Barcelona School of Economics

Master's Degree in Specialized Economic Analysis: Economics of Public Policy Program

"Taking a Rain-Check: The Effect of Droughts on Educational Attainment in Brazil"

Authors: Dario Maddaloni, Lucille McKnight, Francesco Serra, and Iqbal Wibisono

Advisors: Ada Ferrer-i-Carbonell and Joan Llull

June 2022

ABSTRACT IN ENGLISH (100 words):

Droughts are thought to lower the opportunity cost of investing in education. However, there is limited evidence on the impact of droughts on human capital investment in children's education, especially in developing countries. Our study contributes to the literature by providing empirical evidence to a theoretical model of the trade off between investment in education and labor during negative rainfall shocks. Using the Brazilian Schools Panel, we estimate a fixed effects specification and find that higher drought intensity leads to higher investment in schooling. Additionally, we find that longer droughts lead to a larger magnitude of the effect on educational investment. We also conduct heterogeneity analysis to understand the differential effect on rural and urban areas, as well as public and private schools.

ABSTRACT IN CATALAN/ SPANISH (100 words)

Existe una creencia generalizada sobre cómo las sequías tienden a reducir el coste de oportunidad de invertir en educación. No obstante, hay evidencias limitadas del impacto que tienen las sequías sobre la inversión de capital humano en la educación de los niños, especialmente en países en vías de desarrollo. Nuestro estudio contribuye a la literatura existente proporcionando evidencia empírica a un modelo teórico de compensación entre inversión en educación y trabajo durante choques negativos de precipitaciones. Utilizando el Brazilian Schools Panel y mediante una especificación de efectos fijos, hallamos que las sequías más duraderas conducen a una mayor inversión en educación. Adicionalmente, encontramos que las sequías más largas conducen a una mayor magnitud del efecto sobre la inversión educativa. Asimismo, realizamos un análisis de heterogeneidad para comprender los efectos diferenciales en áreas urbanas y rurales, así como en escuelas públicas y privadas.

KEYWORDS IN ENGLISH (3):

Drought, Education, Brazil

KEYWORDS IN CATALAN/ SPANISH (3):

Sequía, Educación, Brasil

Contents

1 Introduction

As the temperature of the planet increases, and natural disasters become more commonplace, understanding their role in social and economic decision-making and outcomes is of the utmost importance. Climate change is thought to be most strongly felt in agricultural communities in developing countries (Piya et al. 2019). Negative rainfall fluctuations lead to lower crop yields and lower income as a result. Additionally, where access to formal credit and insurance mechanisms are limited, households may need to use other methods to smooth consumption. This consumption smoothing decision may in turn affect human capital formation.

From 2010-2012, Brazil experienced their most extreme drought event for 50 years (Brito et al. 2018). Figure 1 maps precipitation across Brazil in the relevant period, showing which areas were most affected by the drought On the map, a value of -0.5 or less indicates a drought

