survey. Figure 30 shows that the pattern of conception across months is almost flat. In Table 2, we examine whether conception might is correlated with mother's schooling or mother's birth month. We find that mother's years of schooling and mother's birth month are not correlated with the conception month of children.

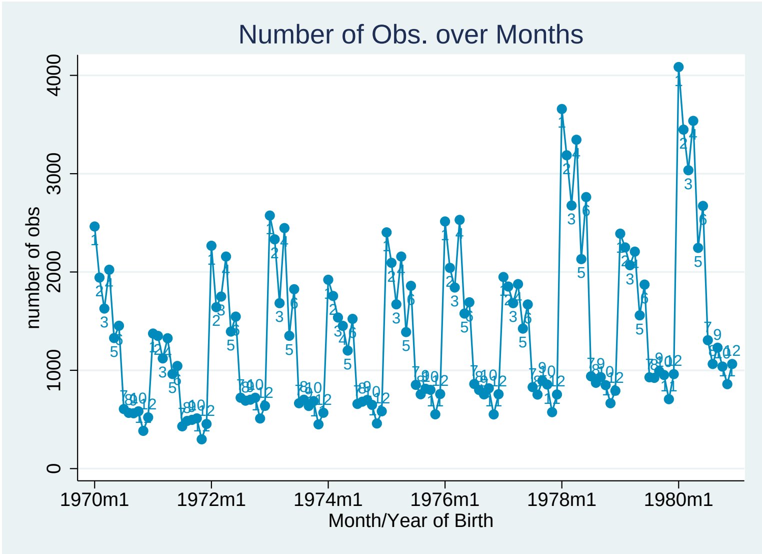
On the other hand, Figure 30 shows that those who are expected to be born in the second half of each year have a higher pregnancy termination rate. Thus, Figures 23, 24 25 and 30 strongly suggest that those who were born in the second half of each year experience the severe selection.

Figure 23: The number of alive population across birth month:1960 to 1970



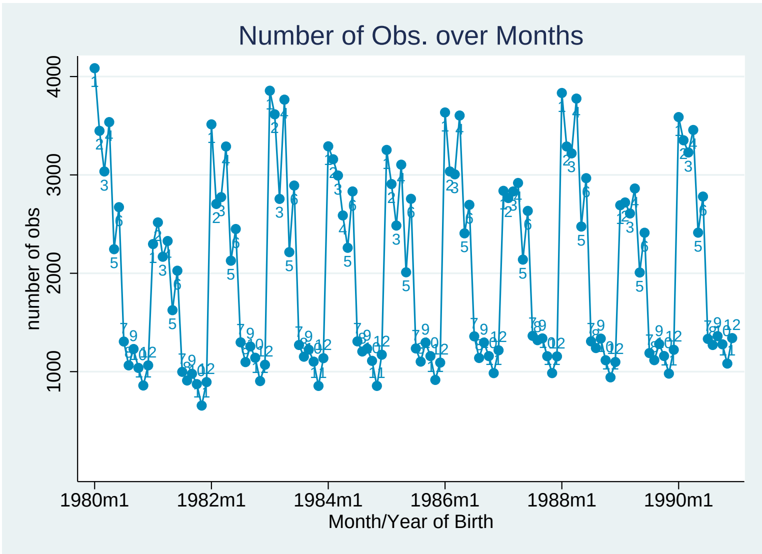
Note: The data source is MPHC 2008.

Figure 24: The number of alive population across birth month:1970 to 1980



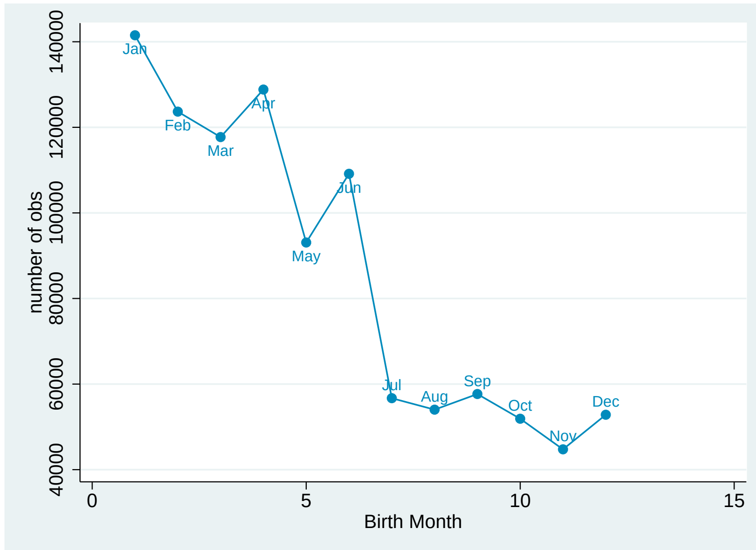
Note: The data source is MPHC 2008.

