2 Rainfall, Early Life Health and Education in the Semiarid Northeast Brazil

There is a consolidated view that health at birth has long lasting impacts on human well-being. For instance, twin fixed effects models provide evidence that birth weight has positive effects on infant health, educational attainment and future labor market outcomes (Black et al. (2007), Almond et al. (2005), Behrman & Rosenzweig (2004)). Besides the recent advances on the question of whether early life health affects future outcomes, less is known about why this is so. One likely mechanism is that children in poor health may have poor education, which reduces their adult earnings potential and health later in life (as suggested in Case & Paxson (2010)). In fact, there is supportive evidence that early life health has positive effects on educational attainment (see review in Almond & Currie (2010)). Yet, economists have found that impacts of early life health on future earnings are larger than those on educational attainment (Currie (2009), Black et al. (2007)). As Currie (2009) argues, the lower estimated effect on education suggests that maybe infant health affects future outcomes through mechanisms other than educational attainment. This could be the case if early life health impacts not only educational attainment, but also cognitive development and academic achievement. Many studies in mental health and psychology have suggested impairment in various cognitive and neurodevelopmental tests due to low birth weight and other early life deprivations, even when confounding factors are controlled for (Shenkin et al. (2004), Linnet et al. (2006), OConnor et al. (2000)). However, it is not yet clear whether infant health shocks have causal effects on cognition and learning (Almond & Currie (2010)).

This paper investigates the human capital impact of weather fluctuations in the Semiarid Northeast Brazil. More specifically, we study how and to what extent idiosyncratic rainfall shocks during the time in utero have affected infant health at birth, educational attainment and children academic achievement. The Semiarid Northeast Brazil consists of a relevant case study in the developing world. The region has long been subjected to harsh climate conditions and intermittent events of drought, water scarcity and food insecurity (SUDENE (1981), Ab'Sáber (1999), Áridas (1995)). Poverty plagues 8 in each

10 children, while infant mortality reaches 31 per thousand births.[1] Around 53% of its 20 million dwellers live in rural areas, where poor households unable to insure against idiosyncratic shocks are more vulnerable to adverse natural phenomena.

We organize this study in two separated analysis. First, we examine whether rainfall idiosyncratic shocks during time in utero affect a range of infant health indicators, such as birth weight, number of weeks of gestation and mortality rates. Second, we ask whether these shocks have long-lasting impacts on human capital accumulation. More specifically, we study to what extent the individual's exposure to weather fluctuations while in utero affects her performance at school in childhood, as measured by academic test scores, grade repetition and dropout probabilities.

This research has considerable data requirements. First of all, it necessitates historic information on high-frequency and site-specific weather conditions. We use gridded time series on precipitation and temperature to build a weather dataset in a municipality-by-month level for the Semiarid Northeast. This information is then used to build two parallel datasets. The first one is a municipality by month of birth panel on weather conditions and health. It combines information on rainfall fluctuations before birth with newborns health indicators and infant mortality rates. The second one is a dataset of pooled cross-sections at the student level, for different cohorts. It contains education outcomes linked with the individual's exposure to rainfall fluctuations while in utero.

Our identification strategy relies on the fact that temporary rainfall deviation from its historic average is uncorrelated with any latent determinants of infant health and human capital. Thus, we are able to identify the causal effects of rainfall variations on early life outcomes. This strategy follows closely the ones adopted in related papers, as discussed below.

We find that areas hit by positive rainfall fluctuations experience in the following months lower incidence of low birth weight, preterm gestation and infant mortality rates. The results indicate that a 10% higher level of rainfall (relative to the local historic average) is associated with a drop in infant mortality rates of between 0.53 and 0.76 standard deviation of the mean, or 44 and 63 per 1,000 births. This effect is particularly driven by a drop in mortality due to intestinal infections and malnutrition. We also find that rainfall before

[1]Municipality averages. Authors' calculations for infant mortality rates up to 1 year old based on data from Datasus. Ipeadata calculations for poverty figures based on the 2000 Census. The exact municipality average of the headcount poverty ratio is 81.4%, which relates to the share of individuals aged between 0 and 14 years living in households with per capita income below the poverty line.

birth is associated with higher academic test scores, and lower incidence of delayed enrollment, dropout and grade repetition among young students (5th graders). Finally, we do not find any significant impacts on test scores among older students (9th graders). As we argue, this result is likely driven by sample selection due to improved educational attainment of rural students in later grades.

Our results provide a novel support for the mechanisms linking infant health and future outcomes. First, we reinforce previous results on the impact of prenatal health insults on educational attainment. Second, we show that these detrimental effects are not restricted only to attainment, but also reach academic achievement in childhood. Third, we provide evidence that improvements in overall educational attainment are potentially driven by lower dropout probabilities of students from more deprived socioeconomic status, so that selection into education also play an important role in this process. This can help us to better understand why early life health impacts on educational attainment have appeared to be smaller than those on labor market outcomes.

There is only very scant empirical evidence on the causal link between early life health and student achievement. The most appealing evidence comes from an unusual empirical setting. Almond et al. (2009) show that children born in regions of Sweden more exposed to the Chernobyl fallout performed worse in secondary academic tests. Twins and sibling fixed effects models also provide some evidence, although mixed. Lin & Liu (2009) show by using twin fixed effect estimates that birth weight has a positive impact on grades at age 15. On the other hand, Oreopoulos et al. (2011) find little effect of infant health on language arts test score at age 17. Conley et al. (2007) find no impact of birth weight on cognitive assessments when using sibling fixed effects.

This paper also speaks to a recent stream of research that studies the causal microeconomic effects of weather fluctuations on health and education. Deschenes & Moretti (2009) find impacts of extreme temperature fluctuations on overall mortality in the United States. Deschenes et al. (2009) show also in the US that exposure to extreme hot temperatures during pregnancy leads to lower birth weight. More closely related to our work on precipitation shocks, Kudamatsu et al. (2010) estimate the impacts of rainfall fluctuations on overall infant mortality in Africa. The authors show that infants born in arid areas who experienced extreme events of rainfall shortage when in utero face a higher risk of death. Maccini & Yang (2009) find that higher rainfall in early childhood leads to long run improved health, schooling and socioeconomic status for adult women in Indonesia.

Our paper expands on the results of this previous work in the following ways. First, we examine the effects of rainfall fluctuations during time in utero on an improved range of infant health indicators, which includes birth weight and mortality by cause of death. Unlike the previous literature, we also study these effects on children education. Second, we provide new evidence that contradicts some mixed existing results. Maccini & Yang (2009) find that higher precipitation in the first year of life has positive long run outcomes. However, no significant impact is found for rainfall fluctuations exposure when in utero. Kudamatsu et al. (2010) find a puzzling association between higher rainfall exposure when in utero and higher infant mortality in arid areas. They also estimate positive impacts of severe episodes of drought on mortality. In this paper we do find positive impacts of severe drought on infant mortality. On the other hand, we also estimate negative effects of more typical rainfall shortage during time in utero, differently to Maccini & Yang (2009) and Kudamatsu et al. (2010). Thus, our results lay in between the current evidence and raise concerns about generalizability.

This paper is organized as follows. In the next section we discuss conceptual issues. Section 2.2 describes the empirical context, the dataset, and provides descriptive statistics. Section 2.3 details our empirical strategy, while Section 2.4 presents the results. Section 2.5 concludes.

2.1 Conceptual Background

Fetal growth, the length of gestation and birth weight are usually health variables associated with improved health outcomes of offsprings. Fetal growth is principally regulated by fetal nutrition. Maternal nutrition does not necessarily lead to changes in fetal nutrition, as the fetus lies at the end of a long supply line (Bloomfield & Harding (1998)). However, where there is significant maternal undernutrition, fetal substrate may not meet fetal demand, necessitating a slowing of the fetal growth trajectory (Bloomfield et al. (2006)). Low maternal body-mass index and intrauterine growth restriction are considered risk factors for neonatal conditions. According to medical reviews, poor fetal growth is rarely a direct cause of death, but rather can contribute indirectly to neonatal deaths, particularly those due to birth asphyxia and infections (sepsis, pneumonia, and diarrhoea), which together are estimated to account for about 60% of neonatal deaths in the world (Black et al. (2008)).

Fetal growth trajectory and birth weight are not only associated with early-life physiological status, but can also have long run pervasive effects. The fetal origins hypothesis proposes that environmental influences while the fetus is in utero cause permanent changes in organs development and pre-disposition to chronic diseases in adult life (Barker (1998a), Barker (1998b)).

One mechanism behind this relation is suggested to be fetal undernutrition during critical periods of rapid cell division. Among other disruptive effects, fetal protein malnutrition affects brain development (as reviewed in Morgane et al. (1993)). In fact, while causal evidence is faulty, many studies in mental health and psychology have suggested impairment in various cognitive and neurodevelopmental tests due to low birth weight and other early life deprivations, even when confounding factors are controlled for (Shenkin et al. (2004), Linnet et al. (2006)).