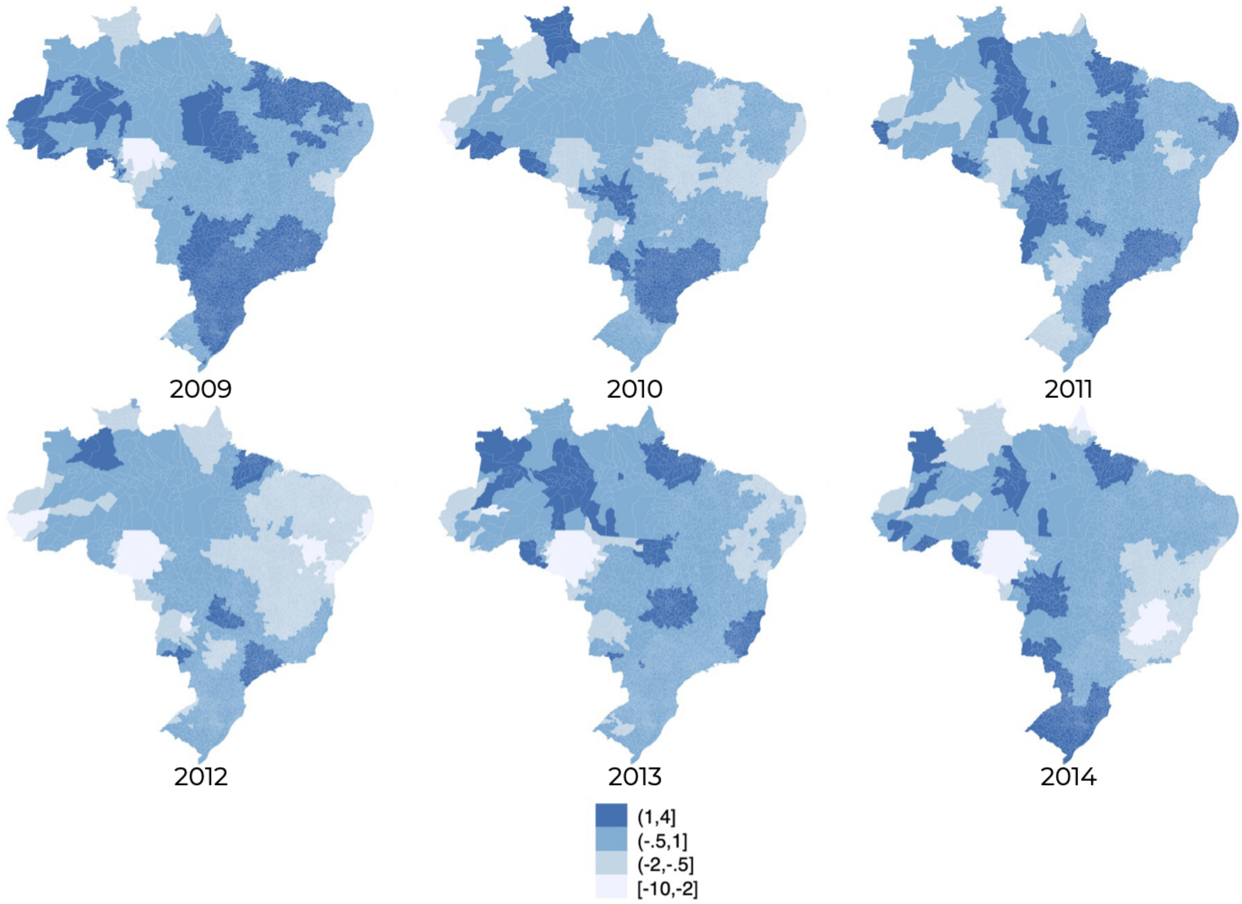

Figure 1: Variation in Precipitation (SPI) by Municipality

Brazil also experienced severe economic hardship as a result of the drought, as a large number of residents are rural farmers and their income comes from crop harvests. Additionally, national educational outcomes in Brazil are poor as compared to those of other middle income and developing countries (Soares et al. 2016). Thus, understanding the factors that contribute to decreased educational performance in the context of a changing climate is of economic and social importance.

This paper examines the impact of drought events on school outcomes in the Brazilian primary education system. Brazil is an appropriate setting for our study due to the richness of longitudinal data at the school level, as well as significant national variation in exposure to drought and rainfall shocks.

We begin by reviewing the existing body of literature on determinants of school outcomes, the effect of weather shocks on schooling and, more generally, on human capital investment dynamics in developing countries. We then propose a theoretical model measuring the effect of a change in wage (based on a weather-related income shock) on the choice of the optimal level of schooling. We hypothesize that a decrease in the wage reduces the opportunity cost of investment in education, raising the optimal level of schooling. Using the Brazilian Schools Panel, we confirm our hypothesis from a fixed effects panel regression at the school level. We conclude with a series of robustness checks to identify potential channels of heterogeneity.

While there is a limited literature on the role of rainfall shocks in rural areas, there are fewer whole country analyses, and none we know of examining all of Brazil. Our paper proceeds as follows, Section 2 summarizes relevant literature, Section 3 introduces our theoretical model, Section 4 provides summary statistics, Section 5 details our empirical strategy, and section 6 details our results along with heterogeneity analyses. Section 7 concludes.

2 Literature Review

We explore the relationship between negative rainfall shocks and educational outcomes. A rich literature investigates this relationship in several countries. Holmes (2002) observed that in North Carolina a series of storms decreased the overall performance of affected schools relative to non-affected schools by as much as 15%. Additionally, in Madagascar droughts and cyclones are found to reduce the probability that children, and specifically girls, attend school, leading them to enter the workforce (Marchetta et al. 2018). Likewise, negative rainfall shocks appear to reduce Ugandan children's school attendance by almost 10% (Agamile & Lawson 2021)). Conversely, Zimmermann (2020) observes that negative rainfall shocks *increase* enrollment rates in primary schools in India.

There are a limited number of studies set in Brazil on the impact of drought on health for young children. Chacón-Montalván et al. (2021) find that exposure to severe amounts of rain leads to lower birth weights in Amazonia. However in the Semiarid region, Rocha & Soares (2015) find that exposure to *negative* rainfall shock leads to adverse health outcomes and lower birth weights, but find these effects are lessened with adequate water and sanitation services. The only Brazil-specific education focused study to our knowledge has been conducted by Branco (2018), who focuses only on students' achievements based on Prova Brasil, a nationally standardized test of Math and Portuguese. Droughts are associated with lower scores in both subjects. We expand Branco's analysis by looking at a wider array of educational outcomes.

Further, weather shocks have a large effect on household consumption and wellbeing in agricultural communities because of the impact on livestock and crops (Alem & Colmer 2022). In the developing world, households employ a number of informal consumption smoothing tactics, such as selling off agricultural assets or skipping meals (Wik 1999). Janzen & Carter (2013) find that this behaviour lessens in Northern Kenya when households have access to programmes such as micro-insurance. In the Brazilian context, Branco & Féres (2021) find that during extreme weather shocks households alter their labor allocation away from primary jobs, such as farming, towards secondary occupations. There is limited evidence from Brazil on how child labour allocation changes, as well as using education as a potential consumption smoothing mechanism.

Overall, our study fits within a larger literature that attempts to explain under-participation and under-investment in education and human capital. There is an ambiguity on whether the effect of drought on human capital investment is positive or negative. We specifically contribute to this literature in that we create a novel interpretation of the model of human capital investment with child labor, and test the hypothesis in the context of Brazil using a robust schools data set.

3 Theoretical Model

We build our analysis moving from a simple model of household's investment in human capital. Following Shah & Steinberg (2017), our household is composed of the parents and one child, and the model envisages three periods. In the first one, the child is too young to either go to school or work. In the second period, the child has a unit of time that the parents can decide to allocate between school and work. In the third and last period, the child works, and the parents benefit from the child's accumulated human capital. The household's utility function can be written as:

$$U(c_1, c_2, c_3) = u_1(c_1) + \theta u_2(c_2) + \theta^2 u_3(c_3)$$

Where c_t represents the household consumption in period t with $t \in [1, 3]$ and θ is a time discount factor. The utility function in each period is increasing in consumption, with decreasing marginal effects:

$$\frac{\partial u_t}{\partial c_t} > 0, \frac{\partial^2 u_t}{\partial c_t^2} < 0$$

Consumption in the three periods has the following structure:

$$c_1 = w_1 h$$

$$c_2 = w_2[h + (1 - s_2)e_2] - ks_2$$

$$c_3 = w_3(h + e_3)$$

The key feature of these consumption functions is that the wage w_t is a function of rainfall in time t. In other words, we expect that as rainfall increases (decreases) the output per unit-of-work increases (decreases). Then, h represents parents' human capital (which is fixed over time), e_t is the child's human capital in time t, k is the cost of schooling, and $s_2 \in [0, 1]$ is the portion of the child's time the parents decide to allocate to education in the second period. The child's human capital is a function of consumption and schooling in the previous period and – following Cunha & Heckman (2007) – of human capital in the previous period as well. The human capital functions are:

$$e_1 = 0$$

$$e_2 = f_2(c_1)$$

$$e_3 = f_3(e_2, s_2, c_2)$$

Human capital can be understood as the productivity of the child if employed to work, and it is increasing in all three factors, with diminishing marginal returns. The productivity is therefore zero in the first period, a function of consumption in $t = 1$ in the second period, and a function of consumption, human capital, and schooling in $t = 2$ in the last period.