Figure 25: The number of alive population across birth month:1980-19980



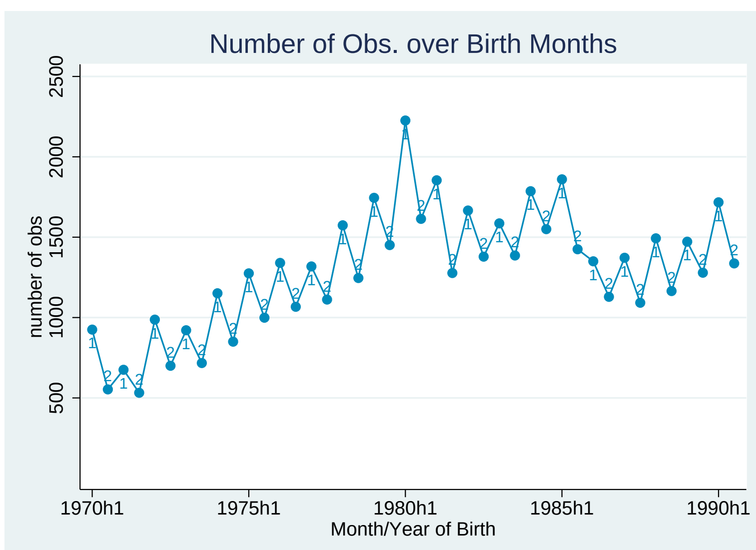
Note: The data source is MPHC 2008.

Figure 26: The number of population across birth month in the census



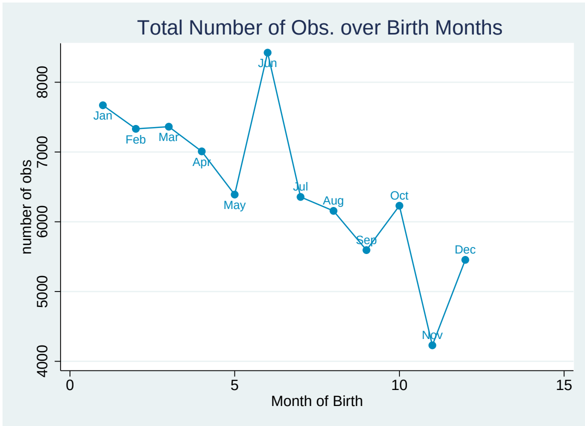
Notes: The data source is MPHC 2008. The census was conducted in June 2008.

Figure 27: The number of population across birth month and year, aggregated in the first and 2nd half of each year(DHS data)



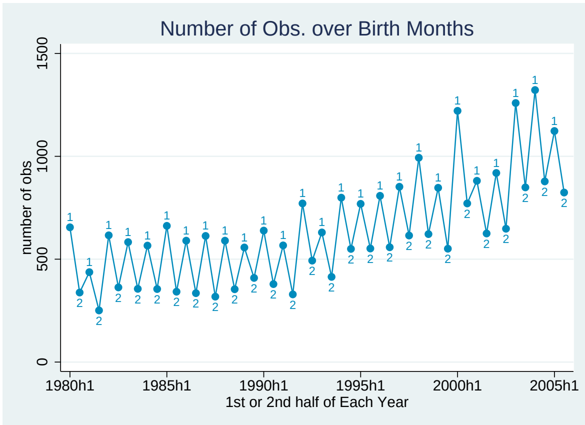
Notes: The data source is DHS 2000, 20004, 2010.