In order to give a formal treatment to this discussion, we present a simple analytical framework for the relationship between early life health and education outcomes.[2] We begin with a health production function. Consider a two-period model, where $t = 1$ is the time at conception and $t = 2$ is the childhood period. Health at time $t = 1$ is determined by genetic characteristics G, as well as by idiosyncratic environmental circumstances R, household socioeconomic status in $t = 1$, W_1, and community characteristics in $t = 1$, X_1 (such as infra-structure, local government policies etc). Thus, health at conception is described by

$$H_1 = H(G, R, W_1, X_1) \tag{2-1}$$

In this paper we focus on the impacts of rainfall fluctuations. Thus, environmental circumstances R are supposed to improve human well-being in important ways through higher levels of precipitation. For instance, positive rainfall variations increase agricultural productivity and the supply of food and nutrients at the local community (Suliano et al. (2009), Maccini & Yang (2009)).[3] It also increases the supply of drinkable water, which helps to avoid the burden of diseases due to parasitoses (Parry et al. (2007), Luna (2007), Kudamatsu et al. (2010)). Hence, rainfall fluctuations are directly associated with maternal health via the supply of food, safe water and the disease environment.[4]

Health in childhood can be analogously described by the function $H_2 = H(H_1, W_2, X_2, I_{h,2})$, where $I_{h,2}$ is family investments in infant health between conception and childhood. We suppose that $I_{h,2}$ varies with H_1 once parents can react to lower health status at birth investing more in health inputs.[5] As

[2]See Glewwe & Miguel (2008) for a complete treatment of the formal links between infant health and education.

[3]For instance, Suliano et al. (2009) estimated that a 1% increase in year precipitation levels is associated with a 0.4% rise in overall agricultural production in the Brazilian State of Ceará. See Maccini & Yang (2009) for a discussion on this issue for rural areas in Indonesia.

[4]The existing research provides descriptive evidence supporting the role of these factors. However, we are not able to identify specific channels in this paper.

[5]Alternatively, families can neglect sick infants and allocate most resources to healthier children. For a formal discussion on responsive investments and the potential bias they might

discussed above, the direct effects of H_1 on H_2 can be both physiological and cognitive.

Finally we describe the production of academic skills A_t in $t = 2$ as a function of health. A simple specification is

$$A_2 = A(G, H_t, W_t, X_t, I_{E,2}, SC, YS \mid t = 1, 2) \tag{2-2}$$

Where $I_{E,2}$ denotes family investments in education inputs, SC is school characteristics and YS is years of schooling attained in period 2. A large body of literature has shown that school attainment as the number of completed grades responds to early life health insults (both in utero or during childhood; see the reviews in Almond & Currie (2010), and Glewwe & Miguel (2008)). A less compelling evidence supports school absences as a mechanism for health to affect education (Currie (2009)); thus, if poor health among children has an effect on acquisition of skills, it is more likely to come through impairing children's ability to learn while they are in school (as proposed by Currie (2009), p.101). This can be the case where idiosyncratic health insults in utero have detrimental cognitive and physiological effects in the postnatal period.

2.2 Data and Descriptive Statistics

2.2.1 Climate and weather in the Semiarid Northeast Brazil

The Northeast Brazil comprises 9 Brazilian states and almost 1,800 municipalities according to the latest national administrative division. The Semiarid Northeast is located in the inner Northeast Brazil. Its official definition includes 1,048 municipalities and follows technical criteria.[6] The municipality is considered officially part of the Semiarid region if at least one of three climate characteristics is satisfied (SUDENE (2008)). First, if it is within the boundaries of the 800mm isohyet, i.e., the line on a map joining points of historic average precipitation below 800mm (yearly precipitation records between 1961 and 1990). Second, it it has average Thorntwaite Index below 0.50. This indicator uses indices such as a humidity index and an aridity index to determine an area's moisture regime based upon its average temperature, average rainfall, and average vegetation type. The lower the value of the index in any given area, the drier the area is. Third, if it has an index of risk of drought

generate, see Almond & Currie (2010) and Glewwe & Miguel (2008). In this paper we are not able to control for intra-household allocation.

[6]This definition is publicly available and has been recently reviewed. To see the new municipalities included in the reviewed official classification see the Ministry of National Integration Portaria nº 89/2005.

above 60%. This index is the share of days under hydric deficit (which balances daily precipitation and evapotranspiration) between 1970 and 1990. Figure 2.1 presents a map of the Northeast Brazil and its Semiarid region highlighted.

The Semiarid Northeast Brazil is one of the most deprived regions in the world. Poverty plagues 8 in each 10 children, while infant mortality reaches 31 per thousand births.[7] Around 53% of its 20 million dwellers live in rural areas. Municipality population has median size of around 12 thousand inhabitants, and mean of 22.4 thousand. Only a few large cities push the mean upwards (only 14 cities have population size above 100 thousand).[8] The Semiarid economy is heavily driven by small-scale agricultural and cattle activities (Áridas (1995)). Productivity is usually low and dependent on weather fluctuations (Ab'Sáber (1999), Áridas (1995), SUDENE (1981)).

It is conventional wisdom among Brazilian scholars and policy makers that the Semiarid region has long been subjected to harsh climate conditions. However, systematic empirical evidence for that is surprisingly scarce. In this paper we raise historic data on precipitation and temperatures for the region by using the *Terrestrial Air Temperature and Precipitation: 1900-2008 Gridded Monthly Time Series, Version 1.02* (Matsuura & Willmott (2009)). This dataset provides worldwide monthly mean temperature and precipitation data at 0.5×0.5 degree resolution. The number of nearby weather stations that influence a grid-node estimate is 20 on average.[9] Each municipality in our Semiarid sample is connected to this weather dataset in two steps. First, each municipality is located within a square defined by the four closest nodes to its centroid's latitude and longitude. Henceforth, we call this square simply by *grid*. Second, monthly mean precipitation and temperatures are calculated for each municipality as the weighed average of the respective weather information recorded at each of the four nodes of this *grid*. The calculation of municipality means is weighed by the linear distances from its centroid to each node.

Figure 2.2 shows time series depicting average annual precipitation across the Semiarid municipalities in comparison to the annual averages across all the Brazilian municipalities, excluding the Semiarid ones. The period comprises 70 years from 1938 to 2008. As we can see, average precipitation in the

[7]Municipality averages. Authors' calculations for infant mortality rates up to 1 year old based on data from Datasus. Ipeadata calculations for poverty figures based on the 2000 Census. The exact municipality average of the headcount poverty ratio is 81.4%, which relates to the share of individuals aged between 0 and 14 years living in households with per capita income below the poverty line.

[8]According to the 2000 Census.

[9]This dataset has been originally used in economics by Dell et al. (2008). Another well-known dataset at the 0.5×0.5 degree resolution based on weather station observations is the one developed by the Climate Research Unit at the University of East Anglia (CRU). However, this dataset was not available by the time of our research.

Semiarid has been about half the national average, and just slightly lower the 800mm mark. We can also see that events of extreme rainfall scarcity have been intermittent throughout the past decades. This is indeed consistent with existing descriptive evidence in the literature. Various authors have documented episodes of historical droughts, such as those occurred in the early 1950s, 1958, 1970, early 1980s, early 1990s and 1998 (SUDENE (1981), Campos (1994), Villa (2000). All these marks are shown in Figure 2.2, which helps to validate our database.

The pattern of weather fluctuation within years is shown in Figure 2.3. There are markedly two main seasons per year. The first semester is the rainy season. Precipitation levels are particularly higher between March and May, when harvest takes place more often. The first semester can be defined also as the hunger season, when food stocks are low and the next harvests are still to come. The second semester is invariably dry, and monthly precipitation levels are often close to zero. Episodes of drought occur when precipitation levels during the first semester are unexpectedly low and irregular. In this situation food security and water supply are jeopardized.

2.2.2 Prenatal exposure to weather fluctuations

In this paper we use two measures of rainfall fluctuation in order to study the relationship between prenatal exposure to weather shocks and early life development. The first measure is defined by the equation

$$R_{m\tau} = Ln\left(\sum_{t=-11}^{\tau} R_{mt}\right) - Ln\left(\overline{R}_m\right) \tag{2-3}$$

Where τ is the individual's month of birth, and the subscript m is the municipality of birth. Thus, we define $R_{m\tau}$ to be the deviation of (i) the natural logarithmic of the total rainfall in the past 12 months up to the individual's birth to the (ii) natural logarithmic of the historic average of the annual rainfall in m, $\overline{R}_m$. This historic average is calculated for each municipality over the years between 1938 and 2008. The measure $R_{m\tau}$ should be interpreted as the approximate percentage deviation from mean rainfall. For instance, a value of 0.01 means that rainfall over the year before birth was around 1% higher than normal. Measures of temperature deviations are constructed analogously, with minor changes.[10]

[10]The measure of temperature deviations is used as a control variable and is defined in the section of empirical strategy.

The second measure is a dummy variable that indicates events of more severe deviation from the norm. We define an episode of drought by using the equation

$$D_{m\tau} = 1, if \sum_{t=-11}^{\tau} R_{mt} < \left(\overline{R}_m - R_m^{SD}\right) \tag{2-4}$$

Where R_m^{SD} is the annual standard deviation of rainfall calculated for each municipality over the years between 1938 and 2008. The value of $D_{m\tau} = 1$ indicates that rainfall over the year before birth was more than one standard deviation below the norm.[11]

We use both measures 2-3 and 2-4 in Figure 2.4 to show that the incidence of rainfall shocks has varied significantly over time and place in the Semiarid. Panel A displays the measure of rainfall log-deviation averaged across municipalities for each month over the period 1938-2008. This variable has mean -0.044, median -0.029 and standard deviation 0.31. In Panel B we display the time series of the measure of drought, which shows that severity of shocks varies geographically for a given month. On average, episodes of drought for a given month occur in 13.7% of the municipalities. However, we observe there have been periods with pervasive events of drought in the Semiarid, hitting more than 50% of the municipalities.

