Birth Month Education Population Puzzle and A Selection Hypothesis

Tom Mtenje *

Ministry of Finance, Economic Planning and Development

Malawi

Hisahiro Naito‡

Graduate School of Humanities and Social Sciences

University of Tsukuba, Japan

October 3, 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 the University of Tsukuba. This article represents the author's opinion. The ministry of Economic planning and development is not responsible for any opinion expressed within this paper. Mtenje appreciates the support of the Joint Japan World Bank Graduate Scholarship.

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

‡I appreciate comments from participants of the Japanese Economic Association Meeting, Canadian Development Economics Study Groups Session (during the Canadian Economic Association Conference in Ottawa and Montreal), and seminar participants at Keio University. I especially appreciate comments from Hao TongTong and Hiroaki Mori, Ryo Nakajima, Yoshito Takasaki.

Abstract

We find that the number of the living population born in the second half of each year (who experienced malnutrition and malaria infection in utero during the first two trimesters in the gestation period) in Malawi is almost 50 percent lower than those who were born in the first period.

On the other hand, we find that those who were born in the second half of each year have 1.6 more years 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 demographic groups. We find similar patterns in the literacy rates and probability of accomplishing higher grades. This suggests that those born in the second half of the year experience severe conditions while in utero compared to those who were born in the first half. However, they perform better in educational scenarios.

To explain these puzzling patterns, we propose the following selection hypothesis. When a developing organ, either in utero or within a newborn baby, experiences malnutrition and malaria infection, only individuals with certain genetic traits that are positively correlated with health and educational outcome can survive. Consequently, individuals born in the second half who are alive today have those innate traits, and they have better educational outcomes.

We prove our hypothesis by examining conception pattern, correlation between parents' and children's birth months, and correlation between the degree of selection and educational outcome. We also demonstrate that, due to this selection effect, children of parents who were born in the first half (who did not experience selection pressure) have much higher mortality when they are born in the second half, suggesting this particular groups need for government intervention.

1 Introduction

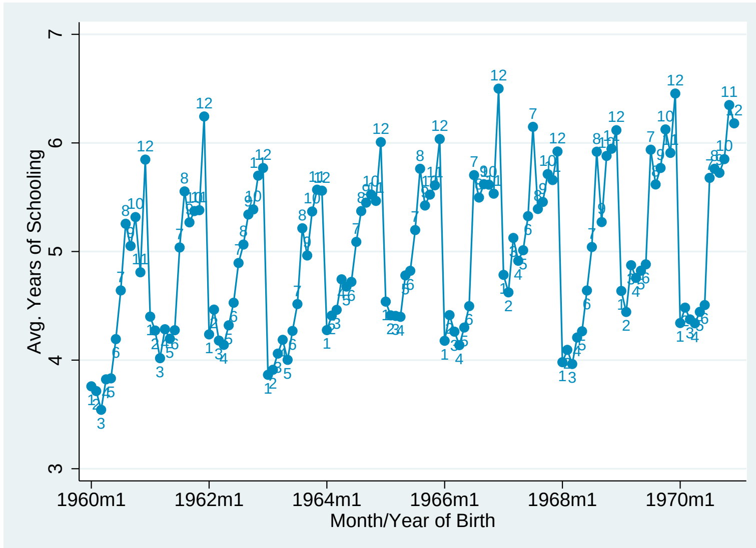
Educational outcomes, such as years of schooling and literacy rate, can vary across birth months due to the nutritional and health condition during pregnancy and the period after birth (Kramer, 2003; Neggers and Goldenberg, 2003). In Malawi, one of the worlds poorest countries, educational outcomes appear to substantially and persistently vary across birth months. Those born in the second half of each year have 1.6 more years 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 schooling years—between those born in December of a particular year and those born one month later(in January of the next calendar year)—is, on average, 1.6 years. Those differences are quite substantial, given that the average years of schooling is six years(Figures 1–3). Similarly sharp and distinct patterns are seen in the literacy rate and probability of accomplishing higher grades (Figure 10 and Figure 11). This pattern is distinct for forty years, at least, regardless of gender, urban-rural location, region, drought or non-drought districts, and religion (Figure 6–9)[1].

There are several possible mechanisms that could generate such a sharp variation of educational outcomes across birth months. As we mentioned above, one possible mechanism is the effect of nutrition or malaria infection, of the pregnant mother, on brain development. First, in developing countries, seasonal variation of food production generates seasonal nutritional intake due to the inability to smooth consumption. The resulting periodic insufficient nutrition intake during pregnancy or the period after birth affects individuals long term outcomes, such as cognitive ability (Kramer, 2003; Neggers and Goldenberg, 2003; Rayco-Solon et al., 2005; Ceesay et al., 1997; Moore et al., 2004; Verhoeff et al., 2001; Alderman et al., 2006; Almond and Currie, 2011; Glewwe et al., 2001). Similarly, the Malaria infection during pregnancy could induce lower birth-weight for the child, which will affect the brain in later developmental stages (Guyatt and Snow, 2004)

However, the period of low agricultural output and the peak season of malaria in-

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

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

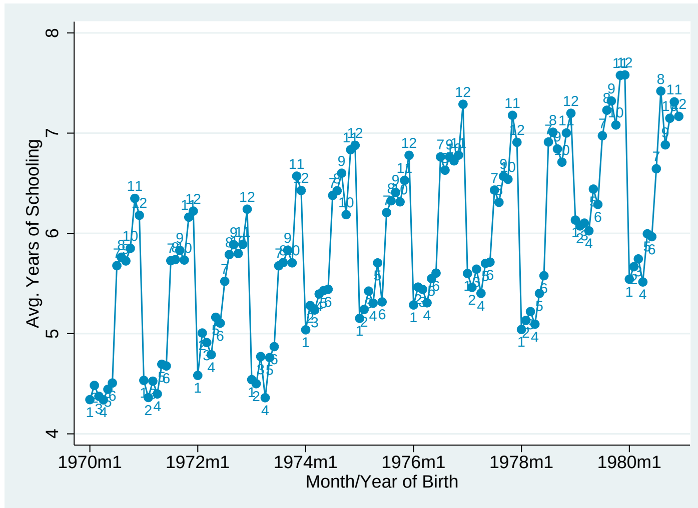


Notes: The source is the MPHC 2008. n=97,392. The years of schooling excludes pre-school years. For all figures below, pre-school years are excluded from the total calculation, unless it provides clarification.

fection in Malawi are not consistent with the observed seasonal pattern of educational outcomes. Simply, those who were born in the second half of each year (who exhibit better educational outcomes) experience malnutrition and relatively high malaria infection during the first two trimesters of gestation. Additionally, in our data set on Malawi, the birth weight of newborn babies across birth months is flat and does not show a consistent pattern.

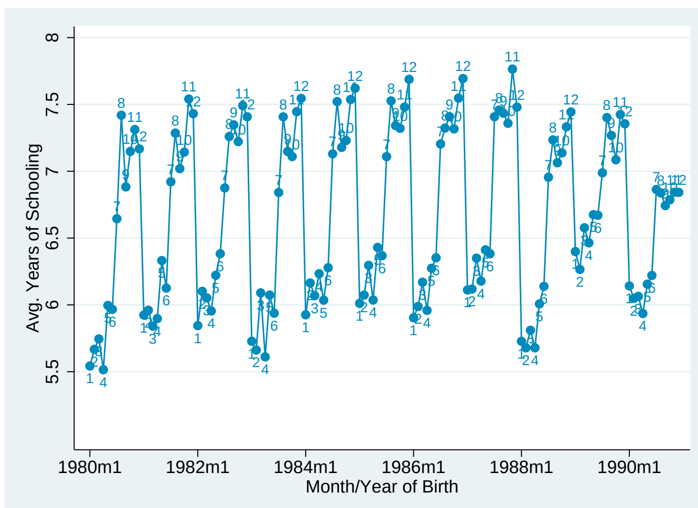
One might argue that academic rules may generate such a pattern. Angrist and Krueger (1991) shows that years of schooling varies substantially depending on the quarter of the birth. If some students are allowed to exit from school early, depending on actual age and not school age, it is possible that schooling years are associated with birth month. In addition, students who were born just before the first month of the academic schedule might perform less successfully than those born just after the first month of the academic schedule. This is especially probable in the early stages of schooling. If academic performance during early stages of schooling affects educational outcomes in later years, it is possible that birth month is a factor. Interestingly, in Malawi, the government changed the start month of the academic schedule twice, in

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.

1994 and 2009. However, as we demonstrate in section 2.1, the variation pattern of educational outcomes across birth months did not change, even after the academic schedules starting month was adjusted. This suggests that the academic schedule is likely not the cause of the variations of educational outcomes.

The reader might think such a variation in educational outcomes must be caused by an error in the data collection process of the census. To assess such a possibility, we examine not only the census but also the Malawi Demographic and Health surveys and Integrated Household Surveys. We show that all data sets exhibit the same pattern.

To explain these discrepancies among the variation in educational outcomes, birth weight across birth months, and seasonal variation of nutrition and malaria infection, we propose the following selection hypothesis. When a developing embryo or fetus in utero experiences malnutrition or malaria infection, only one with certain innate (genetic) traits can survive gestation or the period after birth. Those innate traits are positively correlated with health and educational outcomes. Individuals who were born in the second half of the year, and those who are alive today, have those innate traits. Consequently, they exhibit better educational outcomes.

Several remarks are needed to explain our hypothesis. First, recent literature in 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 the average height is positively correlated with income per capita in middle-high income countries. Deaton (2007) and Bozzoli, Deaton and Quintana-Domeque (2009) argue that this is because of the selection effect. In extremely low income countries, individuals who would otherwise die in early childhood or in utero are able to survive when income increases. As a result, the average height of people of such countries starts to decrease. During the Great Chinese Famine, in which three million people died due to hunger, the height of those who experienced the famine in utero (treated cohort) is as tall as those who did not experience the famine in utero (untreated cohort). Gørgens, Meng and Vaithianathan (2012) shows that the mechanism in which the height of the treated cohort is nearly the same as the untreated cohort is through the selection effect. During the Great Chinese Famine, only those who were inherently tall could survive. Consequently, the observed height of both the treated and untreated cohorts is the same.

Second, from the perspective of malaria infection, the idea that individuals with certain genetic traits respond differently to malaria infection is not ungrounded. In biology literature, it is known that malaria imposes the strongest selective pressure on the human genome, and researchers found that several genes provide a selective mechanism (Kwiatkowski, 2005). For example, the HBs gene causes the genetic disease, Sickle cell disease. Sickle cell disease is fatal in individuals homozygous for Hbs. However, individuals with the HBs gene possess a strong immune mechanism to the malaria infection. Thus, in malaria prevalent areas, individuals with HBs are protected from death caused by malaria.