From this framework we can already see that parents' decision on children schooling is not trivial. Employing your child in work activities (at the expense of education) increases consumption in the second period, which has a positive effect on utility in the same period. The increased consumption has also a positive impact on human capital in the third period, which increases consumption and eventually utility in the third period. However, the effect of utility in the last period is ambiguous, because reducing schooling in the second period will also prevent human capital accumulation, and therefore impact negatively on consumption and utility in the third period. Figure 2 exemplifies these dynamics:

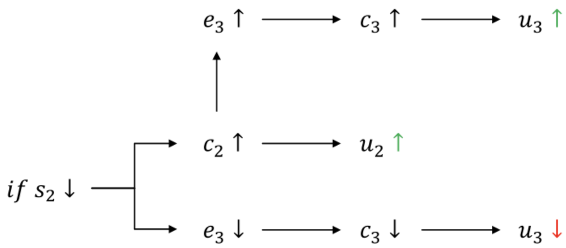

Figure 2: diagram of the effects of reducing schooling

The parents can't take any schooling decision in the first period, therefore their utility maximization function in $t = 2$ is the following:

$$\max_{s_2} \{u_2(c_2) + \theta u_3(c_3)\}$$

$$\text{s.t. } c_2 = w_2[h + (1 - s_2)e_2] - ks_2$$

$$c_3 = w_3(h + e_3)$$

The maximization constraints are expressed with an equal sign because our model does not account for the possibility of intertemporal saving or borrowing. By substituting the constraints in the maximization function we obtain the following:

$$\max_{s_2}\{u_2(w_2[h + (1 - s_2)e_2] - ks_2) + \theta u_3(w_3(h + e_3))\}$$

Deriving with respect to s_2 the result is:

$$\frac{\partial u_2}{\partial c_2} \cdot \frac{\partial c_2}{\partial s_2} + \theta \cdot \frac{\partial u_3}{\partial c_3} \cdot \frac{dc_3}{de_3} \cdot \left[\frac{\partial e_3}{\partial s_2} + \left(\frac{\partial e_3}{\partial c_2} \cdot \frac{\partial c_2}{\partial s_2}\right)\right] = 0$$

$$\frac{\partial u_2}{\partial c_2} \cdot (-w_2 e_2 - k) + \theta \cdot \frac{\partial u_3}{\partial c_3} \cdot w_3 \cdot \left[\frac{\partial e_3}{\partial s_2} + (-w_2 e_2 - k) \cdot \frac{\partial e_3}{\partial c_2}\right] = 0$$

$$\frac{\partial u_2}{\partial c_2} \cdot (w_2 e_2 + k) = \theta \cdot \frac{\partial u_3}{\partial c_3} \cdot w_3 \cdot \left[\frac{\partial e_3}{\partial s_2} - (w_2 e_2 + k) \cdot \frac{\partial e_3}{\partial c_2}\right]$$

At their optimum solution, parents equalize the marginal utility of the consumption deriving from their child not going to school now with the discounted marginal benefits of additional human capital in the third period. It is possible to observe the role of wages – proxied by rainfall variations – within this tradeoff, by observing the optimum equation from a different perspective:

$$\frac{\partial u_2}{\partial c_2} \cdot \underbrace{(w_2 e_2 + k)}_{\substack{\text{Cost in terms } c_2 \text{ of} \\ \text{choosing } s_2 = 1}} = \frac{\partial u_3}{\partial c_3} \cdot \underbrace{\theta \cdot w_3 \cdot \left[\frac{\partial e_3}{\partial s_2} - (w_2 e_2 + k)\frac{\partial e_3}{\partial c_2}\right]}_{\substack{\text{Actualized future benefit in} \\ \text{terms of } c_3 \text{ of choosing } s_2 = 1}}$$

It is easy to see that, by sending their child to school for her entire unit of time, the parents renounce to $(w_2e_2 + k)$ units of consumption in $t = 2$. The impact of a negative rainfall variation on this cost factor is straightforward: $(w_2e_2 + k)$ decreases when w_2 decreases. In other words, a drought reduces the opportunity cost of education. However, we must also consider that utility is increasing in consumption, with decreasing marginal returns. It means that for already low levels of consumption, a further reduction due to a drop in w_2 generates a strong decrease in utility. Despite the lower cost opportunity of education, households may not be able to further smooth their consumption to invest in human capital, and children would have to devote more time to work in order to increase consumption.

On the other side of the equal sign, choosing $s_2 = 1$ has two opposite effects on consumption in $t = 3$. On one hand, it increases e_3, which increases consumption in the last period. On the other hand, it decreases c_2, which has a negative impact on e_3 and eventually c_3. A negative rainfall variation has a positive impact on the actualized future benefits of schooling in terms of c_3, since w_2 is included in the equation with a negative sign. This can be intuitively explained by the fact that w_2 is the mean of transforming child's work $(1 - s_2)e_2$ in consumption (c_2): the lower w_2 the smaller the impact of the time devoted to work on c_2. This in turn means that the negative impact of schooling on e_3 due to a reduction in c_2 is relatively lower when w_2 decreases. However, this effect is mediated by three other elements.

The first one is $\frac{\partial e_3}{\partial c_2}$, i.e. the marginal effect of an additional unit of consumption in $t = 2$ on human capital in the last period. The magnitude of this term depends on the level of c_2. Intuitively, for very low levels of consumption (e.g. a starving kid) an additional unit of consumption will have a substantial impact on human capital accumulation, while for already high levels of consumption, the impact of an additional unit will be substantially lower. The second element one is w_3, which is exogenous and therefore adds a level of uncertainty to the impact of a negative shock in w_2 on overall future benefits from schooling. The third one is the discount factor θ, which depends on the value that each household assigns to future utility compared to actual utility.