2.2.3 Gestation, birth weight and infant mortality

The dataset on health at birth and infant mortality is built by combining microdata from the National System of Information on Birth Records (Datasus/SINASC) and the National System of Mortality Records (Datasus/SIM). The first one gathers information on every event of birth occurred in Brazil since the mid-1990s. Only for the Semiarid region the number of birth events recorded between 1996 and 2008 is around 13.7 million. This information includes at the newborn level her birth weight, length of gestation, and other health indicators such as APGAR indices. A few variables containing the mother's socioeconomic characteristics are also available, such as occupation and education, but missing cases are frequent up to the mid-2000s. Finally, the dataset also provides the exact date of birth, the municipality of birth and the mother's municipality of residence. All this information allows us to construct a

[11]Both measures are closely related to the ones used in Maccini & Yang (2009), once these authors also construct the logarithmic of rainfall deviations from the norm. However, while we calculate our measures on a monthly basis, the authors used rainfall fluctuations aggregated into seasons (6 months). We believe that our approach leads to a more robust test once it relays on smoother weather variation between months.

municipality-by-month of birth panel dataset over the period 1996-2008. Thus, for each municipality and month over the entire period we calculated the total number of births and cell means for birth weight and other health variables.

The system of mortality records gathers information of every event of death officially recorded in Brazil. This dataset contains information on cause of death and some socioeconomic characteristics of the individual. It also contains data on date of birth, municipality of birth and municipality of residence. We selected all the death events of individuals up to one year old in the Semiarid region between 1996 and 2008. During this time there were around 378 thousand infant deaths in the region. We then build an analogous municipality-by-month of birth panel dataset over the period 1996-2008. For each municipality and month of birth over the entire period we calculated the total number of infant deaths, and totals per cause of death.

These panels on birth and infant mortality are then merged by municipality of residence and month of birth. The new dataset allows us to calculate infant mortality rates by municipality and month of birth. Finally, we combine this dataset with our weather dataset. In doing so, we link the health and mortality indicators by month of birth and municipality of residence with site-specific measures of rainfall over the months before a given month of birth. Table 2.1 presents the summary statistics of this final dataset, where all statistics are averages across municipalities. As we can see, the average number of births per month across municipalities is 28.9, although standard deviation is high (the median number of births is only 14). Average birth weight is 3,721, while 94% of the pregnancies lasted 37 or more weeks. The average number of infant deaths per month is about 0.84. Infant mortality rates across municipalities is on average of 31.26 per 1,000 births. In this sample, the average incidence of rainfall in the 12 months before birth is 800mm. Average rainfall log-deviation is positive in 2%.

2.2.4 Education and student achievement

We build two separate datasets at the individual level linking weather and educational outcomes. The first one uses education microdata from the Prova Brasil, a national standardized exam applied to 5th and 9th graders in each two years. The exam is composed of two tests that measure math and language skills. In addition, students answer a survey about their past school achievement and socioeconomic profile, which includes date of birth, race, sex, mother's education, household conditions and others. Questions about past achievement include whether the student has ever repeated a grade or dropped out. It also includes whether her first enrollment at school was before the first year of elementary school (i.e., day care or pre-school). We use the available microdata at the student level for the years 2005 and 2007. Once we do not have information on municipality of birth, we suppose that children study and dwell where they were born. This is a reasonable assumption for young children. Another caveat relates to the fact that the Prova Brasil was applied in 2005 and 2007 only at urban schools with 5th or 9th grade sizes above 20 (or 30 students, in 2007). This excludes students from small schools situated in inner rural areas, supposedly more prone to health insults due to weather shocks than the urban ones.

This education dataset is then merged with our weather dataset. Each student is linked with the weather conditions prior to her birth at the municipality where she studies. Descriptive statistics are provided in Table 2.2. Our sample includes around 330 thousand 5th graders and 320 thousand 9th graders that took the achievement tests in 2005 or 2007. We see that socioeconomic conditions are poor for both groups of students, but worse for 4th graders. This relates to the fact that enrollment in early grades are close to universality, while dropout throughout time is more common among students from low socioeconomic status.

The second dataset is based on microdata from the Yearly National Educational Census (Censo Anual da Educação Básica, Inep). The Educational Census has included information at the student level since 2007, such as grade of enrollment, date of birth, municipality of birth and residence, and a few other socioeconomic characteristics (sex, race, and whether the student area of residence is rural or urban). Due to data limitations by the time of this research, variables of achievement (grade repetition) and dropout at the student level are constructed only for 2008. This is done by tracking the student identification number between 2008 and 2009. This dataset allows us to better understand potential biases that arise from limitations in the Prova Brasil dataset (for instance, the profile of students enrolled in urban schools in comparison to the universe of pupils). Moreover, we are also able to examine the relationship between weather conditions and student outcomes per grade. This is done by using a dataset that again links each student with the weather conditions prior to her birth at the municipality where she was born. Table 2.3 provides the summary statistics for this dataset. Note that we exclude from our analysis the first and the last grades (1st and 9th grades). This is so due to data limitations. First, because we are not able to disentangle students enrolled in pre-school and literacy grade from first graders. Second, tracking students between the 9th grade and the first grade of the secondary school appeared to be problematic.

2.3 Empirical Strategy

In this section we describe our empirical strategies and the potential caveats. First, we present the baseline econometric specification we follow to identify the health impacts of rainfall fluctuations. In the second and third subsections we focus on the identification of the educational effects. We discuss potential caveats at the end.

2.3.1 Infant health and mortality

The analysis of the health impacts is based on a municipality by month of birth panel dataset. Our baseline specification is given by the reduced-form equation

$$H_{cmy} = \alpha + \beta R_{cmy} + \pi T_{cmy} + \phi_{cm} + \lambda_y + \varphi Trend_{cm} + \epsilon_{cmy} \tag{2-5}$$

Where H_{cmy} is a infant health outcome of municipality c, averaged across those children born in month m and year y. Our variable of interest is rainfall fluctuation R_{cmy}, that varies by municipality, month and year of birth. This variable takes the form of either the measure of past rainfall log-deviation, or the indicator of droughts, as defined respectively by equations (2-3) and (2-4). The variable ϕ_{cm} is a fixed-effect for municipality c in calendar month $m = 1, 2, ..., 12$. This specification follows closely the one used in Kudamatsu et al. (2010). In doing so, we control for 12 monthly means in each of our 1,048 municipalities, which results in more than 12,5 thousand fixed-effects. Thus, we identify the parameter of interest β from the rainfall deviation within each municipality from its monthly mean. We also use the year fixed-effect λ_y to allow for a non-parametric trend of infant mortality in the Semiarid region. The term T_{cmy} controls for variation in site-specific temperature deviations from the norm.[12] Thus, we isolate the rainfall channel from another potential confounding weather effect. Finally, in full specifications we also control for the term $Trend_{cm}$, which includes linear time trends specific to the grid (monthly time trends) and the state (yearly). These trends should absorb long run trajectories in the outcome that can vary across municipalities or states.[13]

[12] We build the measure of temperature deviations analogously to the rainfalls one, with minor alterations. First, in the formula of equation 2-3, we use mean temperature over months instead of the sum of precipitation over months. Second, we calculate separately temperature deviations by trimesters instead of years. This is done once Deschenes et al. (2009) have shown that temperature variations impact birth weighs only during the second and third trimesters of pregnancy. Thus, T_{cmy} includes exactly two variables: the log of mean temperature deviation from the norm over the first and second trimesters before birth.

[13] As we mentioned before, we call *grid* as the square defined (in our original 0.5×0.5

The term ϵ_{cmy} is an idiosyncratic error term. Given that we build our weather variables at the grid level, serial and spatial correlation in error terms may be at work. Thus, in all specifications we compute robust standard errors clustered at the grid level. Our identification strategy relies on the fact that temporary rainfall deviation from its historic average is uncorrelated with any latent determinants of infant health and human capital. Thus, we are able to identify the causal effects of rainfall variations on early life outcomes.

2.3.2 Student achievement

The main analysis of the educational impacts of rainfall fluctuations is based on a dataset that comprises two pooled cross sections at the student level (i.e., that includes two rounds of the Prova Brasil achievement tests for 5th and 9th graders, in 2005 and 2007). Our main specification is given by the reduced-form equation

$$A_{icmy} = \alpha^* + \beta^* R_{cmy} + X_i'\theta + \pi^* T_{cmy} + \phi^*_{cm} + \lambda^*_y + \varrho PB + \varphi^* Trend_{cm} + \nu_{icmy} \quad (2\text{-}6)$$

Where i is the subscript for the individual i born in municipality c, month m and year y. The dependent variable A_{icmy} is the individual's educational outcome (test scores, for instance). The starred coefficients are analogous to the non-starred ones defined in equation (2-5), with minor changes. Here, λ^*_y is a fixed-effect that controls for the individual's cohort common characteristics (i.e., it is a year of birth fixed effect). The variable PB is a fixed-effect controlling for the Prova Brasil edition. Pooling two editions allows us to compare the performance of different cohorts (therefore, exposed to different weather shocks when in utero) taking the achievement test scores at the same grade, but in different years. Finally, the term X_i includes controls for the individual's socioeconomic status. In all specifications we compute robust standard errors clustered at the grid level. We estimate all regressions separately for 5th and 9th graders.