To show empirically that our selection hypothesis explains the puzzling patterns of the variation of educational outcomes and the population across birth months in Malawi, we provide several pieces of evidence. First, we show that the number of individuals born in the second half of each year who are alive today is about 50 percent lower than the number of individuals living today who were born in the first half of each year. Such a huge reduction of the number of those living today who were born in the second half demonstrates that there is a substantial selection on those born, or those who were to be born, in the second half of each year.

Second, we directly examine the relationship between the degree of selection (in which the number of individuals born in the second half is lower than the number of individuals born in the first half) and extra-duration of schooling years (the rate in which years of schooling is longer for those born in the second half compared with those born in the first half). More specifically, we have noticed that the degrees of selection differ in various years. This is likely because Malawi's economy frequently experiences famine. This suggests that we should observe the instances when extra-duration of schooling years increases in a year when the degree of the selection rises. Similarly, we have noticed that there is substantial cross-sectional variation regarding the degree of the selection. This is likely because different districts have unique malaria prevalence rates, implying that extra-duration of schooling years should increase in areas where the degree of selection is higher.

Third, we show that parents birth months strongly affect the birth month of their children. If our selection hypothesis is true, then individuals alive today who were born in the second half of each year must have passed the selection. Thus, they must possess innate traits that help them survive. Consequently, children of those individuals are likely to inherit those innate traits. Therefore, if our hypothesis is true, those children should not show a varied population across birth months. On the other hand, if individuals were born in the first half, they did not experience the selection. Hence, those individuals include ones who do not have those innate traits to help them survive.

Therefore, children of those individuals do not have those innate traits, implying that—if our hypothesis is true—those children should display a seasonal population variation across birth months. In fact, we have observed such patterns.

Our study contributes to the existing literature in several ways. First, we show that the selection mechanism exists in the indices, other than height. In the previous studies, Bozzoli, Deaton and Quintana-Domeque (2009) and Gørgens et al. (2012) found that the selection mechanism exists in height in Africa and during the Chinese Great Famine. In this paper, we find that the selection mechanism exists in educational outcomes for at least thirty years for all cohorts, regardless of gender, urban-rural, different religion, or different area. As we discuss in the conclusion section, the presence of selection mechanisms is important when contemplating the effects of policy intervention.

Second, our finding suggests serious policy discussion about the fact that the number of those living today who were born in the second half is 50 percent lower than the number of those alive today who were born in the first half. This is despite the fact that conception occurs evenly across months, implying substantial pregnancy termination or child and infant mortality of those who were born or were to be born in the second half of each year. The fact that conception happens equally suggests that the local population does not know this fact. We discuss the detailed implications of our finding in the conclusion section.

The remainder of this paper is organized as follows. In Section 2.1, we look at the variation in the number of schooling years 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 its GDP per capita at only $320 USD, as of 2013. Malawis income per capita being 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. As of 2008, the population of Malawi was 13 million. Fifty percent of the population is considered poor (World Bank, 2014). Education outcomes in Malawi are also very weak. Although gross primary enrollment 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 the 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. 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 sampling structure, IPUMS-MPHC 2008 is the main source of our analysis.

The second data set is the Demographic and Health Surveys (DHS)from 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 ages 15-49 years are individually interviewed. However, in among a third of these households, men ages 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 records, if they live in the same household. For children born to the female respondent in the five years before the interview date, the DHS collects birth weight data. The DHS also collects fertility data for each female respondent for the five years before the interview date. This data consists of every pregnancy in the last five years, the term of the pregnancy, and whether the pregnancy ended in termination or a live birth. Children are matched with their parents from this data.

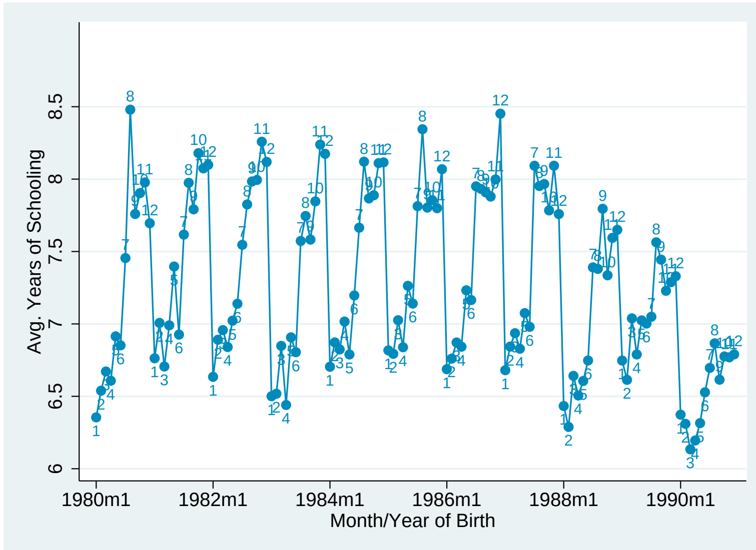
3 Analysis

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.

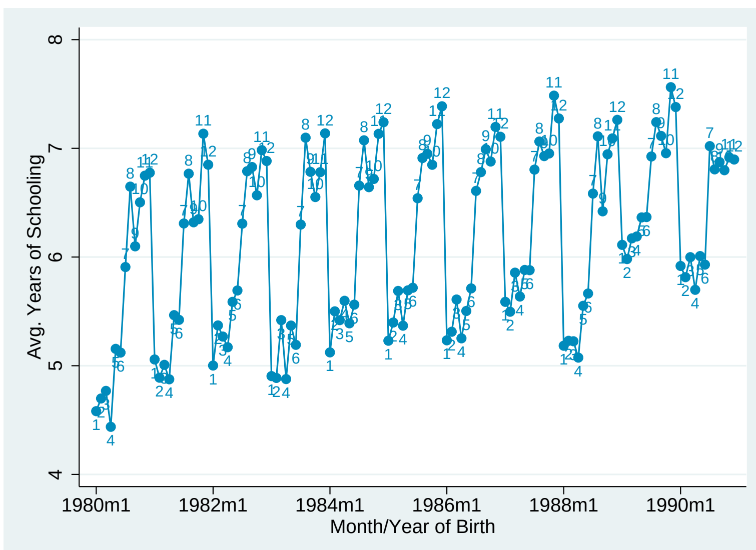
In Malawi, the country can be divided roughly into three regions. Also, we can categorize all districts into non-drought district and drought district. One might wonder some region-district specific factors such as weather or drought cause the seasonal variation of years of schooling in Malawi. Figure 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

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



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

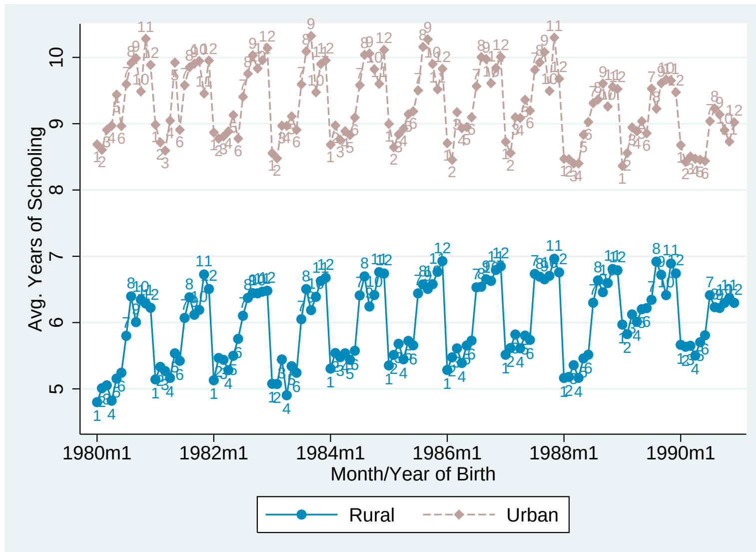
Figure 5: Average Years of Schooling for Female 1980-1990 Cohorts



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

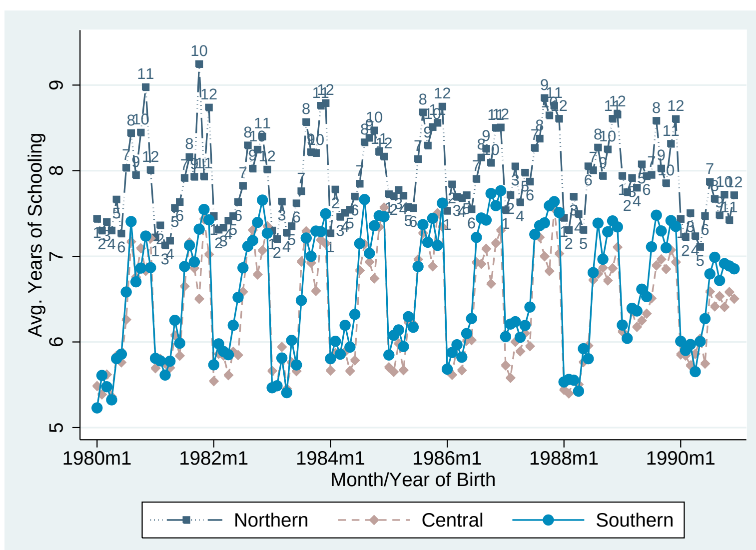
of years of schooling across birth months.