It is worth noting that heterogeneity might be contained in some element of our model. First of all, the real impact of rainfall on wages will likely vary according to the geographical and socioeconomic context. Secondly, the human capital accumulation function $e_3 = f_3(e_2, s_2, c_2)$ does not specify the combination and weights of the different elements. To provide an intuition of the reason behind this choice, we can imagine that the returns of schooling on a human capital that needs to be employed in a rural context are lower than the same returns in a urban context, where the range of tasks that a working adult can perform is wider and generally more complex. The last source of heterogeneity lies in the discount factor θ, since households might vary in their valuation of future versus present utility. All these potential sources of heterogeneity might affect the choice of schooling. We attempt to account for these confounders in the heterogeneity analysis section at the end of the paper.

4 Data

4.1 Brazilian Schools Panel

The Brazilian Schools Panel dataset we use in our study merges and simplifies twenty years of data from the Brazilian School Census, educational testing, and educational indicators. The database is constructed from three different sources (Huberts & Machado (2017)). The first one is the *Censo Escolar*, which covers all schools providing fundamental education. The second one is the INEP [1] education indicators portal,

[1] *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira*, the agency connected to the Brazilian Ministry of Education in charge of evaluating educational systems and the quality of education in Brazil.

which provides some information unavailable in the census data after 2006, such as class size, pass and failure rates, dropout rates. The third source is the *Prova Brasil*, which measures the academic performance of students in the 4th and 8th grade of primary education in their final Portuguese and Math exam (these data are available only for years 2005, 2007, 2009 and 2011). The panel covers all the schools that provide fundamental education (*Ensino Fundamental I* and *Ensino Fundamental II*) included in the census between 1996 and 2015 (also including those schools that became inactive over the years), for a total of 4,066,530 year-school observations, from 412,194 different primary schools.

The dataset is structured in six sections: identifying variables, school characteristics and size, school facilities, teachers/staff, ratios/performance, and test scores. Our outcomes of interest are included in the second to last section. Although it would have been interesting to analyse the impact of droughts on pupils' performance in the Portuguese and Math tests as well, not all schools in the sample take the Prova Brasil exams. Specifically, only schools with more than 20 students enrolled at the tested grade level (4th or 8th grade) are subject to the exams, and most private schools are also excluded. Due to the class size requirement, the schools assessed by *Prova Brasil* tend to locate more in urban areas, and this subsample is therefore less representative of the entire rural school population in the panel, i.e., those schools that are more likely to be affected by a drought in their municipality.

We choose three variables to proxy schooling decision (s_2 in our model): abandonment rate [2], passing rate and rate of overage students in the following year. Passing rate is positively related to schooling, while the opposite is true for the other two variables. If a student drops out, or happens to be older than the expected age the following year, we consider it as she invested less time in education. Schooling decision, in fact, does not concern enrollment and attendance only, but more broadly the overall time devoted to educational activities.

4.2 Rainfall Data

Our rainfall data is obtained from the Climatic Research Unit (CRU) at the University of East Anglia (Harris et al. 2020). The Gridded CRU Time Series data provides information on precipitation and other climatic variables monthly at the half degree level (one data point for every 35 kilometers). Version 4.05 of this data set provides information on precipitation every month from 1901-2021. Using the coordinate point of the municipality seat in each Brazilian municipality, we use the rainfall level from the half degree the coordinate falls in, to measure the raw precipitation level for the given municipality. Figure 3 shows the locations of coordinate points across the country. Thus, we have monthly rainfall data from 1901-2021 for all 5568 mainland Brazilian municipalities. This allows us to construct our precipitation-based measure of drought.

[2] Abandonment rates and dropout rates are used interchangeably henceforth.

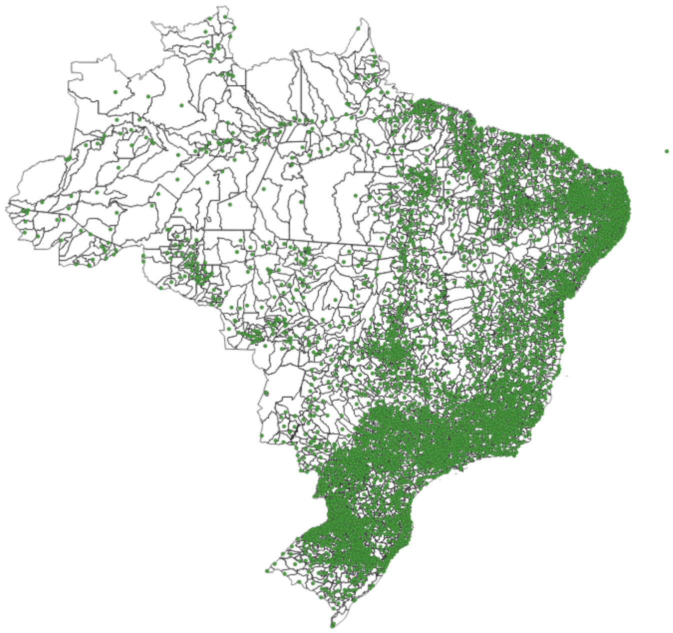

Figure 3: Reference Coordinates for Each Municipality

4.3 Municipality Data

The final dataset we draw upon are municipal level population projections disaggregated by sex and age. These data are estimates of population, extrapolated from the Brazilian National Census, which occurs every ten years. It is available every year from 2000-2021. We have approximately 88,000 observations in total. We utilise this dataset to create variables on the number of children aged 0-4, 5-9, and 10-14 within every municipality.