2.3.3 Educational attainment

We examine educational attainment impacts by using a cross-sectional dataset. As mentioned before, information from the 2008 and 2009 School Censuses allow us to track students between years and create for each student dummy indicators for dropout and grade repetition. First, we suppose the student dropped out between 2008 and 2009 if we are not able to find her

degree weather dataset) by the four closest nodes to the municipality of birth centroid's latitude and longitude. Thus, this time trend is common to all municipalities included in a given *grid*.

identification number in the 2009 Census dataset. Second, we suppose the student repeated a grade if we find her identification number enrolled in the same grade in both years. We select into the sample only the students who are in the correct grade for their ages. This is done in order to eliminate confounding rainfall effects on past attainment. A given child enrolled in the 5th grade in 2008, for instance, is included in the sample if she was born between July 1998 and June 1999. Because we have only a cross section dataset, the main specification is now given by the simpler linear equation

$$A_{icmys} = \alpha^* + \beta^* R_{cmy} + X_i'\theta^* + \pi^* T_{cmy} + \phi_c^1 + \phi_m^2 + \lambda_y^* + \varphi^* T + \nu_{icmy} \quad (2\text{-}7)$$

The main difference between model (2-6) and (2-7) is that in the later we separate the fixed-effect for municipality c in calendar month m into (i) municipality of birth fixed-effects ϕ_c^1, which means a dummy variable for each municipality; and (ii) month of birth fixed-effects ϕ_m^2, which means the introduction of a dummy variable for each calendar month. This is so because we are not able to compare the outcomes of individuals enrolled in the same grade, born in the same calendar month, but in different years. Second, in this model the vector X_i includes only a few socioeconomic variables (such as sex, and place of residence in rural *vs* urban areas). Yet, we are now able to control for other important individual and school characteristics. For instance, whether the child studies and/or live in the same municipality where she was born, whether the school is located in an urban area, its pedagogic system, its size in terms of the overall number of students and per grades, and other variables.

2.3.4 Caveats

Potential biases may arise in our analysis. First, random measurement errors in weather data can bias estimated coefficients towards zero. Maccini & Yang (2009) raise this issue and identify the bias direction by using instrumental variables regressions. The authors use variables for rainfall measured at more distant weather stations as instruments for rainfall in the individual's birthplace and year of birth. They do find attenuation bias in OLS estimated rainfall coefficients. In this research we maintain our baseline linear specifications and take for granted the possibility of attenuation bias.

Another source of attenuation bias concerns fetus selection due to adverse weather conditions. This caveat is recurrently mentioned in the birth weight literature (Currie (2009), p.106). The problem is that only surviving fetuses are recorded in our dataset. Hence, shocks which tend to cull weak fetuses might lead the population of surviving newborns to be stronger than it would have been otherwise. As Currie (2009) argues, the fetal selection argument suggests that estimated coefficients may understate the true negative effects of health insults.

Fetus selection joints forces with infant mortality due to weather shocks in order to strengthen the attenuation bias in education regressions. This survivor-bias has long concerned empirical research on the relationship between health and economic welfare (Gorgens et al. (2011)). For example, Friedman (1982) suggests this type of bias as a possible explanation for the increased height of slaves in Trinidad. Bozzoli et al. (2009) find that population height increases with the mortality rate for countries whose infant mortality rate exceeds a threshold level. Gorgens et al. (2011) use data from the 1959-1961 Great Chinese Famine and find that taller children were more likely to survive the famine. They also find no apparent pattern of stunting amongst famine cohorts. However, when controlling for selection, the authors estimate that children who survived the famine grew up to be shorter.

Another potential caveat relates specifically to our educational data on achievement tests for 9th graders. As mentioned above, Prova Brasil was taken only by urban schools in 2005 and 2007. Young students have access to early grades at a large number of relatively smaller schools in either urban or rural areas. For instance, as shown in Table 2.3, more than 40% of all Semiarid young students are enrolled in rural schools (up to 5th grade). However, the number of schools that offer the second term of the elementary grades (from 6th to 9th grades) is otherwise smaller and more concentrated in urban areas. For this reason, only 25% of the older students enrolled in late grades study in rural schools. Moreover, around half the students change schools between the 5th and 6th grades. Not surprisingly, the share of rural students enrolled in urban school rises from around 10% in the first term to more than 25% in the second term. Thus, heterogeneous attainment or mobility impacts throughout grades can result in sample selection among the 9th graders. We investigate further this hypothesis in the next section.

2.4 Results

2.4.1 Infant health impacts

Table 2.4 presents the core results for the relationship between rainfall fluctuations and overall infant mortality rates. The regressions follow equation (2-5). The first column is our baseline specification, which includes fixed effects for municipalities in calendar months, year of birth fixed-effects and state time trends. Panel A shows that rainfall during the time in utero (positive log deviations, as defined in equation (2-3)) has a negative and significant relationship with infant mortality rates. In column 2 we introduce grid time trends, while in column 3 we additionally include controls for past temperature deviations. In both specifications we find a similar magnitude for the estimated coefficients.

The two following columns deal with concerns about outliers. On the one hand, rainfall effects may be less pervasive in large municipalities, where urban structure and health services should be more easily accessible. In column 4 we exclude municipalities where the average number of births per month is above the 99 percentile of its distribution over the period (i.e., municipalities where the average number of births exceeds 258 newborns per month). The result is quite similar to the previous ones. On the other hand there is some concern that municipalities with a quite small number of births per month could introduce noise in the estimations. For instance, this can be the case if small municipalities are also those that underreport (or misreport) the true number of births. Thus, in column 5 we additionally use a weighed regression (weighed by the average number of newborns per month). This full and weighed specification reports a stronger negative effect. The magnitude of the coefficient is substantial. A 10% increase in rainfall leads to drop in the mortality rate of approximately 63 per 1,000 newborns, which means a fall of 0.77 standard deviation of the mortality rate. This result supports the view that variations in infant mortality in the Semiarid region are strongly associated with rainfall fluctuations.

Panel B presents the results for regressions that use our measure of droughts as variable of interest (as defined in equation (2-4)). We find positive effects, which is consistent with Panel A. However, robust coefficients are found only in the most complete specifications.

We also investigate the timing of the rainfall effects. Medical studies of the relation between maternal health and birth weigh suggest that the period before and around conception is important for the fetus growth trajectory. Birth weight of a newborn is reported to be correlated with the pre-pregnancy weight of the mother (as reviewed in Bloomfield et al. (2006)). The early postnatal conditions can be also critical once infant health is vulnerable to shortages of potable water and the disease environment. On the other hand, as argued by Kudamatsu et al. (2010), breast-feeding is known to lower mortality risk. As long as it is not very severe, maternal malnutrition has little impact on the volume and composition of breast milk (Brown & Dewey (1992) apud.

Kudamatsu et al. (2010), p.19).

Thus, lag and lead rainfall fluctuations may be at work in our empirical setting, and need to be addressed. We calculate lag rainfall fluctuation as the difference between the log of the rainfall over the period between 24 and 12 months before the month of birth and the historic average. That means the rainfall fluctuation that occurred before the time in utero. Lead rainfall is defined analogously, but relates to the rainfall in the next 12 months after the birth. The last column of Table 2.4 presents the results of regressions that also include controls for lag and lead rainfall. We see that lag and lead adverse rainfall fluctuations are indeed negatively associated with infant mortality. However, while rainfall during the time in utero is still significant, the effects of lag and lead fluctuations appear not to be robust.

In Table 2.5 we study the rainfall effects on other infant and maternal health indicators. All regressions follow our preferred specification, which includes the full set of controls for time and place fixed-effects, temperature deviations, time trends and outliers. In the first column we see that rainfall has a positive and significant relationship with birth weight. A 10% increase in rainfall is associated with more 70 grams in birth weight. The second column shows another consistent result. It shows that rainfall has a positive effect also on the gestation length. The share of pregnancies that last more than 36 weeks (i.e., that are not pre-term) increases, although the magnitude of the effect is small. Panel B shows that these results are also robust when the measure of drought is used instead of the rainfall log-deviation.

We then investigate the relationship between rainfall and infant mortality per cause of death. In columns 3 and 4 we find robust rainfall effects on newborn deaths due to intestinal infections and malnutrition. This is consistent with the hypothesis that rainfall increases food availability and safe drinkable water, which improves the maternal and the newborn health status. Thus, healthy newborns should be more resistant to health insults in early life. Interestingly, we also find a negative and significant effect on deaths due to other non reported causes. These reports are usually associated with non traumatic infant deaths that occurred without any kind of medical assistance (at home, for instance). This can be more recurrent among poor families that live far from hospital facilities. Finally, we do not find any effect on deaths due to congenital effects.

2.4.2 Academic achievement impacts

Table 2.6 presents the core results for the relationship between rainfall fluctuations and achievement test scores. The regressions follow equation (2-

6). Panels A and B display the results for 5th graders. The first column is our baseline specification. As we can see in Panel A, rainfall during the time in utero has a positive and significant relationship with achievement test scores in math. Column 1 shows the result from our baseline specification, while the second column includes grid time trends and controls for temperature deviations. Column 3 additionally includes students socioeconomic characteristics, which lead to an increase in the estimated coefficient. This result suggests that once socioeconomic status is positively correlated with achievement, rainfall when in utero and socioeconomic status should be negatively correlated. This is consistent with low income students being more selected into the sample due to better rainfall conditions when in utero. In the last column we see that the results are robust to controlling for outlier municipalities.[14] Panel B displays consistent results of regressions that use the drought indicator instead of the rainfall log deviation. The magnitude of the coefficients found for 5th graders is substantial. A 10% increase in rainfall when in utero leads to an increase of approximately 0.38 to 0.47 standard deviation of the math test scores.

On the other hand, Panels C and D show no significant rainfall effects on test scores for 9th graders. We see in the first 3 columns of Panel C that the estimated coefficients appear surprisingly to be negative, although not significants. Only when we control for outliers in the last column the coefficient changes sign. Before discussing further the potential explanations for this result, we explore in Table 2.7 the relationship between rainfall and other educational variables available in the Prova Brasil survey. In all regressions we use the full and weighted specification.