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



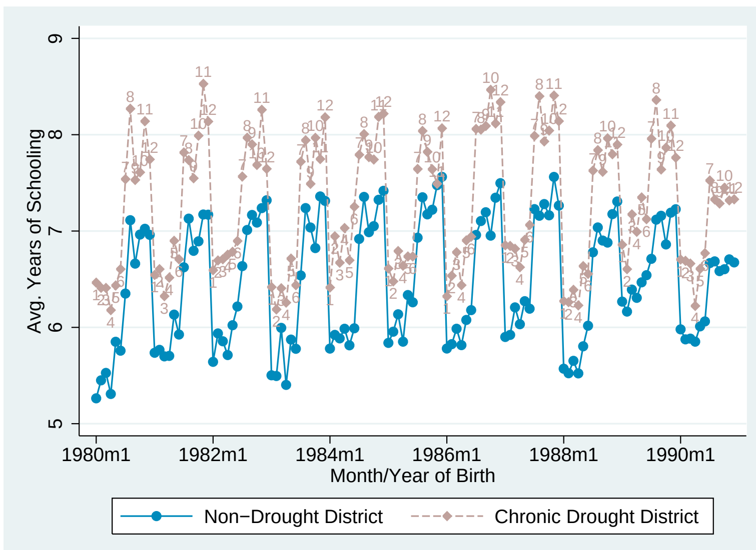
Notes: The source is the MPHC 2008. n=264,957

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

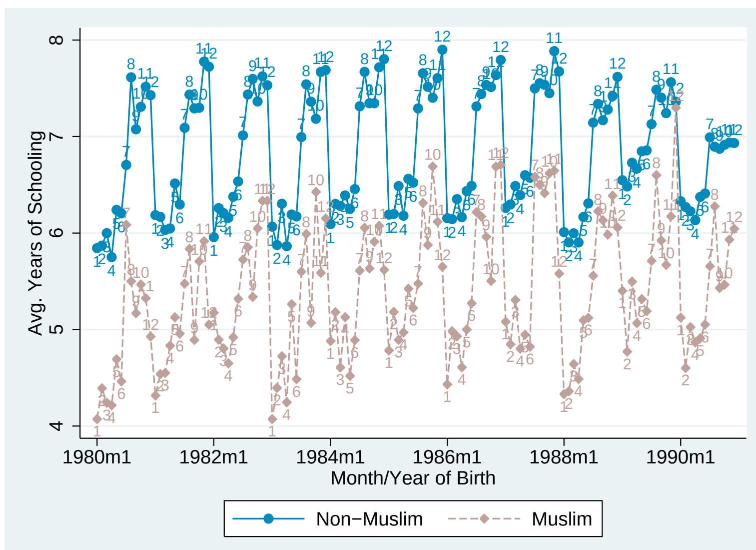


Notes: The source is the MPHC 2008. n=264,957. Drought prone districts are Zomba, Chiradzulu, Blantyre, Mwanza, Phalombe, Chikwawa, Nsanje, Balaka, Neno.

In Malawi, 15 percent of the population is Muslim.[2] 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.

The seasonal years of schooling can be seen at both ends of the distribution. Figure 10 shows the literacy rate across birth month and year. It shows that those who were born in the first half of each year have lower literacy rate than those who were born in the second half of each year.

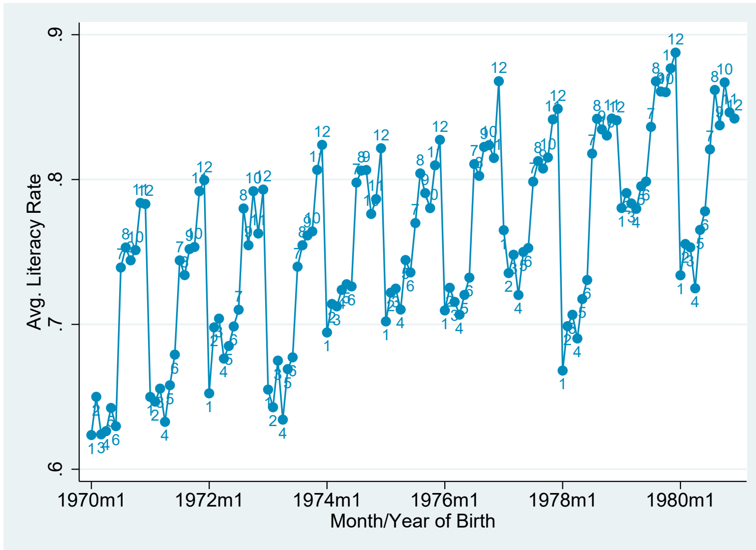
Figure 9: Average Years of Schooling for 1980-1990 Muslim and Non-muslim Cohorts



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

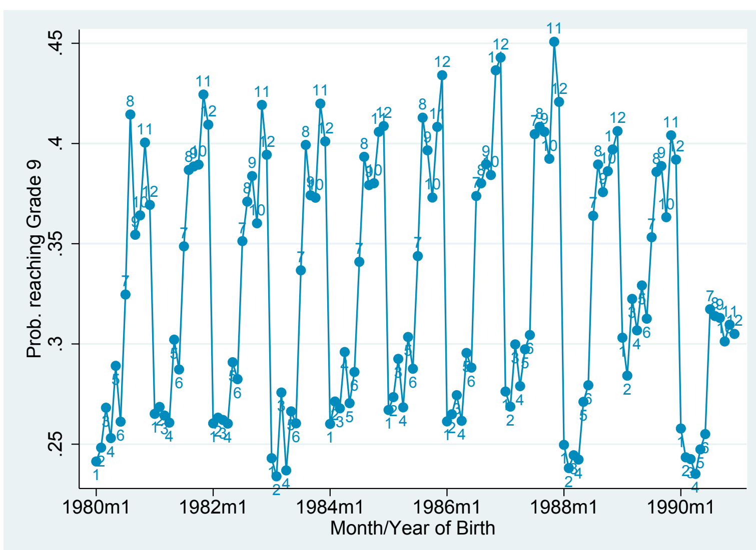
[2]In 10 percent census, the percentage of muslim is 14.57%

Figure 10: Average Literacy Rate for 1980-1990 cohorts



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

Figure 11: Probability of Completing at least Grade 8 for 1980-1990 cohorts



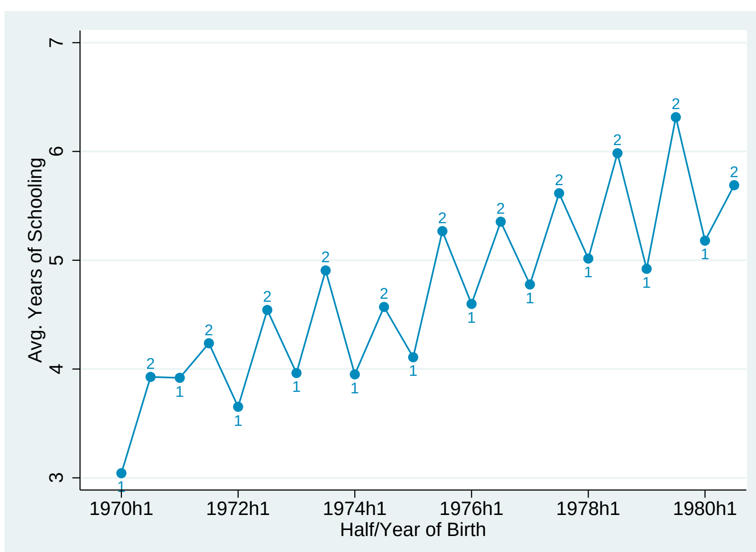
Notes: The source is the MPHC 2008. n=264,957.

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 12, 13 and 14 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 12, 13 and 14 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.

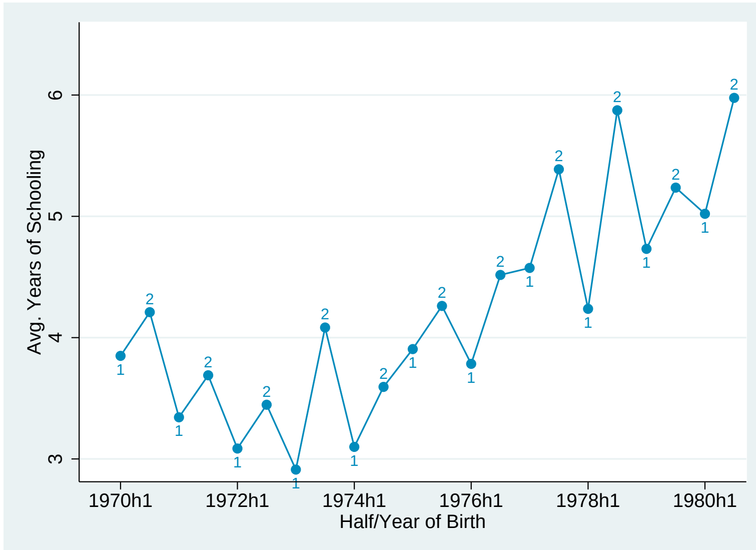
Figure 12: Average Years of Schooling for 1970-1980 Cohorts in DHS 2000



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

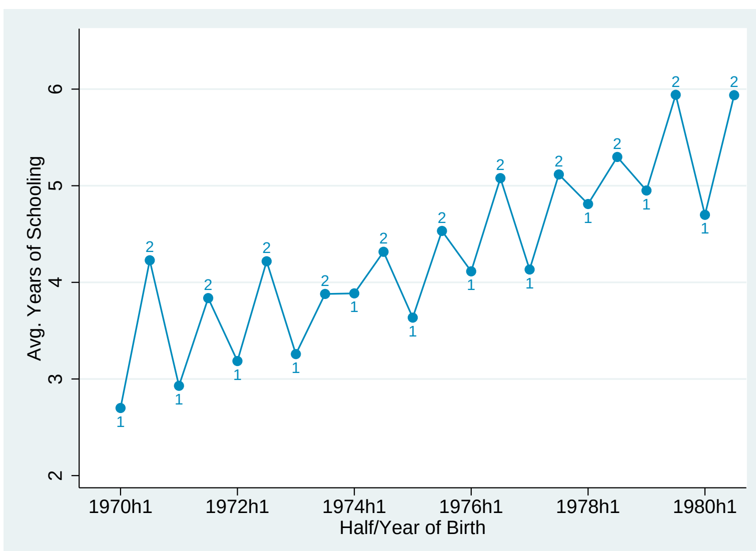
The fact that DHS data sets show the same seasonal variation of years of schooling

Figure 13: Average Years of Schooling for 1970-1980 Cohorts in DHS 2004



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

Figure 14: Average Years of Schooling for 1970-1980 Cohorts in DHS 2010



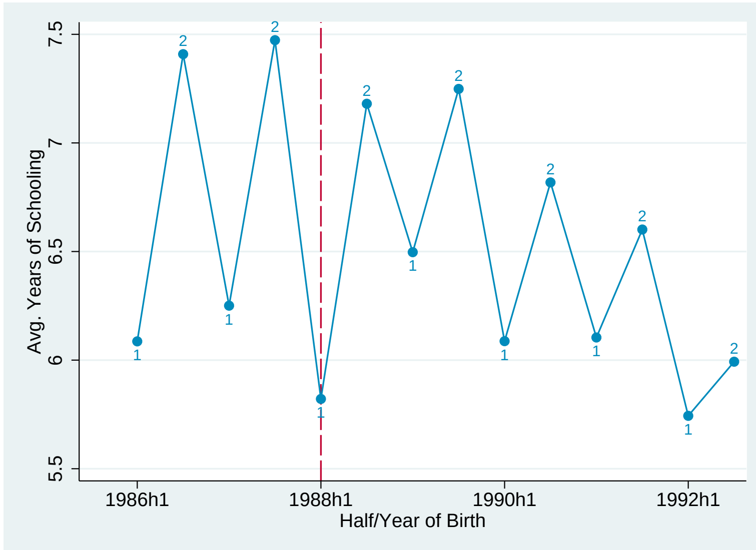
Notes: The source is the DHS 2010. n=5,369.