Figure 28: The number of population across birth months: DHS data



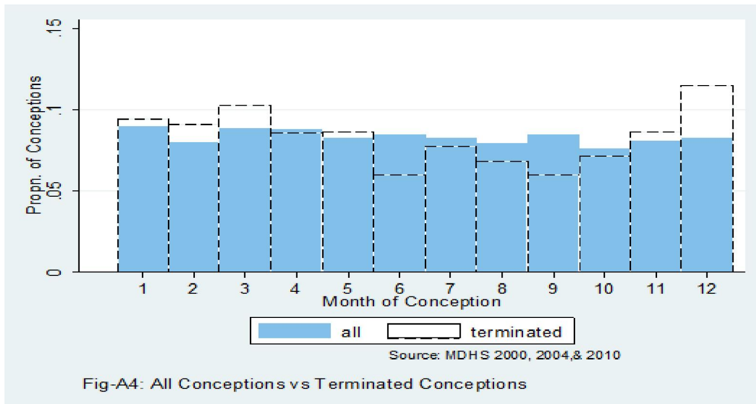
Notes: The data source is DHS 2000, 20004, 2010.

Figure 29: The number of observation across birth month and year, aggregated in the first half and 2nd half of each year(Integrated Household Survey)



Notes: The data source is IHS3. 1 indicates that the number of observation born in the first half of the year. 2 indicates the number of observation born in the 2nd half of the year.

Figure 30: The proportion of conception over months and pregnant termination



Notes: The data source is DHS 2000, 20004, 2010.

Table 2. The Effect of Education and Birth Month of Mother on the Conception Month

Dependent Variable	Conception Month of Pregnancy			
	(1)		(2)	
Birth Month of Mother	0.00657	(0.00514)	0.00564	(0.00515)
Years of Schooling of Mother	-2.79e-06	(0.00489)	-0.00726	(0.00548)
Year of birth			0.00634	(0.00388)
Constant	6.409***	(0.0387)	-6.114	(7.801)
Demographic Characteriscs	No		Yes	
N	40,840		40,840	
R-squared	0.000		0.003	

Notes: Robust standard errors in parentheses. The source is DHS 2000, 2004 and 2010. The sample is restricted to women who experienced pregnancy in the last five years of the interview. Demographic characteristics include region dummies, dummies of mother's age (34 dummies) and urban-rural dummy. *** p<0.01, ** p<0.05, * p<0.1

Figure 31 is the second clear evidence which shows that that those who were born in the second half have higher innate ability than those who were born in the first half of each year. Red line shows the ratio of those conceived during October and March. Note that those who are conceived during October and March are supposed to be born in the second half of each year. The red line show that, the ratio of those conceived during October and March is constant across birth month of mother. The blue line show the ratio of those who are born in the second half of each year across mother's birth month. The blue line shows that the ratio of those born in the second half of each year becomes higher if the mother is born in the second half. Since those who are born in the second half of each year experience more hardship in nutrition and Malaria infection, the blue line shows that mothers who were born in the second half of each year have higher innate ability than those who were born in the first half of each year.

Figure 31: Ratio of Those Conceived during October and March and of Those Born in the 2nd half across Mother's Birth Month

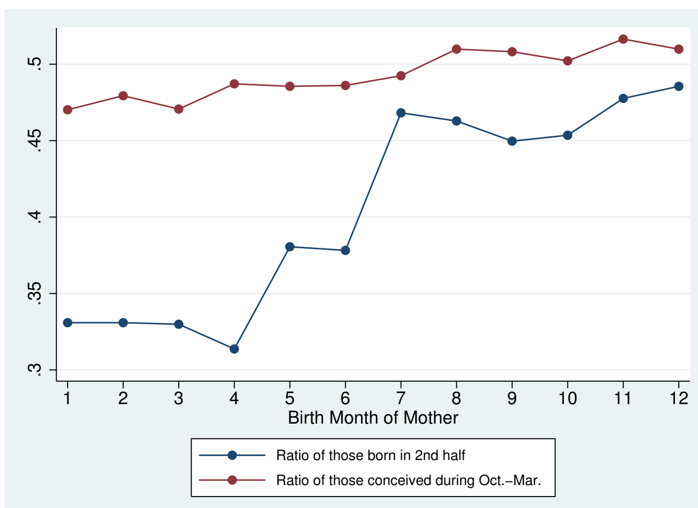


Table 3 shows the summary statistic of the data that we use for the regression analysis. We restrict children of the census sample who are aged from 6 to 18 and who live with both mother and father. In this data set, we have 227,715 observation. The average years of schooling is 3.14 years, average age is 10.8 years and, boys and girls are more or less evenly represented.