In the first two columns we investigate whether rainfall has impacts on child labor. In the Prova Brasil survey students were asked whether they worked outside the household (dummy variable) and how many hours they performed any kind of domestic work (categorical). We do not observe any significant impacts for either 5th or 9th graders. In the third column we use as dependent variable whether the student first enrolled in school at daycare or pre-school (or instead, only in the first grade of elementary). Column 3 shows that rainfall has positive and robust effects on early enrollment for 5th graders. However, we find a negative and also robust effect for 9th graders. This result suggests that sample selection can be a serious concern. The samples of 5th and 9th graders respond differently to a question that should be independent

[14]Municipalities are considered outliers if the total number of students taking the achievement tests in both 5th and 9th grades, in a given year, is above 2,980. This threshold marks the 99 percentile of the distribution of municipalities in terms of the number of students taking the exam. Moreover, weighted regressions use as weight the number of students taking the exam per school. This might smooth potential noise coming from smaller schools.

of actual grade of enrollment. In Columns 4 and 5 the dependent variables are respectively dummy variables indicating if the student has ever repeated a grade, and whether she has ever dropped out. As we can see, rainfall is associated with less repetition and dropout among 5th graders, while the opposite appears to be true for 9th graders.

Next section investigates the relationship between rainfall and school attainment in urban schools by using our Educational Census dataset. This exercise helps us to better understand the puzzling results found for 9th graders. In particular, we examine the following selection hypothesis. First, positive rainfall shocks might select a larger number of students from more deprived households into the sample because of better health conditions. This can be the case for rural students, which are supposedly more prone to the risks associated with rainfall fluctuations. Thus, while rainfall improves health status and achievement, it may also select more students from low socioeconomic status into the sample. This should attenuate the positive effect of rainfall on achievement. However, we still need to explain why we should expect a stronger attenuation bias for 9th graders. As Table 2.3 shows, access to primary education (1st to 5th grades) is widespread in rural areas. However, the supply of fundamental education (6th to 9th grades) is more concentrated in urban schools. This explains why the share of rural students enrolled in rural schools drops from 40% in the 5th grade to 25% in the 6th grade. Analogously, the share of rural students enrolled in urban schools rises from 10% to 27% between the 5th and 6th grades. If attainment among rural students is improved due to better health status, a relatively larger number of rural students should inflow into the later grades of urban schools because of early rainfall. If rural students have poorer school achievement records in general, rainfall should be associated in later grades with less achievement in urban schools because of selection. Finally, because dropout rates usually increase with grade (see Table 2.3), the marginal effects of rainfall on the selection of students enrolled in later grades can be reinforced. Thus, because coverage is very high at early ages, the effect of rainfall manifests itself more strongly through achievement performance. At later ages, coverage is much lower, what makes selection effects relatively more important.

2.4.3 Educational attainment impacts and selection

The analysis of the relationship between rainfall fluctuations and educational attainment follows equation (2-7). We restrict our sample only to students enrolled in urban schools. This is done to make the results from this analysis comparable with those from the achievement regressions. We adopt grade fixed effects in all regressions that include students from more than one grade. Table 2.8 presents the impacts of early life rainfall on dropout. The first column presents the results of regressions that include students enrolled in all grades of urban schools. In Panel A we see that rainfall has a negative effect on the probability of dropping out. However, this effect is heterogeneous across grades. In the remaining columns we see that this result is entirely driven by the impacts of rainfall on dropout probabilities among students enrolled in the fundamental school (i.e., 6th to 8th grades, last column). In Panel A we also highlight that rural students have higher probabilities of dropping out. Again, this effect is stronger and more robust in later grades.

In order to examine the plausibility of the selection hypothesis, we first examine in Panel B the role of the interaction between rainfall and rural students. The results provide evidence that rainfall selects more students from relatively lower socioeconomic status, i.e. rural dwellers, into the sample. First, the coefficient of the interaction is negative. Second, the result is again totally driven by the sample of students enrolled in the fundamental school. Thus, rainfall seems to countervail the rural status effect and helps to retain more rural students in later grades.

We then examine whether rural students have poorer school achievement records in general, and whether rainfall should be associated with less achievement in later grades because of selection. Unfortunately, the Prova Brasil survey does not ask the student whether she lives in a rural or urban area. Yet, we use the School Census to construct a dummy indicating whether the student succeeded in advancing to the next grade between 2008 and 2009. Table 2.9 presents the results of the analysis of the relationship between the probability of passing a grade and rainfall.

The first column of Panel A shows that rainfall has positive effects on the probability of passing. Consistently with the results from the analysis of dropout, we see that this result is driven by later grades. Yet, rural status has now pervasive effects on both the elementary and the fundamental grades. In Panel B, we examine the interaction between rainfall and rural students. Interestingly, we find opposite results for elementary and fundamental grades. Differently to the younger ones, those rural students enrolled in the fundamental grades and exposed to positive rainfall shocks have lower performance.

Jointly with the results for dropout, the evidence is so far supportive of the following selection hypothesis. Rural students that otherwise would have dropped out of the fundamental school, are kept attending school because of better health status. This effect should attenuate the positive achievement impacts of rainfall in our estimations once rural students are lower performers in general. Yet, there seems to exist an additional effect. Rainfall appears to select not only more rural students into the sample, but also to select relatively more the most deprived amongst the rural student. This effect may reinforce the potential effects of sample selection. This hypothesis is consistent with the results found in the achievement regressions shown in the previous section.

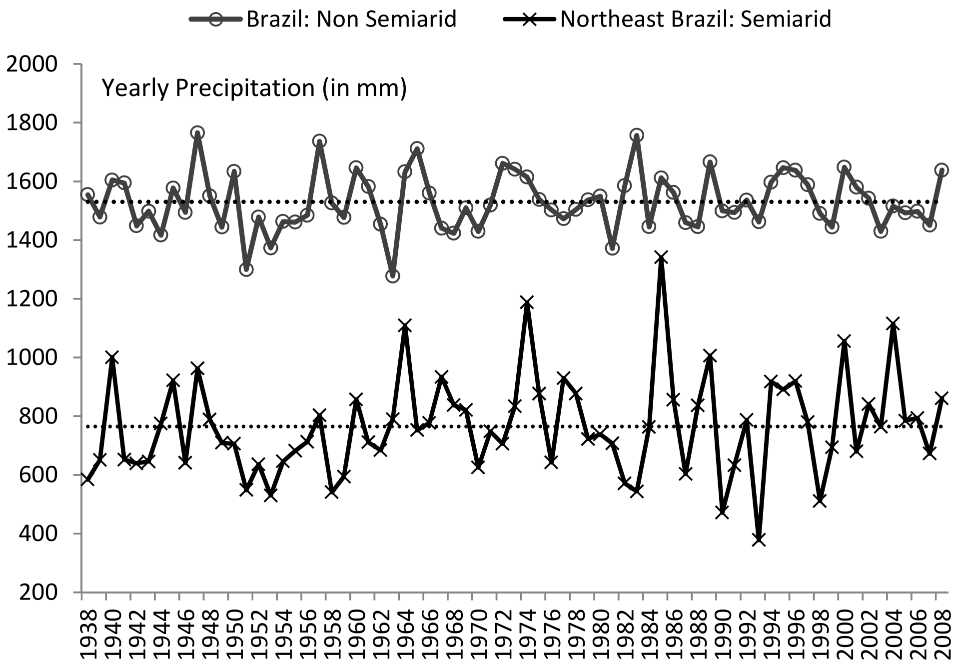
2.5 Final Comments

This paper contributes to the literature on labor and development economics with two main sets of results. The first one concerns more centrally the literature on infant health impacts on human well-being. First, we reinforce previous results on the impact of prenatal health insults on educational attainment. We show that these detrimental effects are not restricted only to attainment, but also reach academic achievement in childhood. We also provide evidence that improvements in overall educational attainment are potentially driven by lower dropout probabilities of students from more deprived socioeconomic status. This can help us to better understand why early life health impacts on educational attainment have appeared to be smaller than those on labor market outcomes. Hence, our results provide a novel support for the mechanisms linking *via education* infant health and future outcomes. As Almond & Currie (2010)'s literature review has recently stated, "*while it is clear that shocks to health have long term effects on domains such as education and earnings, it is not clear whether health shocks have direct effects on cognition or learning, or whether they act mainly by affecting future health*" (see p.56). This paper provides supportive causal evidence linking early life health and childhood learning.

The second set of results relates to the stream of literature that studies the relationship between weather and development. This paper provides statistically significant evidence on a negative relationship between adverse rainfall fluctuations and a range of infant health indicators, as measured by birth weight, gestation length and infant mortality rates per cause of death. Importantly, we find substantial impacts. This set of results suggests that maternal health should be a priority focus of policy interventions during episodes of adverse natural phenomena.