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

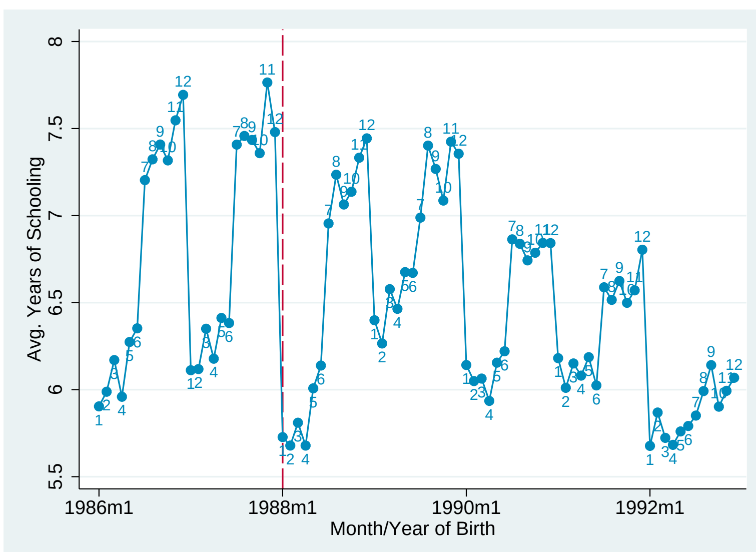
In one of the most cited papers regarding the effect of years of schooling on earnings, Angrist and Krueger (1991) argue that in the United States, the compulsory educational law induces the variation of years of schooling across birth months. This is because if a person reaches a certain age, they are exempted from the compulsory educational 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 from September. But after 1994, the school calendar started from January. Figure 18 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 18 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 18 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 15: Average Years of Schooling around 1988 Cohorts (1986-1994 Cohorts)



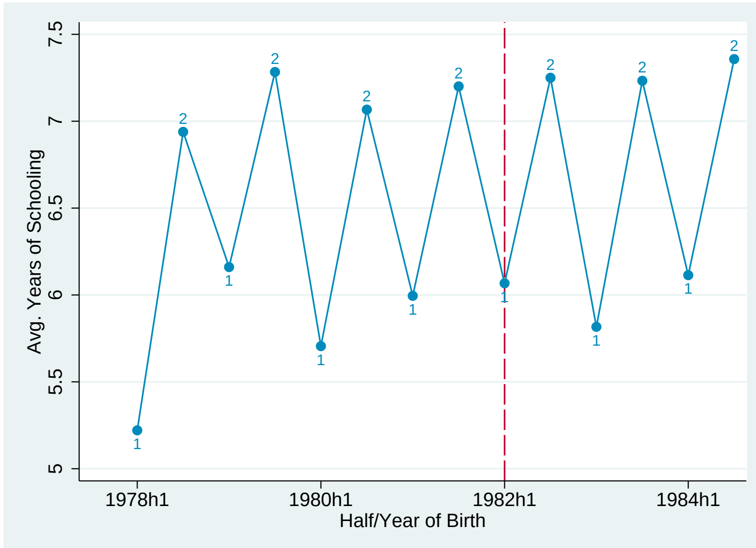
Notes: The source is the MPHC 2008. n=173,780. Cohort 1988 was supposed to be grade 1 in 1994 when the academic calendar was changed.

Figure 16: Average Years of Schooling around 1988 Cohorts(disaggregated)



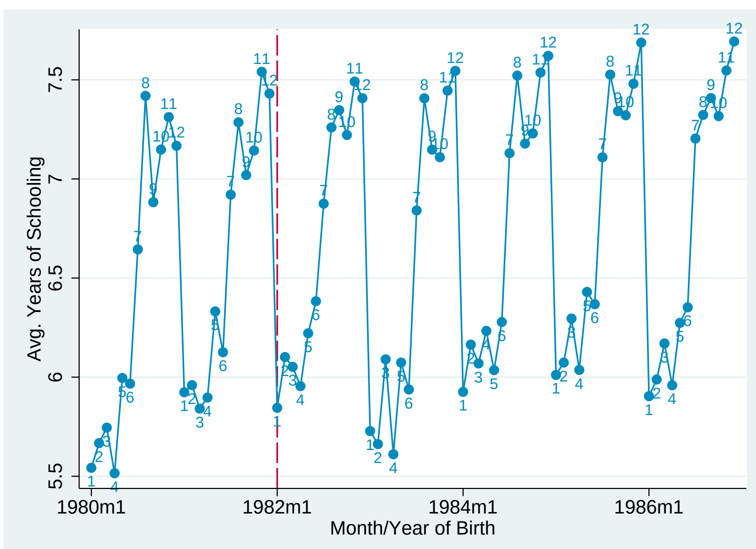
Notes: The source is the MPHC 2008. n=124,314. 1988 Cohort is supposed to be grade 1 in 1994 when the academic calendar was changed.

Figure 17: 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 18: 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 Malnutrition and Malaria Infection on Brain Development

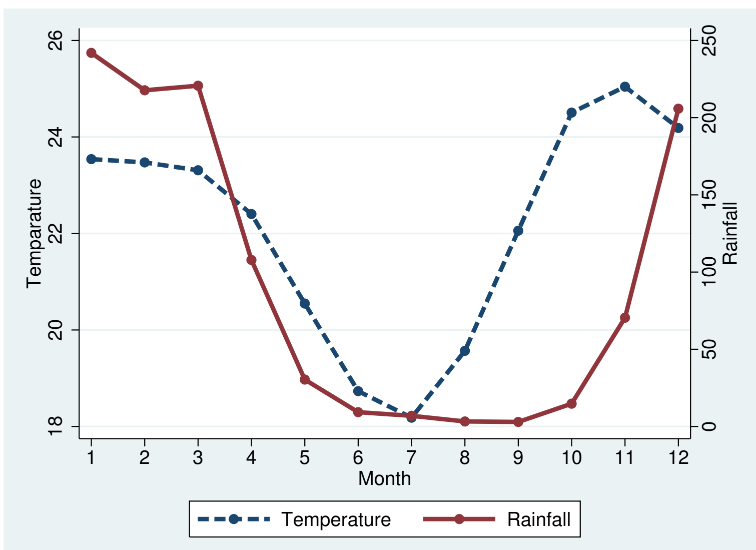
Malaria Infection, Nutrition during Pregnancy and Birth Weight

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 (Figure19). 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.

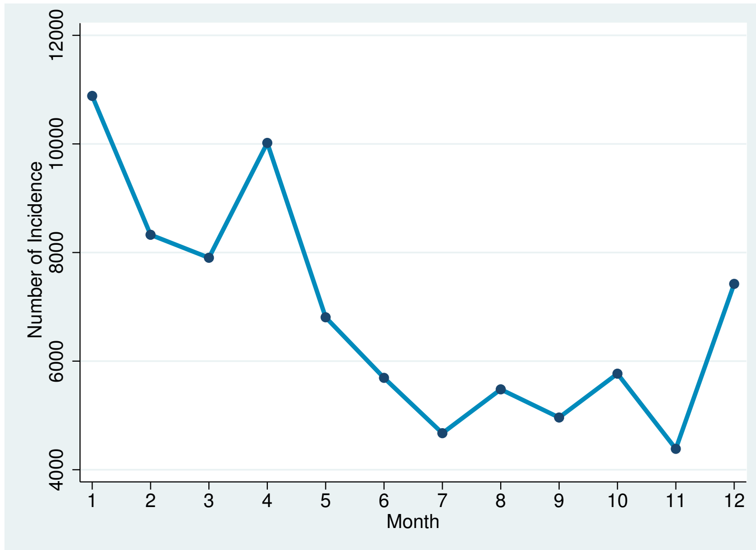
Figure 19: Average Monthly Temperature and Rainfall in Malawi (1901-2018)



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 20). The infections is particularly high during the rainy

Figure 20: Monthly Malaria Incidence in Malawi (2012-2015)



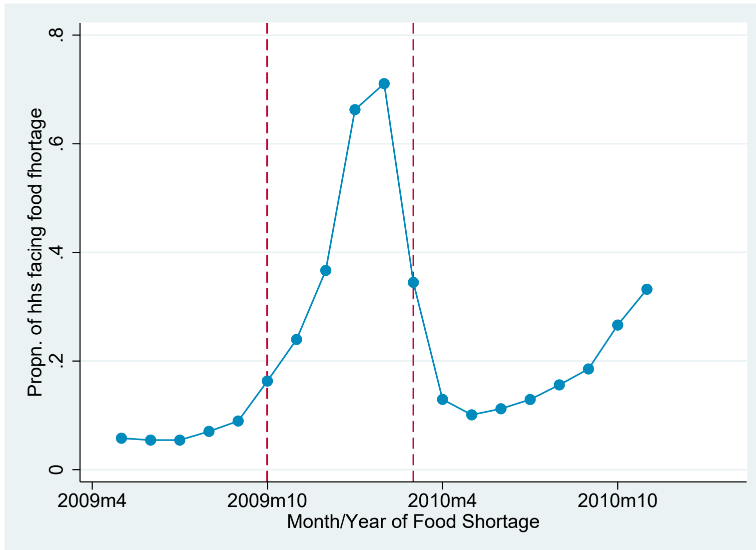
Notes: The source of the data is Hajison et al. (2017).

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.

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 21 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. Fig-

Figure 21: The Experience of Hunger over Months

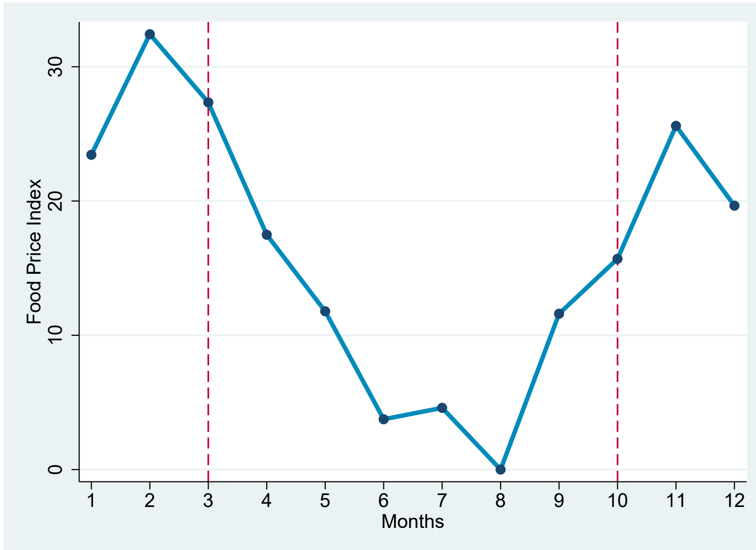


Notes: The vertical axis measures the share of the household who experience the food shortage in a particular month. Source: Third Integrated Household Survey Malawi 2010-11. Those who conceive during October to March will have birth in the second half of year.

ure 22 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.