Table 3: Summary Statistics

VARIABLES	mean	sd	min	max	N
Years of schooling	6.224	3.021	0	14	17,563
Month of birth	5.708	3.407	1	12	17,563
Age	16.85	0.822	16	18	17,563
Sex	0.539	0.499	0	1	17,563
Years of Scholing of Mother	4.013	3.854	0	18	17,563
Birth Month of Mother	4.932	3.151	1	12	17,563
Age of Mother	36.45	4.978	16	48	17,563
Birth Month of Father	4.585	3.138	1	12	17,563
Region	2.256	0.660	1	3	17,563
Urban-rural status	0.159	0.365	0	1	17,455

Notes: Age is calculated based on academic calendar. The sample is resctriected to children whose age is between 16 and 18 and who live with both parents. The source if MPHC 2008.

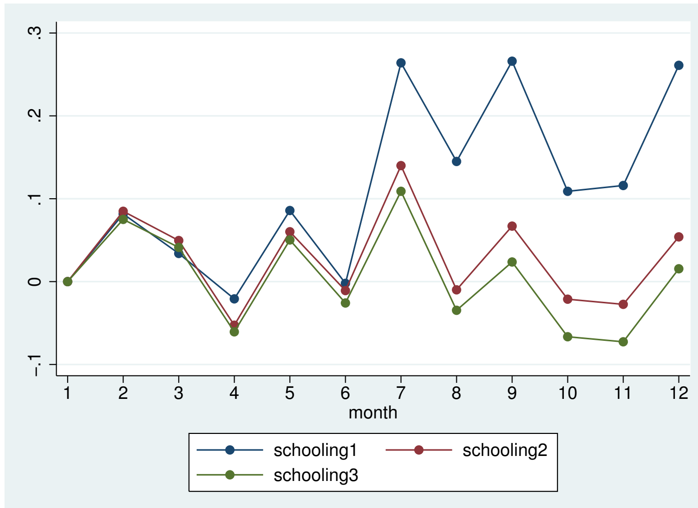
Table 4 shows the result of the main regression. In the first column, child's years of schooling is regressed on parents"s birth month without any covariates. The coefficient shows that as the parent is born in latter months, the children's average years of schooling becomes longer. In the column (2), we control children's birth month and parent education. Controlling parents' education is important since the parent's education level is correlated with parents' birth month and it also affects the child's year of schooling. Also, controlling children's birth month is important since children's birth months affect the child nutrition intake directly. In the column (3), we add children's grade for age and parent's age and time dummy, time dummy interacted with region dummy, region dummy interacted with child's birth month dummy as the additional control. In column (4), we add the region and urban-rural location as additional covariates. Those all columns shows that the years of schooling is upward sloping regarding parents' birth month and mother's birth month. The regression results implies that mother's birth month affect child's years of schooling substantially even after controlling parents' education and other covariates.