5 Summary Statistics

Our sample consists of 331,347 schools attached to 5559 *municipios* across Brazil. We lost approximately 80,000 schools from the raw dataset as we eliminated those schools that had only one year-observation. 283,471 (81.91%) of these schools are publicly owned, with 230,064 schools being administered by municipalities, 52,806 by the states and 601 by the Union; 62,614 schools (18.09%) are privately owned.[1] 53.42% of the schools are located in rural areas. 84,312 schools (25.44%) are situated in the *Sertao*, the semiarid region in the Northeast where droughts are more common. Table 1 below presents summary statistics of the outcomes of interest and the regressors used in all specifications. Since we have significant amounts of missing values in our outcome variables and covariates, we only use data from 2003 onward in our main specification.

Figure 4 plots the mean approval rates, dropout rates and rates of overage students in urban and rural schools from 1997 to 2014.[2] From the figure we see that the educational outcomes we are interested in, namely dropout, overage students and passing rates, show two salient characteristics in the school panel dataset.

[1]These numbers sum together yield a total amount of 346,085 schools all over the country. This indicates that 14,738 schools changed their administrative statues in the timespan of the panel survey.

[2]Data on the educational outcomes was not collected in 2000, 2001 and 2002. Thus Figure 4 is to be interpreted only before 2000 and from 2003 onwards

Table 1: Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
Outcome Variables				
Abandonment Rate	5.456	8.768	0	100
Overage Rate (t-1)	28.592	21.765	0	100
Passing Rate	80.781	17.024	0	100
Covariates				
Drought Intensity	0.045	0.937	-4.530	9.011
Long Drought	0.005	0.068	0	1
Teachers Years of Education	13.144	1.679	8	15
Students/Teacher Ratio	20.052	9.024	0.167	233
Students/Class Size	21.744	8.982	1	233
Library	0.337	0.473	0	1
Computer Lab	0.287	0.452	0	1
Science Lab	0.111	0.314	0	1
Sports Field	0.302	0.459	0	1
Television	0.634	0.482	0	1
Antenna	0.246	0.431	0	1
Computers	0.563	0.496	0	1
Internet	0.319	0.466	0	1
Water	0.933	0.250	0	1
Electricity	0.853	0.354	0	1
Sewer	0.888	0.316	0	1
Post-2007	0.532	0.499	0	1

Since the beginning of the panel, there are higher rates of dropouts and overage students in rural areas than in urban, whilst passing rates are lower. These trends support the notion that the Brazilian primary school system is underperforming in rural areas as compared to urban areas. The urban-rural differential, however, substantially diminishes in the second half of the 2000s.

A second key feature of the dropout and overage student data is that between 2006 and 2007 there is a sharp drop in overage and dropout averages, while approval rates rise. This trend holds for all schools in Brazil, both urban and rural. Dropout rates, for instance, after being consistently above 10% from 1997 to 2006, collapse to 5% in 2007, and flatten in the years after. After 2007, the variation in each outcome, as well as the urban-rural differential is significantly smaller. This downward trend is explained as a school response to the introduction of the Education Quality Index, which is examined in detail in the next paragraphs.

The 2007 change characterizes only certain school indicators such as dropout and approval rates. Other variables, such as student/teacher ratio or class size, do not exhibit such trends. As schools directly report data to the Ministry of Education to construct the school panel, the drop in negative indicators may be connected to some policy change that occurred in Brazilian education between 2006 and 2007.

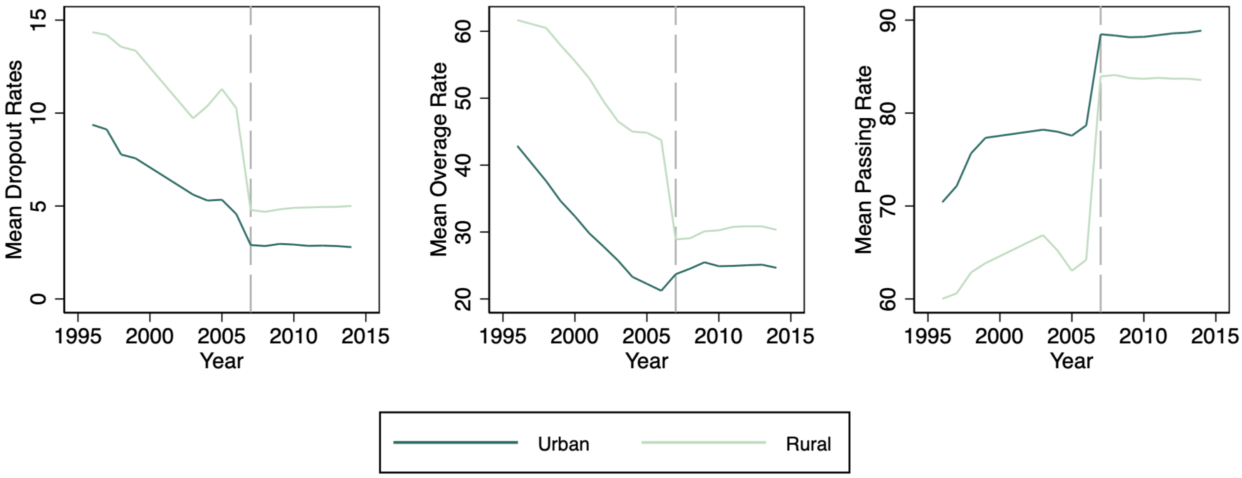

Figure 4: Trends in Mean Dropout, Overage Students and Passing Rates (1997-2014)

Our explanation for the major change in trends occurring in 2007 is the implementation of the Education Development Plan (*Plano de Desenvolvimento da Educação*, PDE) by the Lula da Silva Government. The PDE involved a series of reforms of the public education system aimed at improving the quality of education throughout Brazil (Haddad 2008) . Critically, the PDE included the ranking of all Brazilian schools based on a Education Quality Index (*Indice de Desenvolvimento da Educação Básica*, IDEB). The index was constructed as $IDEB = QF$. Q is a school average of student proficiency in *Prova Brasil*. F is the average passing rate in each school.