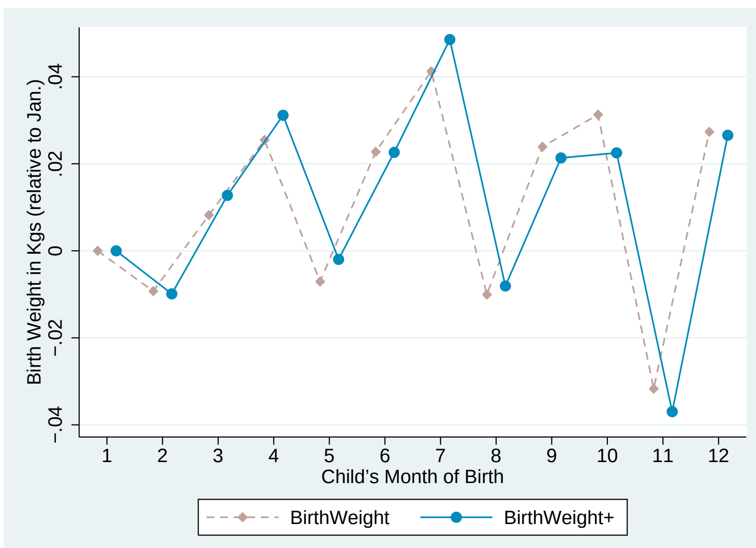
To support our argument further, we also examine the birth weight of new born children over birth months. If the infection to Malaria or malnutrition during pregnancy is the cause of generating the variation of years of schooling across birth months, then it should be reflected on birth weight across 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 23, we plot the birth weight across birth months. In Figure 23, 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

Figure 22: Seasonality of General Food Price in Malawi (2007-2018)



Note: Source is General Food Price Index in Malawi (FAO (2019)). Those who conceive during October to March will have birth in the second half of year.

Figure 23: 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.

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.

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 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 an egg is conceived during October to March, it is very likely that developing embryo and fetus in utero experience severe conditions such as mal-nutrition and malaria infection. Our hypothesis is that developing embryo or fetus in utero with certain innate (genetic) traits can eventually survive during gestation period or the period after birth. Those innate (genetic) traits are positively correlated with health and educational outcomes. Individuals who were born in the second half and those who are now alive have those innate (genetic) traits. As a result, they exhibit better educational outcomes.

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 directly examine the degree of selection and years of schooling premium (the rate in which years of schooling is longer for those who were born in the second half than for those who were born in the first half of each year). More specifically, we have noticed that the degrees of selection (the ratio in which the number of population who were born in the second half of each year is lower than the number of population who were born in the first half of each year) is different in different years. This is likely due to the fact Malawi quite frequently experiences famine. This suggests that we should observe years of schooling premium (the rate in which years of schooling is longer for those who were born in the second half than for those who were born in the first half of each year) is larger in a year when the degree of the selection is higher. Similarly, we have noticed that there is substantial cross-sectional variation regarding the degree of the selection (the ratio in which the number of population who were born in the second half of each year is lower than the number of population who were born in the first half of each year) for different district. This is probably because different district has different malaria prevalence rate since different district has different weather pattern and altitude. This implies that in an area where the degree of selection is higher, then years of schooling premium becomes higher.

Thirdly, we show that birth month of parents affects strongly the birth month of children. If our selection hypothesis is true, then individuals who were born in the second half of each year and who are alive now must have passed the selection. Thus, they must have innate traits that help them to survive. As a result, children of those individuals are likely to inherit those innate traits. Thus, those children will not show variation of population across birth months. On the other hand, if individuals were born in the first half of each year, they did not experience the selection. Thus, those individuals include those who do not have those innate traits that help them to survive. Thus, children also do not have those innate traits. This implies those children show the seasonal variation of number of population across birth months. In fact, we show that the data shows this predicted patterns.

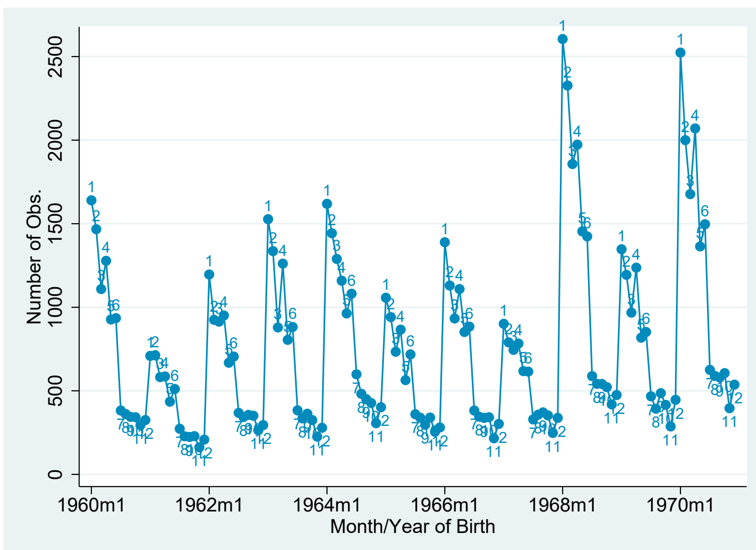
Results

Figures 24–26 show the first group of evidences of our selection hypothesis. They 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 24–26 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– 28 show similar patterns in the DHS dataset and the Integrated Household Survey.

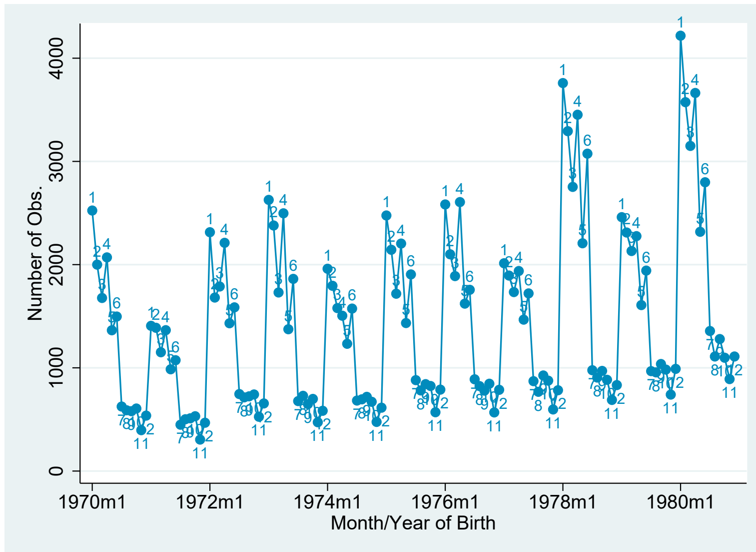
One might think that such a huge variation of the number of birth across birth month might come from parents' selective choice of a birth month. I Figure 29 examines the conception pattern across birth month. The blue solid line show the share of the conception across conception months. On the contrary, the conception pattern is quite flat and it does not show a patten consistent with Figures 24–26. On the other hand, those who conceived during October and March have higher pregnancy termination rate. This suggests that substantial selection is happening for those who conceived during October to March.

Figure 24: The number of alive population across birth month:1960 to 1970



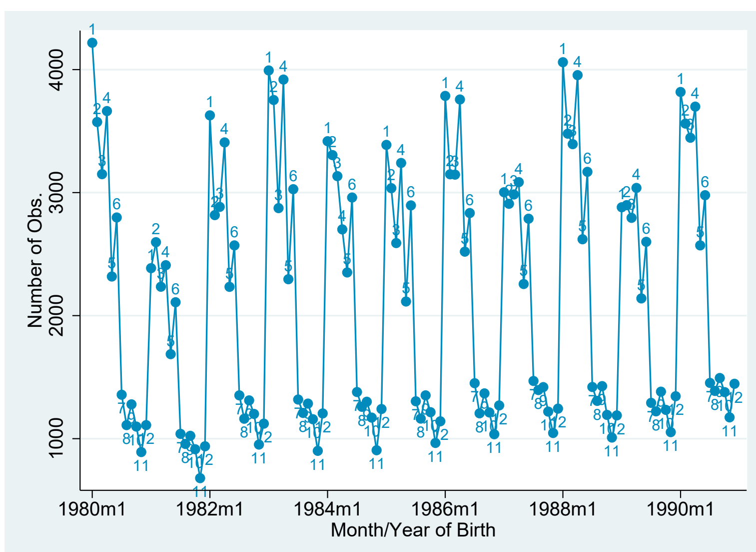
Note: The data source is MPHC 2008. On the graph k indicates the number of observation in k-th month of each year.

Figure 25: The number of alive population across birth month:1970 to 1980



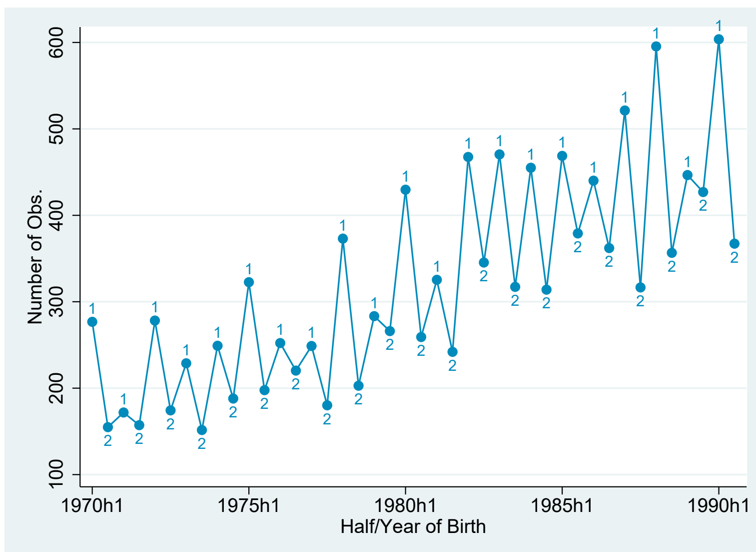
Note: The data source is MPHC 2008. On the graph k indicates the number of observation in k-th month of each year.