Table 4: The Effect of Birth Months on Child's Years of Schooling

	(1)		(2)		(3)	
Child's Birth month						
Feb	0.0817	(0.0948)	0.0849	(0.0852)	0.0751	(0.0852)
Mar	0.0340	(0.0937)	0.0496	(0.0849)	0.0413	(0.0849)
Apr	-0.0208	(0.0942)	-0.0526	(0.0841)	-0.0606	(0.0840)
May	0.0858	(0.101)	0.0602	(0.0911)	0.0503	(0.0911)
Jun	-0.00208	(0.0963)	-0.0108	(0.0858)	-0.0258	(0.0859)
Jul	0.264**	(0.106)	0.140	(0.0956)	0.109	(0.0957)
Aug	0.145	(0.104)	-0.00989	(0.0927)	-0.0346	(0.0930)
Sep	0.266***	(0.101)	0.0670	(0.0914)	0.0237	(0.0915)
Oct	0.109	(0.103)	-0.0212	(0.0945)	-0.0665	(0.0947)
Nov	0.116	(0.107)	-0.0274	(0.0969)	-0.0726	(0.0973)
Dec	0.261**	(0.102)	0.0540	(0.0921)	0.0155	(0.0924)
Mother's Birth month						
Feb					-0.182**	(0.0764)
Mar					-0.0192	(0.0785)
Apr					-0.130*	(0.0746)
May					-0.0270	(0.0837)
Jun					0.0481	(0.0800)
Jul					0.0443	(0.102)
Aug					0.102	(0.100)
Sep					0.314***	(0.102)
Oct					0.190*	(0.0978)
Nov					0.130	(0.112)
Dec					0.289***	(0.101)
Father's Birth month						
Feb					0.0947	(0.0681)
Mar					0.0820	(0.0744)
Apr					-0.00566	(0.0717)
May					0.0528	(0.0804)
Jun					0.0583	(0.0781)
Jul					0.0338	(0.102)
Aug					0.00319	(0.105)
Sep					0.167	(0.103)
Oct					0.154	(0.0987)
Nov					-0.0159	(0.129)
Dec					0.374***	(0.104)
Parents' education	no		yes		yes	
r2	0.132		0.285		0.288	
N	17,455		17,455		17,455	

Notes: Clustering robust standard error in parentheses. The error term is clustered at month ×year. January serves as the reference month. Sample comprises children between the ages of 16 and 18 living with both their parents. Other control variables are school age dummy, male dummy, urban dummy, region dummy, mother's ageand father's age. Parents' education include father's and mother's years of schooling and their square. The data is Malawi Population Household Census 2009. * p<0.10, ** p<0.05, *** p<0.01.

Figure 32: The Effect of Birth Month on Years of Schooling of Children

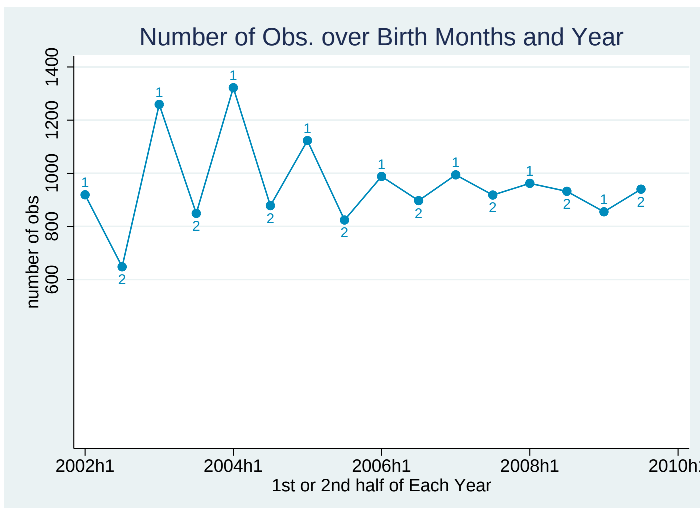


Notes: The figure shows the estimated coefficient of each birth month dummy in different specifications. School 1 controls sex, urban-rural dummy, regions, school age, parents' age. School 2 additionally controls parents' education. School 3 control parents' birth months.

Selection Process

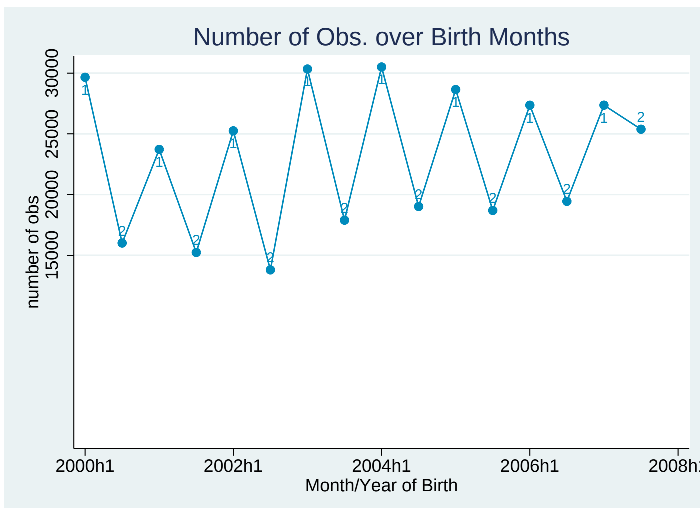
Note that the difference of pregnancy termination rates in the first half and second half of each year is statistically significant but it is not big enough to generate the difference of the population of those who were born in the first half of each year and the second half of each year. This suggests that the death of those who were born in the second half of each year is occurring gradually rather than instantaneously. Figure and 33 and 34 and show the number of individual who were born during a few years before the survey years of the census and Integrated household survey. The graph shows that at the beginning, the number of population who were born in the second half is not so low. But after two years, the number starts to drop substantially.