The IDEB, which was officially launched on April 24, 2007, allowed the federal government to identify underperforming schools. In turn, this index was used to allocate funding. Under IDEB, we expect two potential school responses. First, schools may have increased leniency towards weaker students, resulting in higher passing rates. Second, in order to obtain more funding, schools may have misreported the approval rate leading to higher reported approval rates. In turn, higher passing rates may have led to lower abandonments, as many students tend to drop out after failing a year. The possibility of a school response to the IDEB is recognized in the official Guidelines to Brazil Educational Panel Database, which states that 'the creation of the IDEB and as a consequence the change in the form of data collection seems to have had a significant effect on the way these rates are reported (Huberts & Machado 2017).

The large variation caused by the IDEB introduction could operate as a confounder on the coefficients of drought intensity on educational outcomes, thus leading to biased estimates of the treatment effect. We control for this by introducing a dummy variable for post-2007 observations, as a regressor. As we will see in the results section, this allows us to single out the variation caused by the 2007 policy change.

6 Empirical Strategy

6.1 Measuring Drought Intensity

To transform monthly rainfall data into a representative index for rainfall variations we employ a simplified version of the Standard Precipitation Index (SPI). The SPI, as defined by McKee (1995) and McKee et al. (1993)), is a simple and effective index to proxy dry and wet periods, which requires precipitation as the only input parameter. The SPI is usually calculated on a monthly basis, but we use an annual version of the index, since the frequency of observations in the Brazilian Schools Panel dataset is annual. The first step was therefore to compute average annual precipitation in each municipality, as the average of monthly rainfall in a given year. Then, the formula for computing the annual SPI is the following:

$$SPI_{m,t} = \frac{P_{m,t} - \bar{P}_m}{\sigma_m}$$

$P_{m,t}$ is average annual precipitation in municipality m and year t. $\bar{P}_m$ is the average of average annual precipitations in municipality m, computed over a base period that goes from 1901 (the first year in the rainfall dataset) to 1995 (the year before the first observations in the school panel). Finally, σ_m is the standard deviation of average annual precipitations in municipality m, computed over the base period 1901 to 1995. We calculate the SPI over the period 1996 to 2015. The choice of the base period allows us to use, for computing the index within a given municipality, the same average and standard deviation for each year considered. When average precipitation in a given year is equal to the average over the base period, the SPI takes a value of 0, while it becomes positive (negative) if positive (negative) variations in rainfall happen. To further simplify the interpretation of our results (considering that we are interested in the impact of negative rainfall variations), our drought intensity index $DI_{m,\ t}$ is equal to $-SPI_{m,\ t}$.

We also employ a measure of long term drought, using the monthly SPI measure. Monthly SPI is calculated according to the same logic as the annual index, but without needing to previously set the reference unit as the annual average of monthly precipitation. The formula is therefore:

$$SPI_{m,t,z} = \frac{P_{m,t,z} - \bar{P}_{m,z}}{\sigma_{m,z}}$$

$P_{m,t,z}$ is the value of total monthly rainfall in municipality m year t and month z. $\bar{P_{m,z}}$ is the average of rainfall in municipality m and month z, calculated over the base period. $\sigma_{m,z}$ is the standard deviation of rainfall in municipality m and month z, calculated over the base period. The long term drought measure $LD_{m,\ t}$ is a dummy variable which takes value 1 if, within a given year, there have been more than 6 months with a value of monthly SPI below -0.5. The values of $DI_{m,\ t}$ and $LD_{m,\ t}$ are assigned to all the schools from municipality m in year t, so that for each school i in municipality m, $DI_{i,\ t}$ = $DI_{m,\ t}$, and $LD_{i,\ t}$ = $LD_{m,\ t}$.

6.2 Panel Fixed-Effects Estimation

Our identification strategy relies on a fixed-effect estimation from the school panel data. The sample is composed of school observations repeated over time. We employ a fixed-effect model to control for time-invariant school unobservables that could influence our outcomes of interest. In the most basic specification of the model the educational outcomes (defined in paragraph 4.1) are kept as dependent variables ($Y_{i,t}$), and the measures of drought intensity ($DI_{m,t}$) and long drought ($LD_{m,t}$) as the independent variables alongside a vector of control variables ($X_{i,t}$). Our two main specifications are therefore:

$$Y_{i,t} = \beta_0 + \beta_1 DI_{i,t} + \beta_k X'_{i,t} + \eta_i + v_{i,t}$$

$$Y_{i,t} = \beta_0 + \beta_1 LD_{i,t} + \beta_k X'_{i,t} + \eta_i + v_{i,t}$$

In the first specification, β_1 represents the effect that one standard deviation in DI has on the educational outcome. For example, regressing on abandonment rates, a coefficient of $\beta_1 = 0.5$ indicates that one standard deviation increase in drought intensity implies a 0.5 percentage point increase in dropout rates. In the second specification, β_1 represents instead the effect of a long drought on a given educational outcome.

Relating this to our theoretical model, β_1 our primary coefficient of interest measures the impact of a negative wage (w_2) variation (as proxied by DI and LD) on household schooling decision (s_2), measured by our outcomes of interest.

The vector of covariates $X_{i,t}$ includes all variables available in the dataset that could potentially impact educational outcomes and vary over time. Many of these are related to school facilities such as a water or sewage connection or having a laboratory or library in the schools, while others are educational features such as average education level of the teachers, average class size, and student/teacher ratio. A dummy variable for the observations from 2007 onwards is included to account for the structural changes related to the introduction of the IDEB index. Standard errors are clustered at the village level to account for the possibility that schools in the same municipality experience common shocks at the municipal level.

7 Results

7.1 Main Specification Results

Table 2 below depicts the results from the fixed effects specification that employs the drought intensity index DI as the main explanatory variable. As shown in the first row, drought intensity has a statistically significant effect on all educational outcomes. A one standard deviation increase in DI leads to a 0.06 percentage points (pp) decrease in the abandonment rate, a 0.46 pp decrease in the rate of overage students in the following year, and an increase of 0.1 pp in passing rate. Crucially, the effect of drought intensity is negative on both indicators of a decrease in schooling investment (abandonment and overage rate), whereas it is positive on passing rates. These results suggest that an increase in DI has a positive effect on schooling.