Figure 26: The number of alive population across birth month:1980-19980



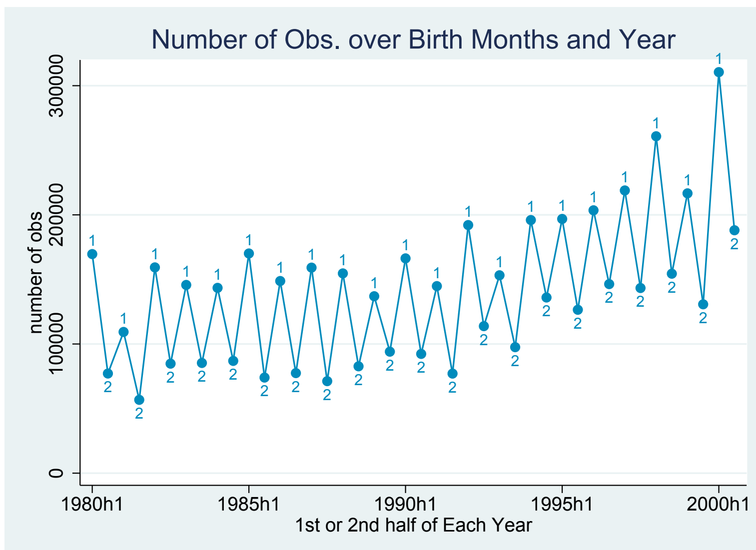
Note: The data source is MPHC 2008. On the graph k indicates the number of observation in k-th month of each year.

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



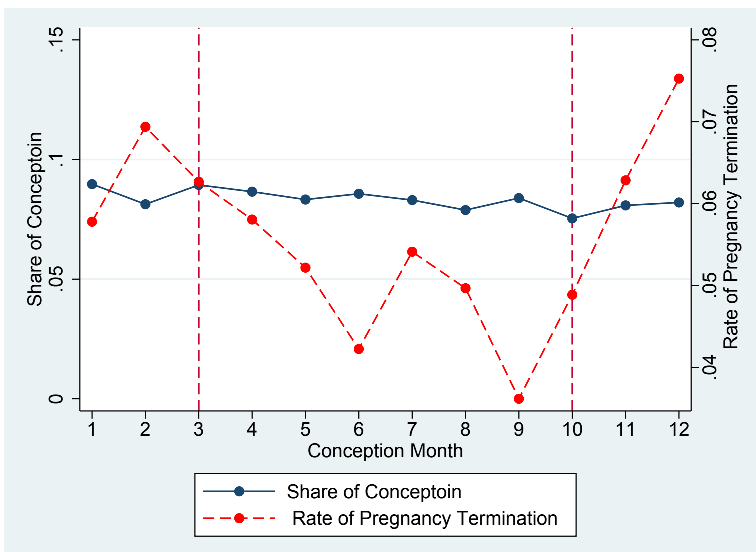
Notes: The data source is DHS 2010. 1 indicates that the number of observation born in the first half of the year. 2 indicates the number of observation born in the 2nd half of the year.

Figure 28: The number of observation across birth month and year, aggregated in the first half and 2nd half of each year(Malawi Third Integrated Household Survey)



Notes: The data source is Malawi Third Integrated Household Survey (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 29: Monthly Share of Conceptions and Pregnant Termination Rate over Months



Notes: The data source is DHS 2000, 20004, 2010. If an egg is conceived from October to March, it is likely that a child is born in the second half of a year.

Figure 30 and Figure 31 are the second group of evidences to prove our selection hypothesis. In Malawi, the degree in which population born in the second half is lower than the population born in the first half is different across years and birth place. This variation mainly comes from food production shock (time-series) and the degree of malaria prevalence rate. This suggests that if our selection hypothesis is true, then we should observe a positive correlation between the degree of attrition of the second cohort and the degree of years of schooling duration premium. Figure 30 shows the time series correlation between the degree of the selection and the degree of years of schooling premium. For the degree of selection in year t, we calculate the percentage in which the number of population born in the second half in year t is lower than the number of population born in the first half in year t. For the degree of the schooling duration premium in year t, we calculate the percentage in which the duration of schooling of individuals born in the second half is longer than the duration of schooling of individuals born in the first half. Figure 30 shows that when the degree of the selection becomes higher, the degree of schooling duration premium becomes larger. Table 1 shows the regression result of OLS. It shows that one percent increase of the degree of selection is associated with 0.89 percent increase of the year of schooling duration premium. Similarly, Figure show the cross sectional correlation between the degree of the selection and the degree of the year of schooling duration premium. Figure 30 and Figure 31 show that there is a strong association between the degree of selection and the degree of years of schooling duration premium.

Figure 32 and Figure 33 are the third group of evidences to prove our selection hypothesis. If our selection hypothesis is true, then individuals who were born in the second half of each year and who are alive now must have passed the selection. Thus, they must have innate (genetic) traits that help them to survive. As a result, children of those individuals are likely to inherit those innate traits. Thus, those children are not likely to show a variation of population across birth months. On the other hand, if individuals were born in the first half of each year, they did not experience the selection. Thus, some of those individuals do not have those innate traits that help them to survive. Thus, some of those children of those individuals do not have those genetic traits. This implies those children should show the seasonal variation of number of population across birth months.

Figure 30: Correlation between the Degree of the Selection and the Degree of Schooling Duration Premium: Time Series Data

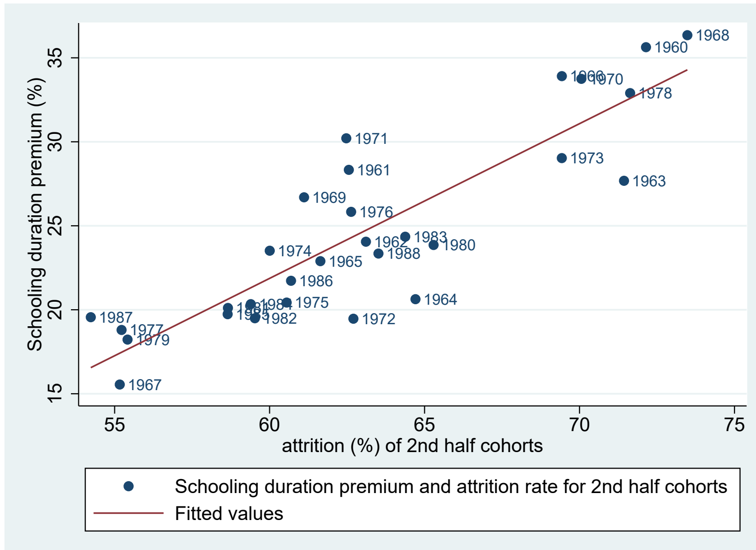


Figure 31: Correlation between the Degree of the Selection and the Degree of Schooling Duration Premium: Cross-sectional Data

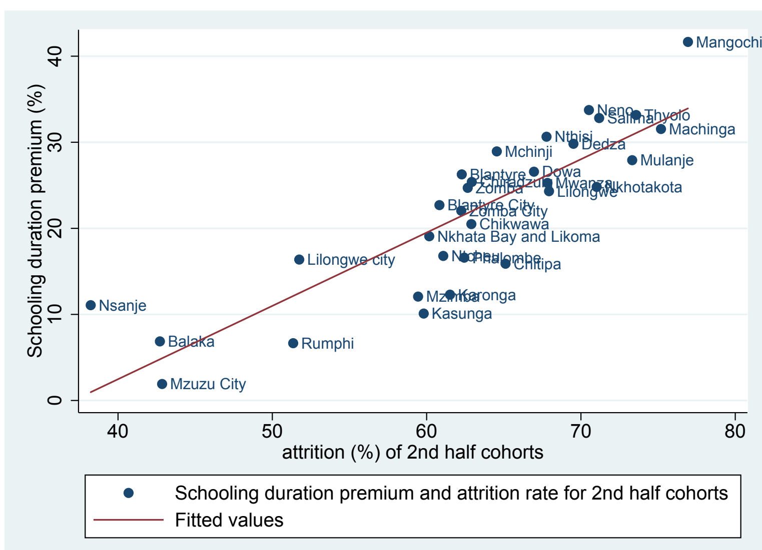
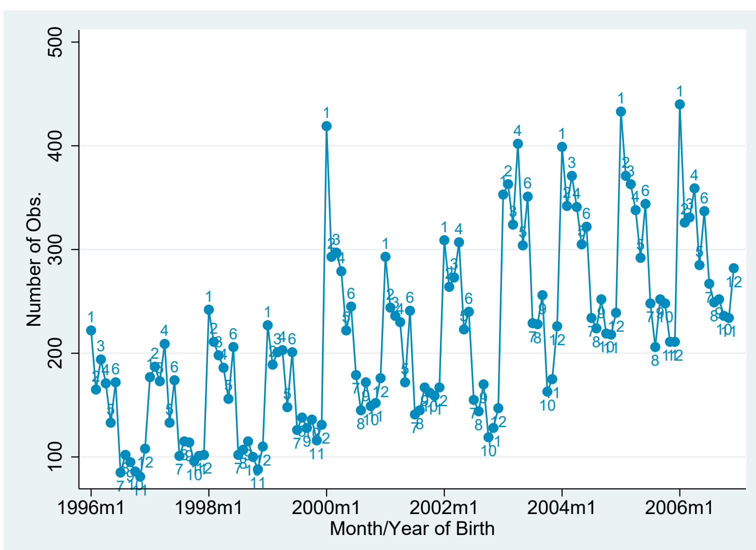


Table 1. Association between the Degree of Selection and Years of Schooling Duration Premium (OLS)

Dependent Variable	Years of Schoolign Duration Premium (%)				
Variable	(1)	(2)	(3)	(4)	(5)
Selection Index (%)	0.921***	0.883***	0.883***	0.883***	0.851***
	(0.0872)	(0.0935)	(0.0954)	(0.0964)	(0.1203)
Linear trend		yes	yes	yes	
Quadratic trend			yes		
N	29	29	29	29	31
R-squared	0.760	0.766	0.766	0.766	0.72

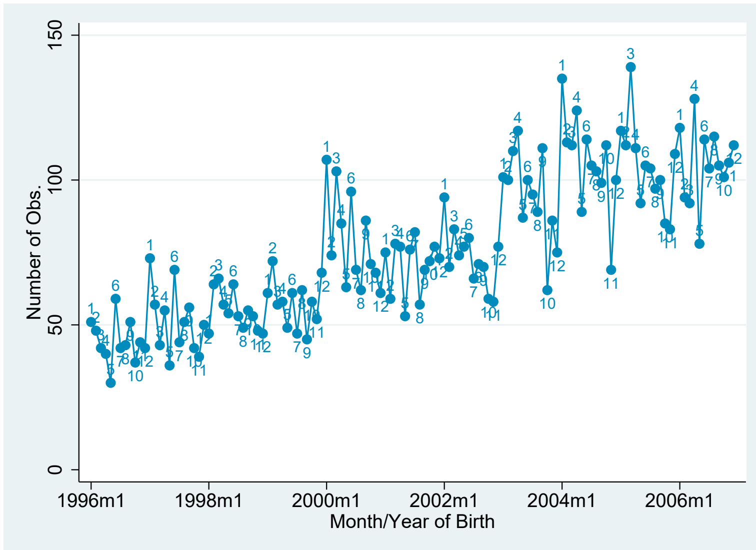
Notes: Column (1)-(4) use the time series data. Column (5) uses the cross-sectional data. Robust standard errors in parentheses except column (4). For column (4), Newy-west standard error with two period lag is calculated. *** p<0.01, ** p<0.05, * p<0.1

Figure 32: Number of Observations across Birth Months and Year When Both Parents were Born in the First Half of Year



Notes: The data source is MPHC 2008. The sample is restricted to children who were born between 1998-2006 and who live with both parents who were born in the first half of any year and parents' years of schooling are 6 or 7 years. The number of children born in the second half is 59 % of those born in the first half in this sample.