Figure 33: The Total Number of observation across birth months (Integrated Households Survey)



Notes: The data source is IHS3. The vertical axis measures the number of observation born in each half of the year. IHS3 was conducted in 2010. The graph shows that the difference of the number of the observation between the first and 2nd half increases as time passes.

Figure 34: The number of population across birth month: a few years after birth



Notes: The data source is MPHC 2008. The census was conducted in June 2008. The graph shows that the number of population born in the second half of each year start to drop substantially 2 years after their birth.

Critical Stage

One might think that which stage of the malnutrition is critical given that the death of those who are born in the second half of each year is occurring gradually instead of instantaneously. One might argue that the malnutrition afterbirth is more critical than the malnutrition during pregnancy. The Table 3 examine which stage is critical for the selection and years of schooling. Note that the huge discontinuous change of the number of alive individuals occurs between those who are born in December and January. If the malnutrition just after birth is critical, then it does not explain the sharp discontinuous change from December to January cohort because those who are born in December and January both experience hunger after birth (through breast feeding).

The Table 5 shows that the critical stage is the second month of pregnancy. As the table 5 shows, the second month of pregnancy is the period when all cohorts who shows a longer years of education and lower number of observations experience hunger. It also explain why there is discontinuity between December cohort and January cohorts.

Table 5. The Second Stage of Pregnancy is Critical

calendar	July	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	June	July	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
least nutrition																						
malaria																						
Conception Month																						
July	p1	p2	p3	p4	p5	p6	p7	p8	p9	due												
Aug		p1	p2	p3	p4	p5	p6	p7	p8	p9	due											
Sep			p1	p2	p3	p4	p5	p6	p7	p8	p9	due										
Oct				p1	p2	p3	p4	p5	p6	p7	p8	p9	due									
Nov					p1	p2	p3	p4	p5	p6	p7	p8	p9	due								
Dec						p1	p2	p3	p4	p5	p6	p7	p8	p9	due							
Jan							p1	p2	p3	p4	p5	p6	p7	p8	p9	due						
Feb								p1	p2	p3	p4	p5	p6	p7	p8	p9	due					
Mar									p1	p2	p3	p4	p5	p6	p7	p8	p9	due				
Apr										p1	p2	p3	p4	p5	p6	p7	p8	p9	due			
May											p1	p2	p3	p4	p5	p6	p7	p8	p9	due		
June												p1	p2	p3	p4	p5	p6	p7	p8	p9	due	
July													p1	p2	p3	p4	p5	p6	p7	p8	p9	due

4 Summary and Conclusion

In Malawi, years of schooling varies across birth months substantially and consistently over thirty years. We have explored the possible mechanism to explain this variations We first established that the compulsory educational law and family characteristics do not explain the variation of years of schooling. Second, we shows that the birth weight does not vary across birth months and the patten of food availability is not consistent with the variation of years of schooling. Third, we proposed a hypothesis that a selection mechanism explains the variation of years of schooling across birth months and the variation of food availability across months. To prove that the selection mechanism working, we have provided two evidences. First, we have demonstrated that the number of individuals were born during the second half of each year, the months that exhibits a longer year of schooling, is 50 percent lower than the number of individuals who were born in the first half of each year. This implies that individuals who were born in the last half of each year experience the hardship during pregnancy or after birth.

Second we regress each individual years of schooling on parents' birth month, parents' education and other covariates. We have shown that when parent is born in the last half of each year, the children's years of schooling is longer than the years of schooling of children whose parents were born in the first half of each year. Our regression result shows that individuals who were bon in the last half of each year have higher innate ability than those who were born in the first half of each year.