We also include a lagged drought intensity coefficient to account for the fact that one drought may occur over two years and that a drought in one year can also have an impact on educational attainment in the following year. The lag of drought is significant in both the abandonment rate and overage rate models. Additionally, including the lag increases the significance and magnitude of the current period drought coefficient in the abandonment rate and passing rate models. Furthermore, we confirm our suspicions about the impact of the 2007 IDEB policy, as the coefficient capturing the effect of an observation taking place during or after 2007 is negative and significant across all three models.

Returning to our theoretical model, we can interpret these findings as follows: a more intense drought (or negative rainfall shock) results in a lower current period wage (w_2). This has two potential effects. First, the opportunity cost of schooling given by the function ($w_2e_2 + k$) is reduced. Second, the overall future benefits in terms of c_3 of schooling increase, as explained in the theoretical model section. These two effects lead to a *higher* optimal choice of s_2 during negative rainfall shocks.

Table 2: Main Specification with Drought Intensity

	Abandonment Rate		Overage Rate (in $t+1$)		Passing Rate	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Drought Intensity	-0.061*	-0.081**	-0.455***	-0.417***	0.100*	0.136**
	(0.033)	(0.039)	(0.078)	(0.089)	(0.058)	(0.068)
Lag of Drought Intensity		-0.094***		-0.340***		0.056
		(0.027)		(0.080)		(0.044)
Teachers Years of Education	-0.175***	-0.179***	-1.030***	-1.092***	0.341***	0.376***
	(0.018)	(0.019)	(0.050)	(0.053)	(0.029)	(0.033)
Students/Teacher Ratio	0.043***	0.038***	0.121***	0.138***	-0.209***	-0.187***
	(0.007)	(0.007)	(0.020)	(0.017)	(0.015)	(0.013)
Students/Class Size	-0.013	-0.005	0.028	0.030	0.159***	0.131***
	(0.008)	(0.008)	(0.025)	(0.022)	(0.017)	(0.016)
Library	-0.089**	-0.041	0.050	0.183	-0.008	-0.117
	(0.043)	(0.044)	(0.127)	(0.114)	(0.086)	(0.090)
Computer Lab	-0.532***	-0.529***	-1.978***	-1.775***	0.471***	0.445***
	(0.073)	(0.073)	(0.146)	(0.131)	(0.150)	(0.159)
Science Lab	-0.233***	-0.221***	-0.911***	-0.891***	0.080	0.113
	(0.070)	(0.070)	(0.131)	(0.118)	(0.204)	(0.213)
Sports Field	-0.209***	-0.197***	-0.963***	-0.905***	0.385**	0.400**
	(0.048)	(0.053)	(0.149)	(0.147)	(0.154)	(0.162)
Television	-0.263***	-0.287***	-0.589***	-0.725***	0.811***	1.045***
	(0.090)	(0.100)	(0.178)	(0.196)	(0.139)	(0.156)
Antenna	0.275***	0.342***	2.170***	2.235***	0.334***	0.344***
	(0.043)	(0.045)	(0.110)	(0.111)	(0.079)	(0.082)
Computers	-0.222***	-0.316***	0.961***	1.100***	0.911***	1.081***
	(0.070)	(0.085)	(0.171)	(0.200)	(0.107)	(0.131)
Internet	0.380***	0.324***	0.139	-0.123	-2.312***	-2.474***
	(0.071)	(0.074)	(0.221)	(0.219)	(0.116)	(0.130)
Water	0.582***	0.574***	0.877*	0.443	-0.482	-0.425
	(0.193)	(0.217)	(0.455)	(0.505)	(0.326)	(0.379)
Electricity	-1.355***	-1.359***	-5.840***	-5.565***	3.010***	2.911***
	(0.150)	(0.159)	(0.423)	(0.427)	(0.263)	(0.274)
Sewage	-0.048	-0.015	0.445	0.463	-0.713**	-0.876***
	(0.170)	(0.187)	(0.346)	(0.362)	(0.285)	(0.307)
2007 Policy	-4.548***	-4.528***	-8.515***	-8.771***	14.772***	14.802***
	(0.163)	(0.165)	(0.625)	(0.618)	(0.351)	(0.351)
Constant	10.620***	10.488***	47.003***	47.120***	66.117***	65.878***
	(0.326)	(0.359)	(0.877)	(0.944)	(0.520)	(0.585)
Observations	856,615	693,421	1,035,855	852,324	1,001,901	787,175
R-squared	0.088	0.089	0.105	0.113	0.214	0.225
Number of Schools	228,211	209,551	222,945	207,416	234,487	217,113

Notes: Robust standard errors clustered at the municipality level.
Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

We replicate the previous specification using a long drought explanatory variable to assess the impact of prolonged dry periods on schooling decisions. Table 3 reports the main outcomes of this specification. The coefficients confirm the positive effects of negative rainfall variations on schooling described above. These effects appear to be intensified, which is consistent with the fact that a prolonged drought determines a more substantial decrease in the opportunity cost of education.

Table 3: Main Specification with Long Drought

	Abandonment Rate		Overage Rate (in $t+1$)		Passing Rate	
Variables	(1)	(2)	(3)	(4)	(5)	(6)
Long drought	-0.646**	-0.824*	-1.494*	-3.138***	0.138	0.531
	(0.314)	(0.437)	(0.794)	(1.105)	(0.592)	(0.540)
Lag of Long Drought		-1.190**		-2.421**		1.194
		(0.564)		(1.118)		(0.774)
Observations	857,346	694,075	753,626	632,315	1,002,675	787,855
R-squared	0.088	0.089	0.055	0.052	0.214	0.224
Number of Schools	228,469	209,794	203,540	184,473	234,746	217,359

Notes: Robust standard errors clustered at the municipality level. Long drought is a dummy variable indicating a drought that occure for more than 6 months in a year. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

7.2 Heterogeneity Analysis

Table 4 and 5 explore the heterogeneous effects on our main model specification. We run the main specification dividing the sample by urban and rural, by private and public schools, and post-2007. For simplicity, we only show results for current period and lagged drought measures. However, all covariates from our main specification are included in the model.