Figure 33: Number of Observations across Birth Month and Year When Both Parents were Born in the Second Half of Year



Notes: The data source is MPHC 2008. The sample is restricted to children who were born between 1998-2006 and who live with both parents who were born in the second half of year and whose years of schooling is 6 or 7 years. The number of children born in the second half is 91 % of the number of children born in the first half in this sample.

To use this identification strategy, several remarks are needed. First, please note that whether children of parents who experienced the selection have genetic traits that help them to survive depends on whether genetic traits exhibit autosomal dominant or autosomal recessive. If they are autosomal dominant, it is possible that even if both parents experience the selection, some of their children do not posses innate traits that help them to survive. Therefore, even if both parents experience the selection, it is possible that some children do not own those parents' traits and, as a result, when those children experience malnutrition and malaria infection, they cannot survive.

Second, since parents' birth month is correlated with parents' education, it is important to control parents' education. Otherwise, it is difficult to see whether the difference of population pattern is due to the difference of innate traits or the difference of parents' education level. To control parents' education in the graph, we pick up children whose parents' years of school is six or seven years.

Figure 32 shows the number of observation across birth month and year when both parents are born in the first half of each year and they are still live and father's years of schooling is seven 6 years or 7 years. In Figure 32 , as the theory predicts, the number of children born in the first half become almost two times higher than the number of children born in the first half when both parents are born in the first half. On the other hand, Figure 33 shows that it does not exhibit such a pattern when both parents are born in the second half.

Table 2 calculates the share of children born in the first half and the second half based on parents' birth month. It shows that the share of the children born in the second half increase by 10 percentage point when both parents are born in the second half compared with the case when both parents are born in the first half. Since the initial share of children born in the second half is about 30 percent, it implies that parents' birth months increases the probability of children born in the second half by about 30 percent.

Table 3 checks this relationship in the regression form by controlling various factors. More specifically, we estimate the following equation in Table 3.

$$Y_{thi} = \beta_0 + \beta_1 \text{Treat}_{th} + \gamma X_{thi} + \delta_t + \delta_h + \epsilon_i \tag{1}$$

where t is index of birth year, h is the index of household and i is index of children. Y_{thi} is a dummy variable which is equal to one if a child i is born in the second half of year and otherwise zero. Treat_{th} is the key explanatory variable which is equal to one if both parents are born in the second half of year and otherwise equal to zero. X_{thi} is a vector of controlling variable such as child's birth year, gender, birth place, parents' education and other demographic characteristics. δ_t is the time fixed effect and δ_h is the household fixed effect. The sample is restricted to children who were born from 1994 to 2004 and who live with both parents who were both born in the second half or the first half of year.

Table 3 shows that when both parents were born in the second half compared with the first half, the share of children born in the second half increase by about 9 percentage point which is quite similar to the result in Table 2.

Table 2. Share of Children Born in Each Half and Parents' Birth Month

	Parents are born in the first half (A)	Parents are born in the second half (B)	Difference (B-A)
Share of children born in the first half	67.9%	57.2%	-10.7%
Share of children born in the second half	32.1%	42.8%	10.7%
Total	100%	100%	

Notes: Sample is restricted to children who were born from 1996 to 2004 and who live with both parents. Both parents' birth months need to be in the first half or the second half.

Table 3. The Effect of Parents Birth Month on Children's Birth Month (OLS)

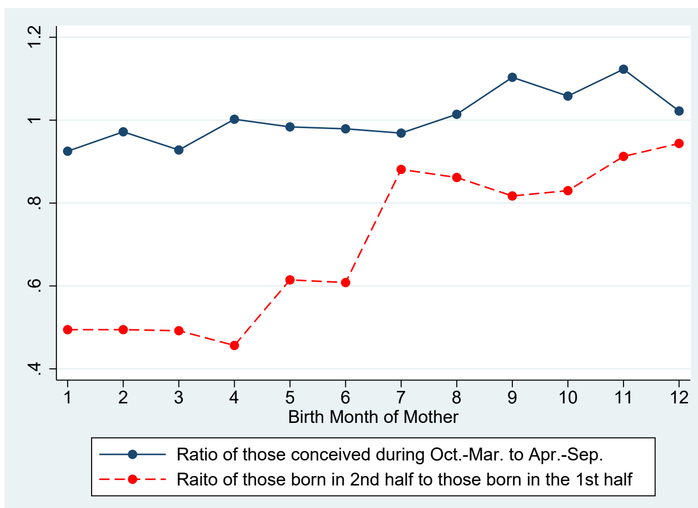
Dependent Variable	A child born in the 2nd half			
Variable	(1)	(2)	(3)	(4)
Parents born in the 2nd half	0.107***	0.107***	0.105***	0.0928***
	(0.00291)	(0.00291)	(0.00292)	(0.00296)
Birth year	yes	yes	yes	yes
Sex		yes	yes	yes
Parents birth year		yes	yes	yes
Parents educaiton			yes	yse
Other Demographic Characteristics				yes
N	163,332	163,332	163,332	163,332
R-squared	0.011	0.011	0.013	0.020

Notes: Sample is restricted to children whose both parents were in the second halr or whose both parents were born in the first half. Children's birth year is restricted from 1994 to 2004. Other demographic characteristics include birth place, district, ditrict times birth year. Clustering robust standard errors in parentheses assuming that the error term is correlated within children in the same family. *** p<0.01, ** p<0.05, * p<0.1

Figure 34 and Figure 35 show another way to look at our third group of evidences.

The blue solid line in Figure 34 shows that the ratio of those conceived from October to March to those conceived from April to September for mother's birth month. The red dot line shows the ratio of those born in the second half to those born in the first half. It show that although the conception month is not affected by mother's birth month, the birth month of currently alive children are affected by mother's birth month. It shows that if a mother is born in January, the ratio of children born in the second half to those born in the first half is almost 50 percent. This suggests that when the mother is born in January, many who were supposed to be born in January died in utero or those who were born in January died after birth.

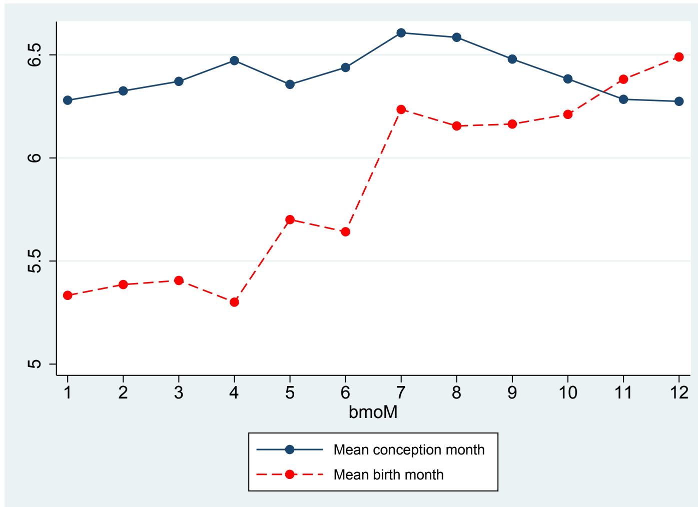
Figure 34: Ratio of Pregnancy that Conceived during October–March to April–September for each mother's birth month and Ratio of Those Born in the 2nd Half to those Born to the First Half for Each Mother's Birth Month



<u>Selection Process</u>

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 **??** and 37 and show the number of individual who were born during a few years before

Figure 35: Correlation between Average Conception Month and Mother's Birth Month vs. the Correlation between Average Birth Month of Children and Mother's Birth Month

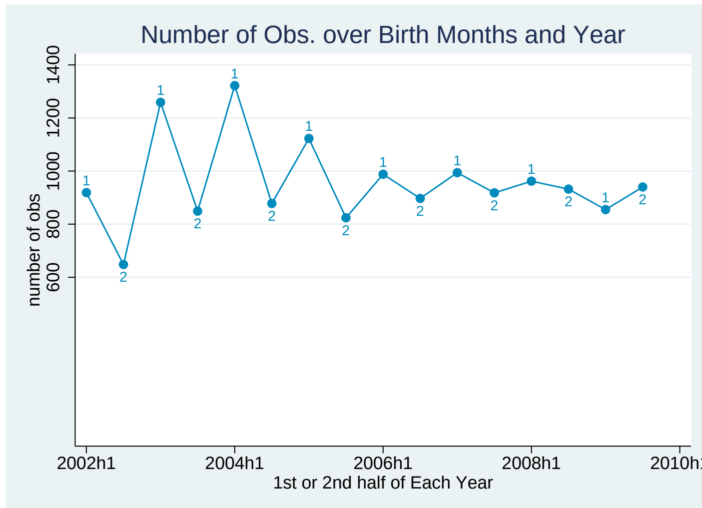


the survey years of the census and Integrated household survey. The graph shows that at the beginning, the number of population who were born in the second half is not so low. But after two years, the number starts to drop substantially.

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

Figure 36: The Total Number of observation across birth months (Integrated Households Survey)

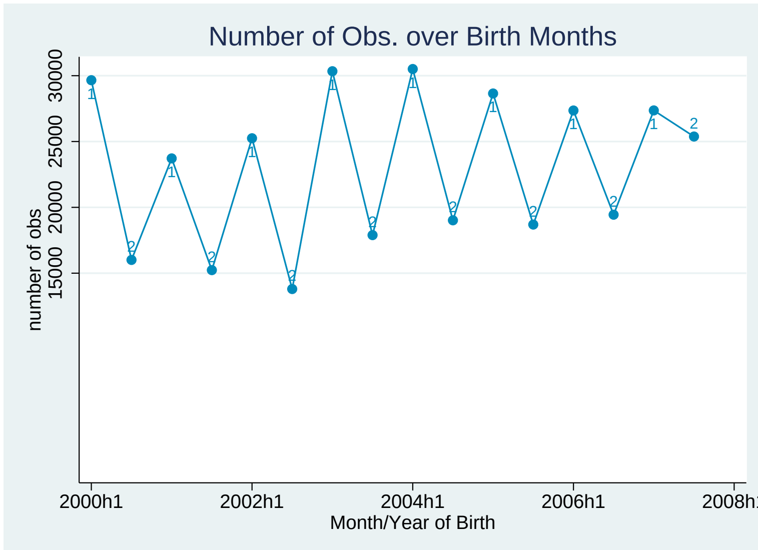


Notes: The data source is IHS3. The vertical axis measures the number of observation born in each half of the year. IHS3 was conducted in 2010. The graph shows that the difference of the number of the observation between the first and 2nd half increases as time passes.

feeding).