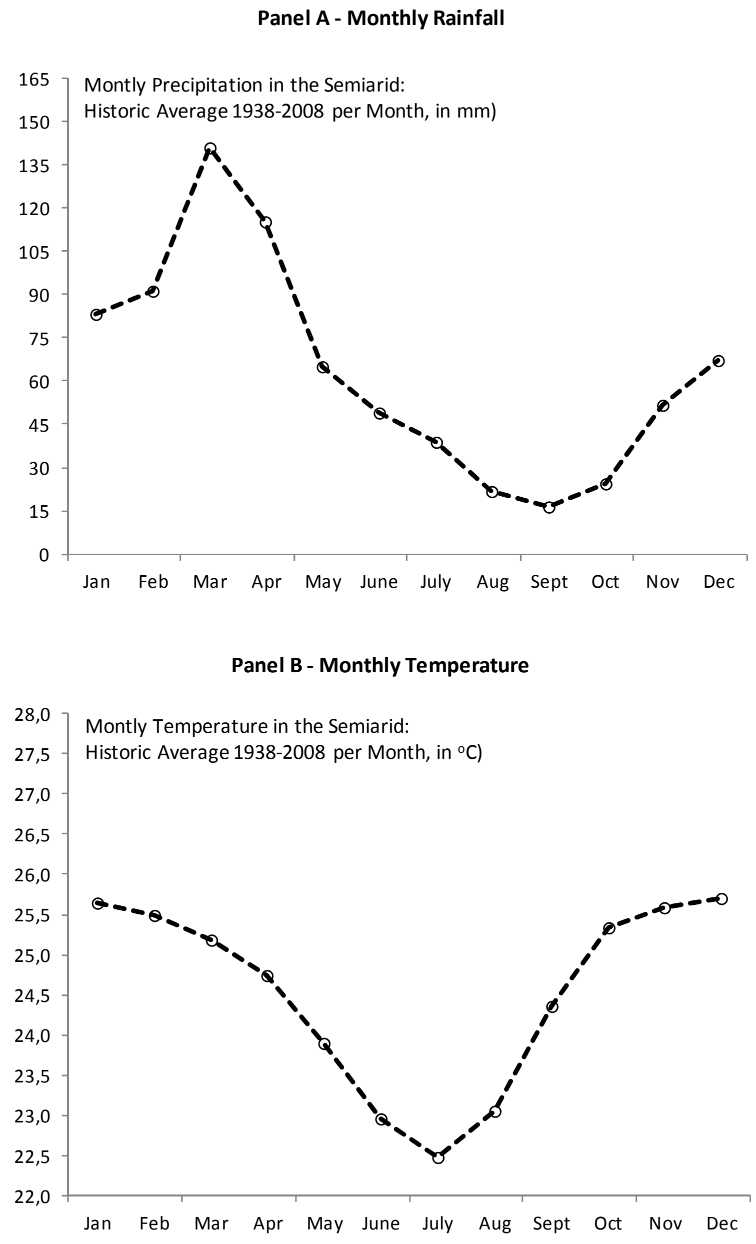
Figure 2.1: Northeast Brazil and the Semiarid Region Highlighted (Source: SUDENE (2008))

Figure 2.2: Yearly Precipitation in Brazil and in the Semiarid Northeast Brazil, Averages Across Municipalities



Notes: This figure depicts the average annual precipitation in millimeters across the Semiarid municipalities in comparison to the annual averages across all the Brazilian municipalities, excluded the Semiarid ones. Original data on precipitation from the *Terrestrial Air Temperature and Precipitation: 1900-2008 Gridded Monthly Time Series, Version 1.02* (Matsuura & Willmott (2009)). Each Brazilian municipality is connected to the weather dataset in two steps. First, each municipality is located within a square defined by the four closest nodes to its centroid's latitude and longitude. Second, monthly mean precipitation and temperatures are calculated for each municipality as the weighed average of the respective weather information recorded at each of the four nodes of this *grid*. The calculation of municipality means is weighed by the linear distances from its centroid to each node.

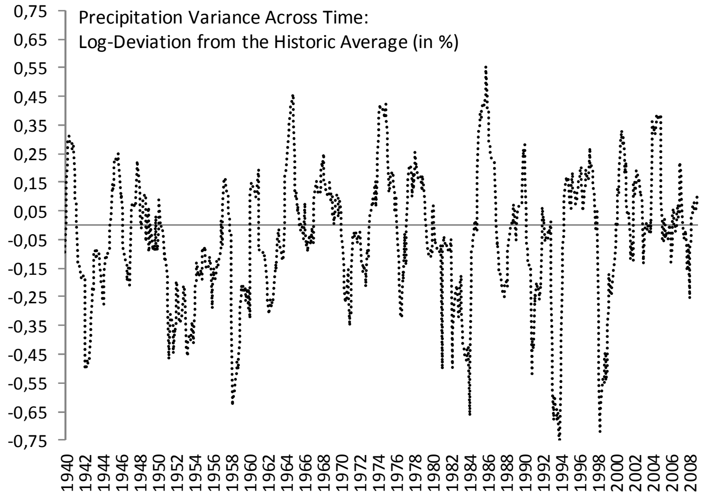
Figure 2.3: Monthly Rainfall and Temperature in the Semiarid Northeast Brazil, Historic Averages



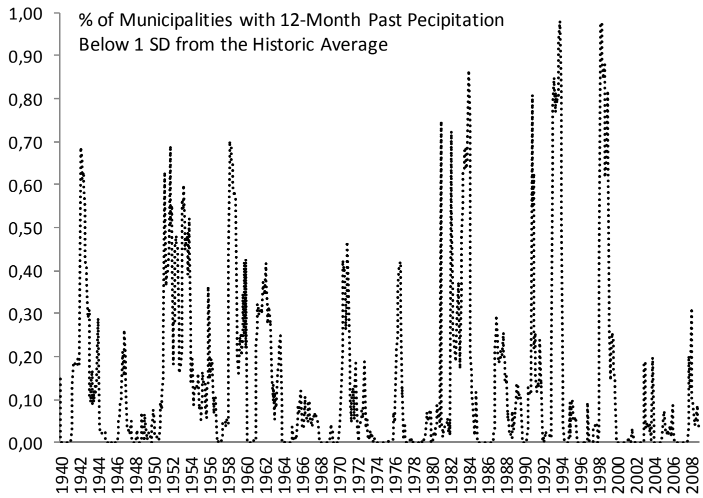
Notes: These figures depict average precipitation (sum in millimeters) and temperature (mean, in degree Celsius) per month. Both trends are calculated by using monthly information at the municipality level between 1938 and 2008. Original data on precipitation and temperature from the *Terrestrial Air Temperature and Precipitation: 1900-2008 Gridded Monthly Time Series, Version 1.02* (Matsuura & Willmott (2009)).

Figure 2.4: Rainfall Idiosyncratic Fluctuations Across Time and Place in the Northeast Semiarid

Panel A - Deviation of Log Rainfall in the Past 12 Months from the Avg.



Panel B - Drought Indicator (Rainfall in the Past 12 Months Below 1 SD the Avg.)



Notes: Variable in Panel A is the average measure of rainfall log-deviation across municipalities. This measured is calculated for each municipality by using equation 2-3 as the deviation of (i) the natural logarithmic of the total rainfall in the past 12 months up to the given month (ii) to the natural logarithmic of the historic average of the annual rainfall in the given municipality. This historic average is calculated for each municipality over the years between 1938 and 2008. Variable in Panel B is the average measure of drought across municipalities. This measure is calculated for each municipality by using equation 2-4, and indicates that rainfall over the year before a given month was one standard deviation below the municipality historic average.

Table 2.1: Summary Statistics: Infant Health and Mortality Rates Across Municipalities, Monthly Data Over the Period 1996-2008

Variables	Mean	Std. Deviation	Min	Max	Number of Municipalities	Number of Observations
Fertility and Health Indicators Per Month, Avgs Across Municipalities:						
Number of Births	28,9	52,4	0	1117	1048	163488
Average Birth Weight	3271	207,0	300	5281	1048	157222
Average % of Births Occurred after 36 Weeks of Gestation	0,94	0,13	0	1	1048	157200
Number of Infant Deaths per Month of Birth	0,84	2,05	0	72	1048	163488
Infant Mortality Rates per Month of Birth (up to 1 year old, *1000):						
Total Infant Mortality	31,26	82,63	0	1000	1048	157403
Due to Intestinal Infections	3,67	29,51	0	1000	1048	157403
Due to Malnutrition	1,09	16,48	0	1000	1048	157403
Due to Pneumonia and Respiratory Infections	1,97	21,48	0	1000	1048	157403
Due to Affections of Perinatal Origin	14,02	51,88	0	1000	1048	157403
Due to Congenital Malformations	2,28	20,56	0	1000	1048	157403
Due to Non Reported Causes	7,37	45,89	0	1000	1048	157403
Rainfall Indicators per Month, Averages Across Municipalities:						
Rainfall in the Past 12 Months	800,3	256,80	80,89	2299,4	1048	163488
Deviation of Log Rainfall in the Past 12 Months from the Log Historic Avg	0,02	0,29637	-1,77	0,96	1048	163488
Drought (Rainfall in the Past 12 Months Below 1 SD the Historic Avg) (0/1)	0,10	.	.	.	1048	163488

Notes: This table presents health and weather indicators averaged at the municipality/month level over the period 1996-2008. Authors' calculation based on microdata originally from: (i) the National System of Information on Birth Records (Datasus/SINASC); (ii) the National System of Mortality Records (Datasus/SIM); (iii) and the *Terrestrial Air Temperature and Precipitation: 1900-2008 Gridded Monthly Time Series, Version 1.02* (Matsuura & Willmott (2009)).