Our results have several implications. First, our result indicates that when the government improves the nutritional condition , it is quite possible that year of schooling is not improved. This does not mean that the government intervention is not effective. It is the result of relaxing the selection effect. Thus, for evaluating the government intervention program, it is important to control the selection effect. Also in the past, often in developing countries, the inelastic response of the outcome to the policy intervention in the field of education and health is observed. It is possible that those inelastic response might come from the selection mechanism. Second, our analysis shows that encouraging birth during the last half of each year is not appropriate. Although, observationally, the average years of schooling of those who were born in the last half of each year is 1.5 years longer than other individuals, this longer years of schooling is archived through the selection. Thus, encouraging birth during such a period is not appropriate from the point of equity and efficiency. Third, our result have shown that the selection mechanism exists in a dimension other than in height. To the best of our knowledge, our study is the first study that find that the selection mechanism exist in a dimension other than in height.

References

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

Almond, D. and Currie, J. (2011). Killing me softly: The fetal origins hypothesis. *The Journal of Economic Perspectives*, 25(3):153–172.

Almond, D. and Mazumder., B. A. (2011). Health capital and the prenatal environment: the effect of ramadan observance during pregnancy. *American Economic Journal: Applied Economics*, 3(4):56–85.

Angrist, J. D. and Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, 106(4):979–1014,.

Bozzoli, C., Deaton, A., and Quintana-Domeque, C. (2009). Adult height and childhood disease. *Demography*, 46(4):647–669.

Brossard, M. (2010). *The Education System in Malawi: Country Status Report*. World Bank.

Ceesay, S. M., Prentice, A. M., Cole, T. J., Foord, F., Poskitt, E. M., Weaver, L. T., and Whitehead, R. G. (1997). Effects on birth weight and perinatal mortality of maternal dietary supplements in rural gambia: 5 year randomised controlled trial. *Bmj*, 315(7111):786–790.

Deaton, A. (2007). Height, health, and development. *Proceedings of the National Academy of Sciences*, 104(33):13232–13237.

Fews Net (2014). Malawi. *Famine Early Warning System Network*.

Glewwe, P., Jacoby, H. G., and King, E. M. (2001). Early childhood nutrition and academic achievement: a longitudinal analysis. *Journal of Public Economics*, 81(3):345–368.

Gørgens, T., Meng, X., and Vaithianathan, R. (2012). Stunting and selection effects of famine: A case study of the great chinese famine. *Journal of development Economics*, 97(1):99–111.

Guyatt, H. L. and Snow, R. W. (2004). Impact of malaria during pregnancy on low birth weight in sub-saharan africa. *Clinical microbiology reviews*, 17(4):760–769.

Kramer, M. S. (2003). The epidemiology of adverse pregnancy outcomes: an overview. *The Journal of nutrition*, 133(5):1592S–1596S.

Ministry of Education, Science and Technology (2008). National education sector plan 2008 2017.

Ministry of Health of Malawi (2011). Malawi health sector strategic plan 2011-2016.

Minnesota Population Center (2013). *Integrated Public Use Microdata Series, International: Version 6.2 [Machine-readable database].* University of Minnesota, Minneapolis.

Moore, V. M., Davies, M. J., Willson, K. J., Worsley, A., and Robinson, J. S. (2004). Dietary composition of pregnant women is related to size of the baby at birth. *The Journal of nutrition*, 134(7):1820–1826.

Neggers, Y. and Goldenberg, R. L. (2003). Some thoughts on body mass index, micronutrient intakes and pregnancy outcome. *The Journal of nutrition*, 133(5):1737S–1740S.

Rayco-Solon, P., Fulford, A. J., and Prentice, A. M. (2005). Differential effects of seasonality on preterm birth and intrauterine growth restriction in rural africans. *The American journal of clinical nutrition*, 81(1):134–139.

Verhoeff, F. H., Brabin, B. J., Van Buuren, S., Chimsuku, L., Kazembe, P., Wit, J., and Broadhead, R. L. (2001). An analysis of intra-uterine growth retardation in rural malawi. *European Journal of Clinical Nutrition*, 55(8):682.

World Bank (2014). Malawi overview. *World Bank*.

World Health Organization (2008). Technical expert group meeting on intermittent preventive treatment in pregnancy (iptp), geneva, 11-13 july 2007.