The Table 4 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.

Figure 37: 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.

Generalizability to other countries

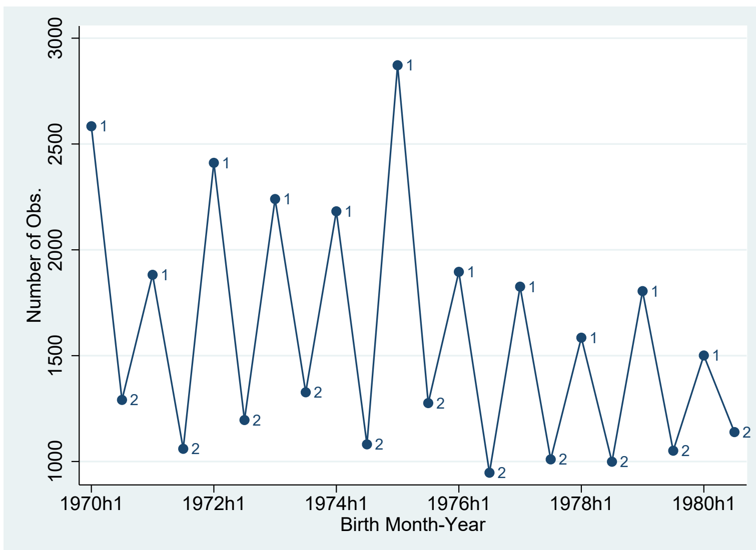
One natural question to our study is to what extent our finding can be applied to other countries. To check, we have looked at other countries. Unfortunately, in most of African countries, the census does not ask the key question for our study, birth month of each individual. Thus, we are restricted to use Demographic health survey which has much smaller sample size. Within this restriction, we found that several countries or region show similar patterns. One of those countries is Mozambique, which is located just east of Malawi and whose GDP per capita is similar to Malawi's GDP per capita. The following table show the number of observation across birth months. Similar to Malawi's case, it shows that the number of observation changes across birth months substantially.

Table 4: The Timing of Negative Shock and Birth Outcome

Calendar Month	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4
Least nutrition																						
Malaria																						
Conception Month																						
July	p1	p2	p3	p4	p5	p6	p7	p8	p9	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11	b12	b13
Aug		p1	p2	p3	p4	p5	p6	p7	p8	p9	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11	b12
Sep			p1	p2	p3	p4	p5	p6	p7	p8	p9	b1	b2	b3	b4	b5	b6	b7	b8	b9	b10	b11
Oct				p1	p2	p3	p4	p5	p6	p7	p8	p9	b1	b2	b3	b4	b5	b6	b7	b8	b9	b9
Nov					p1	p2	p3	p4	p5	p6	p7	p8	p9	b1	b2	b3	b4	b5	b6	b7	b8	b9
Dec						p1	p2	p3	p4	p5	p6	p7	p8	p9	b1	b2	b3	b4	b5	b6	b7	b8
Jan							p1	p2	p3	p4	p5	p6	p7	p8	p9	b1	b2	b3	b4	b5	b6	b7
Feb								p1	p2	p3	p4	p5	p6	p7	p8	p9	b1	b2	b3	b4	b5	b6
Mar									p1	p2	p3	p4	p5	p6	p7	p8	p9	b1	b2	b3	b4	b5

Notes: The above table shows the relationship between the timing of negative shock on birth outcomes. For example, a women who gives births in July starts p1 in October. A women who gives birth in August starts p1 on November. In Malawi, the least nutrition months are from November to March. Malaria peak season is from December to April. Those who give birth from July to December experience selection (yellow color). The table shows that as long as p2 overlap with the period of least nutrition or malaria infection, the birth outcome experience the selection.

Figure 38: The Number of Population across Birth Months in Mozambique (DHS)



Notes: The data source is DHS Mozambique 1997, 2003 and 2011. Due to sampling structure, the sample is restricted to women aged from 20 to 50.

4 Conclusion

We find that in Malawi, years of schooling and the population number varies across birth months substantially and consistently over forty years. We proposed a selection hypothesis that individuals with certain genetic traits survive when they are hit by malnutrition or malaria infection in utero. We have shown three groups of evidences that prove our hypothesis.

First, we have shown that the number of the population who were born in the second half is about 50 percent lower than the number of the population who were born in the first half, despite that the conception occurs equally across months. Second, we have shown that the degree of selection and the degree of the years of schooling duration premium is positively correlated in the time series and across districts. We have shown that one percent increase of the degree of the selection is associated with 89 percent increase of the degree of years of schooling duration premium. Third, we have shown that when parents did not experience the selection, then children of those parents show that some of them do not have innate traits, while when the parents experienced the selection, the children have innate traits.

There are several implications for our study. First, the fact that the number of those who were born in the second half and who are alive is 50 percent lower than the number of those who were born in the first half and who are alive. This implies that in terms of the current cohort size, one third of each cohort died during the gestation period or the period after birth. In 2008, the size of the cohort of age 20 is about 300, 000. Thus, for each cohort, 100,000 were lost during pregnancy or after birth. On the other hand, according to UNAIDS (2019), about 470,000 people died due to HIV in the Africa region in 2018 and the world communities are spending substantial resources to reduce HIV related deaths in Africa. Given this number, 100, 000 deaths for each cohort in Malawi is not small. It is important that the world community takes immediate steps to reduce this size of the death. On the other hand, it is true that a policy that is trying to reduce the deaths related with birth month could boost the population growth substantially in Malawi, which will impose the heavy burden on health, land and educational resources for a poor country like a Malawi. Thus, for implementing a policy that will attempt to reduce deaths related with birth month, additional resources would be needed.

For implementing such a policy, our findings are quite useful. Our findings show that when both parents are born in the first half of the year, the risk of pregnancy and the risks of children are high. Thus, using information on parents birth months, such a policy can be implemented very effectively.

In addition to immediate implications for policy, our study has several academic implications. First, our study shows that a selection mechanism exists in the dimension other than height and that it plays a substantial role in explaining the variation of educational outcomes. Our study shows that a one percent increase of the selection increases the years of schooling duration premium by 0.6 percent. Note that those who were born in the second half of the year are negatively affected from the point of nutrition and malaria infection. Thus, the fact that the selection increases the schooling duration premium is surprising. Secondly, our findings have an important implication for evaluation of negative shocks and evaluation of a policy. For example, using the rainfall shock data during the gestation period in Indonesia, Maccini and Yang (2009) examine the effect of rainfall shock on socio-economic outcome in later life. He could not find any systematic negative effect regarding the exposure to negative shock in utero. However, our findings suggest that an embryo or fetus which experienced negative shock in utero might have died during the gestation period or the period after birth. Only those which have good innate traits might have survived. In such a case, it is possible that researchers find zero effect of negative shock in utero.

Similarly, in many government projects in Africa, often zero net effect or a very small effect of policy intervention is reported (Banerjee et al., 2011). However, our findings suggest that zero effect of policy can be due to the presence of a selection mechanism. When health is improved by government intervention, some individuals who would have died without intervention start to survive. In the presence of the selection mechanism, in such a situation, the mean effect of the policy can be zero or even be negative. This implies that when researchers evaluate the health or nutrition policy of intervention, it is important to take the effects of the selection into consideration.

References

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

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

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

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

Banerjee, Abhijit V, Abhijit Banerjee, and Esther Duflo, *Poor economics: A radical rethinking of the way to fight global poverty*, Public Affairs, 2011.

Bozzoli, Carlos, Angus Deaton, and Climent Quintana-Domeque, "Adult height and childhood disease," *Demography*, 2009, *46* (4), 647–669.

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

Ceesay, Sana M, Andrew M Prentice, Timothy J Cole, Frances Foord, Elizabeth ME Poskitt, Lawrence T Weaver, and Roger G Whitehead, "Effects on birth weight and perinatal mortality of maternal dietary supplements in rural Gambia: 5 year randomised controlled trial," *Bmj*, 1997, *315* (7111), 786–790.

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

Fews Net, "Malawi," *Famine Early Warning System Network*, 2014.

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

Gørgens, Tue, Xin Meng, and Rhema Vaithianathan, "Stunting and selection effects of famine: A case study of the Great Chinese Famine," *Journal of development Economics*, 2012, *97* (1), 99–111.

Guyatt, Helen L and Robert W Snow, "Impact of malaria during pregnancy on low birth weight in sub-Saharan Africa," *Clinical microbiology reviews*, 2004, *17* (4), 760–769.

Hajison, Precious L, Bonex W Mwakikunga, Don P Mathanga, and Shingairai A Feresu, "Seasonal variation of malaria cases in children aged less than 5 years old following weather change in Zomba district, Malawi," *Malaria journal*, 2017, *16* (1), 264.

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

Kwiatkowski, Dominic P, "How malaria has affected the human genome and what human genetics can teach us about malaria," *The American Journal of Human Genetics*, 2005, *77* (2), 171–192.

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

Ministry of Education, Science and Technology, "National Education Sector Plan 2008 2017," 2008.

Ministry of Health of Malawi, "Malawi Health Sector Strategic Plan 2011-2016," 2011.

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

Moore, Vivienne M, Michael J Davies, Kristyn J Willson, Anthony Worsley, and Jeffrey S Robinson, "Dietary composition of pregnant women is related to size of the baby at birth," *The Journal of nutrition*, 2004, *134* (7), 1820–1826.

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

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

UNAIDS, *Global HIV and AIDS Statistics—2019 Fact Sheet* 2019.

Verhoeff, Francine H, Bernard J Brabin, S Van Buuren, L Chimsuku, P Kazembe, JM Wit, and Robin L Broadhead, "An analysis of intra-uterine growth retardation in rural Malawi," *European Journal of Clinical Nutrition*, 2001, *55* (8), 682.

World Bank, "Malawi Overview," *World Bank*, 2014.

World Health Organization, "Technical Expert Group Meeting on Intermittent Preventive Treatment in Pregnancy (IPTp), Geneva, 11-13 July 2007," 2008.