Table 2.2: Summary Statistics: Education and School Achievement

Variables	Mean	Std. Deviation	Min	Max	Number of Observations
	Panel A: 5th Graders				
Math Test Score	172,1	36,4	0	364,1	330884
Never Failed a Grade (0/1)	0,57	.	.	.	315871
Never Dropped Out (0/1)	0,88	.	.	.	317717
Male (0/1)	0,50	.	.	.	321158
Mother Education (Up to Complete Primary) (0/1)	0,61	.	.	.	219192
Household Owns TV (0/1)	0,91	.	.	.	321519
Household Owns Refrigerator (0/1)	0,78	.	.	.	320632
Daily Domestic Work > 1 hr (0/1)	0,36	.	.	.	316341
Child Work (Outside the Household) (0/1)	0,20	.	.	.	317314
Weather Indicators, per Municip. and Month of Birth:					
Rainfall in the Past 12 Months	839,2	342,3	58,3	2460,8	330884
Deviation of Log Rainfall in the Past 12 Months	0,02	0,3	-2,1	0,8	330884
Drought Indicator (0/1)	0,13	.	.	.	330884
	Panel B: 9th Graders				
Math Test Score	225,8	39,0	0	415,1	320499
Never Failed a Grade (0/1)	0,57	.	.	.	313172
Never Dropped Out (0/1)	0,87	.	.	.	313475
Male (0/1)	0,43	.	.	.	314145
Mother Education (Up to Complete Primary) (0/1)	0,66	.	.	.	267487
Household Owns TV (0/1)	0,92	.	.	.	316995
Household Owns Refrigerator (0/1)	0,82	.	.	.	315811
Daily Domestic Work > 1 hr (0/1)	0,53	.	.	.	312804
Child Work (Outside the Household) (0/1)	0,26	.	.	.	305659
Weather Indicators, per Municip. and Month of Birth:					
Rainfall in the Past 12 Months	694,4	263,5	60,5	2149,9	320499
Deviation of Log Rainfall in the Past 12 Months	-0,16	0,3	-1,9	1,0	320499
Drought Indicator (0/1)	0,22	.	.	.	320499

Notes: This table presents educational outcomes and weather indicators averaged at the student level. Authors' calculation based on microdata on: (i) achievement test scores, students characteristics and other educational information originally provided by the 2005 and 2007 Prova Brasil editions (INEP/MEC); (ii) weather data originally from the *Terrestrial Air Temperature and Precipitation: 1900-2008 Gridded Monthly Time Series, Version 1.02* (Matsuura & Willmott (2009)).

Table 2.3: Summary Statistics: Education and Students Characteristics in the Semiarid Northeast Brazil (2008 and 2009 School Censuses)

	Grade of Enrollment in 2008						
	2	3	4	5	6	7	8
Achievement and Attainment:							
Success Rate	0.70	0.71	0.73	0.72	0.62	0.67	0.71
Success Rate (Exclude Dropout)	0.77	0.78	0.81	0.83	0.72	0.79	0.85
Dropout Rate	0.09	0.09	0.10	0.13	0.14	0.15	0.16
Student Characteristics							
Sex (male)	0.56	0.55	0.54	0.52	0.53	0.49	0.46
Rural (Area of Residence)	0.47	0.47	0.46	0.45	0.44	0.42	0.41
Share Rural Enrolled Urban Schools	0.08	0.08	0.09	0.10	0.27	0.25	0.24
Enrolled in Rural Schools	0.44	0.43	0.42	0.40	0.25	0.25	0.24
Does not live in the municip where studies	0.08	0.08	0.08	0.08	0.09	0.10	0.10
Does not study in the municip where was born	0.32	0.32	0.33	0.34	0.34	0.35	0.35
Mobility Between 2008-2009:							
Changed Systems: Regular - Multi-Grade	0.15	0.14	0.12	0.16	0.01	0.01	0.01
Changed Administrative Dependence (Mun. State. Fed)	0.04	0.04	0.05	0.16	0.07	0.07	0.07
Changed School	0.18	0.18	0.18	0.50	0.16	0.15	0.15
Changed System 8-9 Years	0.11	0.12	0.10	0.11	0.08	0.08	0.08
Changed Municipality (of Enrollement)	0.05	0.05	0.04	0.05	0.04	0.04	0.04
School Characteristics:							
Enrolled in Schools with System in 8 Years	0.63	0.58	0.54	0.54	0.51	0.51	0.51
Enrolled in Schools with System in Cycles	0.15	0.16	0.16	0.18	0.17	0.17	0.18
Average Number of Students per School	216.71	226.16	236.00	249.31	398.02	384.31	364.37
Average Number of Students per Grade	41.27	42.03	41.11	42.36	112.94	88.92	74.45

Notes: This table presents educational outcomes and weather indicators averaged at the student level. Authors' calculation based on microdata: (i) originally from educational information provided by the 2008 and 2009 Yearly National Educational Census (Censo Anual da Educação Básica, INEP/MEC); (ii) and weather data originally from the *Terrestrial Air Temperature and Precipitation: 1900-2008 Gridded Monthly Time Series, Version 1.02* (Matsuura & Willmott (2009)).

Table 2.4: Fixed-Effect Panel Regressions: The Impacts of Rainfall Fluctuations on Infant Mortality Rates

	Dep. Var.: Infant Mortality*1,000 per Municipality and Month of Birth					
	(1)	(2)	(3)	(4)	(5)	(6)
			Panel A: Rainfall			
Rainfall (Past 12 Months, Log Deviation)	-4.442	-4.434	-4.061	-4.395	-6.303	-7.578
	(1.365)***	(1.288)***	(1.386)***	(1.393)***	(1.682)***	(2.107)***
Rainfall (Past 12-24 Months, Log Deviation)						-2.026
						(2.418)
Rainfall (Next 12 Months, Log Deviation)						-3.773
						(2.794)
			Panel B: Drought			
Drought (Past 12 Months < 1sd from Avg)	1.777	2.035	1.785	1.988	3.457	4.354
	(1.129)	(1.111)*	(1.118)	(1.108)*	(1.275)***	(1.491)***
Rainfall (Past 12-24 Months, Log Deviation)						-1.057
						(2.350)
Rainfall (Next 12 Months, Log Deviation)						-2.712
						(2.600)
Common Specification:						
Observations	157,403	157,403	157,403	155,688	155,688	143,319
N. of Municipalities*Months	12,576	12,576	12,576	12,444	12,444	12,444
Municipality*Month of Birth FE	Yes	Yes	Yes	Yes	Yes	Yes
Year of Birth FE	Yes	Yes	Yes	Yes	Yes	Yes
States Time Trend	Yes	Yes	Yes	Yes	Yes	Yes
Grids Time Trend	No	Yes	Yes	Yes	Yes	Yes
Past Temperatures (Log Deviation)	No	No	Yes	Yes	Yes	Yes
Exclude Municipalities Top 1% by Newborns Avg	No	No	No	Yes	Yes	Yes
Weighted (by Avg N. of Newborns)	No	No	No	No	Yes	Yes

Notes: Dependent variable is infant mortality (up to 1 year old) per 1,000 births, calculated at the municipality level, per month. Variable of interest in Panel A is the measure of rainfall log-deviation in the past 12 months calculated by using equation 2-3. In Panel B the variable of interest is the dummy indicator of drought in the past 12 months, as defined by equation 2-4. All regressions include municipality by calendar month fixed effects, year fixed effects and annual state time trends. Controls for grid time trends are calculated on a monthly basis. Controls for past temperatures include the first and second past trimesters' mean temperature deviations from the municipality norm. Columns 4-6 exclude municipalities with the average number of newborns per month above 258. Regressions in columns 5-6 are weighted by the municipality average number of newborns per month. All regressions are calculated by using robust standard errors clustered at the grid level. Significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 2.5: Fixed-Effect Panel Regressions: The Impacts of Rainfall Fluctuations on Maternal Health and Infant Mortality by Causa Mortis

			Infant Mortality*1000 per Causa Mortis:					
	Avg. Birth Weight	Gestation >37 Weeks	Intestinal Infections	Mal-Nutrition	Respiratory Infections	Perinatal Origin	Congenital Defects	Other Non Reported
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Rainfall								
Rainfall (Past 12 Months, Log Dev.)	7.058	0.015	-3.725	-1.104	-0.790	-1.775	0.037	-1.605
	(2.532)***	(0.003)***	(0.912)***	(0.423)***	(0.469)*	(0.968)*	(0.276)	(0.732)**
Panel B: Drought								
Drought (Past 12 Months < 1sd from Avg)	-4.856	-0.009	2.229	0.272	0.590	0.831	-0.084	0.654
	(2.085)**	(0.003)***	(0.631)***	(0.193)	(0.333)*	(0.779)	(0.211)	(0.541)
Common Specification:								
Observations	155,507	155,485	155,688	155,688	155,688	155,688	155,688	155,688
N. of Municipalities*Months	12,444	12,444	12,444	12,444	12,444	12,444	12,444	12,444
Municipality*Month of Birth FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year of Birth FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State and Grid Time Trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Past Temperatures (Log Dev.)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Weighted (by Avg N. of Newborns)	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Exclude Top 1% Largest Avg N. Newbrn.	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Dependent variable in Column 1 is the average birth weight of newborns calculated by municipality and month of birth. Dependent variable in Column 2 is the average share of gestations that lasted more than 36 weeks, calculated by municipality and month of birth. Dependent variables in the remaining Columns is infant mortality (up to 1 year old) per 1,000 births, per causa mortis calculated at the municipality level, per month. Variable of interest in Panel A is the measure of rainfall log-deviation in the past 12 months calculated by using equation 2-3. In Panel B the variable of interest is the dummy indicator of drought in the past 12 months, as defined by equation 2-4. All regressions include municipality by calendar month fixed effects, year fixed effects, annual state time trends and monthly grid time trends, controls for past temperatures (which include the first and second past trimesters' mean temperature deviations from the municipality norm). All regressions also exclude municipalities with the average number of newborns per month above 258; and are weighted by the municipality average number of newborns per month. All regressions are calculated by using robust standard errors clustered at the grid level. Significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 2.6: The Effects of Rainfall Fluctuations During Time in Utero and Student Achievement in Math: Regressions at the Student Level Pooling the 2005 and 2007 Prova Brasil Editions

	Dependent Variable: Math Test Scores			
	(1)	(2)	(3)	(4)
	Panel A: Rainfall, 5th Graders			
Rainfall (Past 12 Months, Log Dev.)	1.407	1.563	1.637	1.719
	(0.354)***	(0.380)***	(0.521)***	(0.511)***
Observations	330,884	330,884	203,746	168,524
	Panel B: Drought, 5th Graders			
Drought (Past 12 Months $<$ 1sd from Avg)	-0.854	-1.089	-1.305	-1.167
	(0.344)**	(0.346)***	(0.416)***	(0.434)***
Observations	330,884	330,884	203,746	168,524
	Panel C: Rainfall, 9th Graders			
Rainfall (Past 12 Months, Log Dev.)	-0.339	-0.464	-0.230	0.203
	(0.439)	(0.403)	(0.456)	(0.576)
Observations	320,499	320,499	256,653	220,119
	Panel D: Drought, 9th Graders			
Drought (Past 12 Months $<$ 1sd from Avg)	-0.308	0.011	-0.357	-0.494
	(0.271)	(0.258)	(0.282)	(0.360)
Observations	320,499	320,499	256,653	220,119
Common Specification:				
Municipality*Month of Birth FE	Yes	Yes	Yes	Yes
Year of Birth FE	Yes	Yes	Yes	Yes
States Time Trend	Yes	Yes	Yes	Yes
Grids Time Trend	No	Yes	Yes	Yes
Past Temperature Log Deviation	No	Yes	Yes	Yes
Student Socioeconomic Characteristics	No	No	Yes	Yes
Exclude Municipalities Top 1% by N. of Students	No	No	No	Yes
Weighted (by the N. of Students per School)	No	No	No	Yes

Notes: Dependent variable is the student achievement test score in math in either the 2005 or the 2007 Prova Brasil edition. Variable of interest in Panels A and C is the measure of rainfall log-deviation in the past 12 months before birth calculated by using equation 2-3. In Panel B and D the variable of interest is the dummy indicator of drought in the past 12 months before birth, as defined by equation 2-4. All regressions include municipality by calendar month of birth fixed effects, year of birth fixed effects and annual state time trends. Controls for grid time trends calculated on a monthly basis are included in Columns 2, 3 and 4. Controls for past temperatures in Columns 2, 3 and 4 include the first and second past trimesters' mean temperature deviations from the municipality norm. Controls for students characteristics in Columns 3 and 4 include sex, mother's education and dummy indicators for whether the household owns refrigerator or television. Columns 3 and 4 exclude municipalities with the number of students taking the exam in a given year above 2,980. Regressions in column 4 are weighted by the number of students taking the exam per school. All regressions are calculated by using robust standard errors clustered at the grid level. Significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 2.7: The Effects of Rainfall Fluctuations During Time in Utero on Labor Market and Other Educational Outcomes: Regressions at the Student Level Pooling the 2005 and 2007 Prova Brasil Editions

	Child Work	Daily HH Work > 1hr	Entered Pre-School	Never Repeated	Never Dropped
	(1)	(2)	(3)	(4)	(5)
	Panel A: Rainfall, 4th Graders				
Rainfall (Past 12 Months, Log Dev.)	-0.001	-0.002	0.018	0.030	0.011
	(0.006)	(0.007)	(0.007)**	(0.011)***	(0.005)**
Observations	164,174	163,528	158,345	163,381	164,320
	Panel B: Drought, 4th Graders				
Drought (Past 12 Months < 1sd from Avg)	0.007	0.005	-0.005	-0.019	-0.010
	(0.006)	(0.007)	(0.006)	(0.007)***	(0.005)**
Observations	164,174	163,528	158,345	163,381	164,320
	Panel C: Rainfall, 8th Graders				
Rainfall (Past 12 Months, Log Dev.)	-0.003	-0.003	-0.015	-0.022	-0.005
	(0.006)	(0.007)	(0.006)***	(0.009)**	(0.005)
Observations	213,761	217,610	216,777	217,855	218,058
	Panel D: Drought, 8th Graders				
Drought (Past 12 Months < 1sd from Avg)	-0.002	-0.005	0.006	0.018	0.007
	(0.004)	(0.004)	(0.004)	(0.005)***	(0.003)**
Observations	213,761	217,610	216,777	217,855	218,058
Common Specification:					
Municip*Month FE and Year of Birth FE	Yes	Yes	Yes	Yes	Yes
State and Grid Time Trends	Yes	Yes	Yes	Yes	Yes
Past Temperature Log Deviation	Yes	Yes	Yes	Yes	Yes
Student Socioeconomic Characteristics	Yes	Yes	Yes	Yes	Yes
Exclude Municip. Top 1% by N. of Students	Yes	Yes	Yes	Yes	Yes
Weighted (by the N. of Students per School)	Yes	Yes	Yes	Yes	Yes

Notes: Dependent variables calculated by using microdata from the 2005 or the 2007 Prova Brasil edition. Dep. variable in Column 1 is a dummy indicating whether the student work outside the household; in Column 2, is a dummy indicating if the student perform domestic work in the household; in Column 3, is a dummy indicating whether the student was first enrolled at daycare and pre-school, or alternatively at the first grade of the elementary school; in Column 4 is an indicator of whether the student has ever repeated a grade; in Column 5, indicates whether the student has ever dropped out. Variable of interest in Panels A and C is the measure of rainfall log-deviation in the past 12 months before birth calculated by using equation 2-3. In Panel B and D the variable of interest is the dummy indicator of drought in the past 12 months, as defined by equation 2-4. All regressions include municipality by calendar month of birth fixed effects, year of birth fixed effects and annual state time trends, controls for grid time trends calculated on a monthly basis, controls for past temperatures (which include the first and second past trimesters' mean temperature deviations from the municipality norm). All regressions exclude municipalities with the number of students taking the exam in a given year above 2,980; and are weighted by the number of students taking the exam per school. All regressions are calculated by using robust standard errors clustered at the grid level. Significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 2.8: The Impacts of Rainfall Fluctuations During Time in Utero and Educational Attainment in Urban Schools: OLS Regressions at the Student Level

	Dep. Variable: Dropout (0/1)			
	All	1st-3rd	4th	5th-7th
	Panel A: No Interactions			
Rainfall (Log Dev.)	-0.013	-0.004	0.003	-0.037
	(0.004)***	(0.004)	(0.009)	(0.011)***
Rural Student	0.013	0.006	0.007	0.013
	(0.002)***	(0.003)*	(0.005)	(0.003)***
	Panel B: Interactions			
Rainfall (Log Dev.)	-0.011	-0.003	0.003	-0.029
	(0.004)***	(0.004)	(0.010)	(0.011)***
Rural Student	0.014	0.005	0.007	0.017
	(0.002)***	(0.004)	(0.006)	(0.004)***
Rainfall*Rural	-0.013	-0.008	0.001	-0.024
	(0.005)**	(0.008)	(0.012)	(0.010)**
Common Specification:				
Observations	626,638	275,911	79,330	271,397
Municipality of Birth FE	Yes	Yes	Yes	Yes
Year and Month of Birth FE	Yes	Yes	Yes	Yes
State Time Trends	Yes	Yes	Yes	Yes

Notes: Dependent variable is a dummy indicating whether the student dropped out of school between 2008 and 2009, variable calculated from microdata provided by the 2008 and 2009 Yearly National Educational Census (Censo Anual da Educação Básica, Inep/MEC) . The variables of interest are (i) the measure of rainfall log-deviation in the past 12 months before birth, calculated by using equation 2-3; (ii) a dummy indicating whether the student dwells in a rural area; (iii) and the interaction between (i) and (ii). All regressions include municipality of birth fixed effects, month of birth fixed effects, year of birth fixed effects and annual state time trends. Regressions in all Columns, but Column 4, include fixed effects for grades. All regressions exclude municipalities with the total number of students enrolled in schools above 33,326 (top 1% larger municipalities by the number of students); all regressions are weighted by the number of students per grade and school. All regressions are calculated by using robust standard errors clustered at the grid level. Significance: *** p<0.01, ** p<0.05, * p<0.1.

Table 2.9: Rainfall Fluctuation and School Attainment in Urban Schools

	Dep. Variable: Success (0/1)			
	All	1st-3rd	4th	5th-7th
	Panel A: No Interactions			
Rainfall (Log Dev.)	0.021	-0.011	-0.005	0.047
	(0.006)***	(0.010)	(0.014)	(0.010)***
Rural Student	-0.029	-0.026	-0.030	-0.031
	(0.004)***	(0.005)***	(0.007)***	(0.005)***
	Panel B: Interactions			
Rainfall (Log Dev.)	0.022	-0.012	-0.005	0.059
	(0.006)***	(0.010)	(0.014)	(0.011)***
Rural Student	-0.029	-0.024	-0.029	-0.025
	(0.004)***	(0.005)***	(0.008)***	(0.005)***
Rainfall*Rural	-0.004	0.021	0.006	-0.037
	(0.008)	(0.011)*	(0.014)	(0.014)***
Common Specification:				
Observations	626,638	275,911	79,330	271,397
Municipality of Birth FE	Yes	Yes	Yes	Yes
Year and Month of Birth FE	Yes	Yes	Yes	Yes
State Time Trends	Yes	Yes	Yes	Yes

Notes: Dependent variable is a dummy indicating whether the student succeeded the grade in 2008, variable calculated from microdata provided by the 2008 and 2009 Yearly National Educational Census (Censo Anual da Educação Básica, Inep/MEC) . The variables of interest are (i) the measure of rainfall log-deviation in the past 12 months before birth, calculated by using equation 2-3; (ii) a dummy indicating whether the student dwells in a rural area; (iii) and the interaction between (i) and (ii). All regressions include municipality of birth fixed effects, month of birth fixed effects, year of birth fixed effects and annual state time trends. Regressions in all Columns, but Column 4, include fixed effects for grades. All regressions exclude municipalities with the total number of students enrolled in schools above 33,326 (top 1% larger municipalities by the number of students); all regressions are weighted by the number of students per grade and school. All regressions are calculated by using robust standard errors clustered at the grid level. Significance: *** p<0.01, ** p<0.05, * p<0.1.