From Table 4, we observe that, similar to the main result, drought intensity has an overall positive effect on schooling in urban areas (negative effect on abandonment and overage rate, positive on passing rate), with varying levels of significance. Surprisingly, however, DI appears to have the opposite effect in rural areas (although the results for abandonment and passing rate have little significance). We can try to interpret this result from different perspectives, also using the heterogeneity built into our theoretical model. First, the importance of schooling within the human capital accumulation function varies between urban and rural contexts. In other words, schooling may not be the best technology to transform time in human capital for an individual who will likely be employed in a farm. Furthermore, the human capital function e_3 is increasing in consumption in the previous period (c_2), with decreasing marginal returns. If poor households in affected rural areas have low baseline levels of consumption, an additional unit of c_2 may be a better investment in future human capital than an additional unit of schooling. Additionally, it is possible that for rural households the time discounting factor is smaller, meaning that future consumption is relatively less important than today's consumption.

The effects of drought intensity on schooling in public school is positive (and significant for abandonment and passing rate). This matches the hypothesis in our model. Public schools likely have a higher proportion of students from poor households, who decide to invest in their children when the opportunity cost of education decreases, in order to benefit from the future payoff of an increase in human capital. Meanwhile, for private schools we find positive and significant effects on drought intensity on abandonment rate, and negative and significant effects on passing rates, thus a negative effect on schooling overall. There are two ways this outcome can be explained. First, it is possible that the function w_2 for wealthy families is less

Table 4: Heterogeneity Analysis with Drought Intensity

Sample	Variables	Abandonment Rate	Overage Rate(in t+1)	Passing Rate
		(1)	(2)	(3)
Urban	Drought Intensity	-0.120***	-0.219**	0.080
		(0.038)	(0.100)	(0.052)
	Lag of Drought Intensity	-0.115***	-0.175**	-0.018
		(0.022)	(0.086)	(0.048)
	Observations	376,417	391,095	422,225
	R-squared	0.074	0.048	0.137
	Number of Schools	102,798	90,004	107,143
Rural	Drought Intensity	0.104	0.260**	-0.101
		(0.064)	(0.120)	(0.093)
	Lag of Drought Intensity	0.028	-0.283**	-0.021
		(0.057)	(0.125)	(0.086)
	Observations	317,004	240,585	364,950
	R-squared	0.114	0.098	0.309
	Number of Schools	110,361	97,352	113,739
Private	Drought Intensity	0.118***	-0.025	-0.359***
		(0.039)	(0.115)	(0.089)
	Lag of Drought Intensity	0.056*	0.329***	-0.212***
		(0.029)	(0.127)	(0.070)
	Observations	115,322	120,785	129,319
	R-squared	0.049	0.189	0.039
	Number of Schools	34,228	28,530	36,427
Public	Drought Intensity	-0.115***	-0.151	0.218***
		(0.044)	(0.095)	(0.061)
	Lag of Drought Intensity	-0.113***	-0.362***	0.084
		(0.033)	(0.075)	(0.051)
	Observations	578,099	510,895	657,856
	R-squared	0.120	0.096	0.299
	Number of Schools	175,618	155,994	181,020
Post-2007	Drought Intensity	0.011	0.004	-0.015
		(0.019)	(0.046)	(0.027)
	Lag of Drought Intensity	0.013	0.006	0.039
		(0.020)	(0.051)	(0.029)
	Observations	295,509	202,831	389,263
	R-squared	0.000	0.000	0.000
	Number of Schools	149,967	100,906	167,432

Notes: Robust standard errors clustered at the municipality level. Same specificafion from the main model is used. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

influenced by rainfall shocks. However, this does not explain the significance of the relationship. For the second explanation, consider the coefficient from the opposite perspective to interpret the results: a positive rainfall shock increases the level of schooling. A higher wage generates an increased level of consumption, c_2. However, for wealthier households the level of consumption is sufficiently high that a marginal increase has very little effect. Therefore, investing in an additional unit of s_2 will have a comparatively larger impact on e_3 in the next period. When the wage is high wealthy families still invest in education. On the other hand, we find no significant effects of drought intensity using post-2007 data.

In Table 5, we replicate the previous specification using long drought as the main explanatory variable. As in the main specification, this new setting generally leaves the sign of the coefficients unchanged, while increasing their magnitude. This is not true, however, for rural areas, where the coefficients change sign when switching to a long drought measure. These new results appear to be more in line with the predictions based on our theoretical model. This might suggest that a long drought index may be more appropriate to evaluate the effect on education in rural areas, as the impact of longer droughts on crops may be more intense. On the other hand, we only find that in the post-2007 estimation, only abandonment rate has significant coefficient.

Besides the change of sign in the coefficients for rural areas, the most interesting finding of the table is the substantial increase in the magnitude of the overage rate coefficient. Long droughts (both in years t and $t-1$) appear to considerably reduce the rate of overage students in year $t+1$, in urban and rural public schools. These results suggest that prolonged droughts increases the educational investment in students who would not necessarily have dropped out from school otherwise, but who would not have progressed to the next grade due to low investment in schooling.

In order to account for potential measurement error in dropout rates introduced by the 2007 IDEB reform, we exploit the measure of class size (which is not affected by the policy) to generate a new measure of abandonment rate at the municipality level. We first compute the number of students in each primary school in a given municipality using the ”class size” and ”level size” variables. Then, using Brazilian National Census Data, we compare this number to the total number of school aged children in the municipality, generating the level of non-attendance at the municipality level. We find that one additional standard deviation of drought intensity leads to 13.1 pp decrease in non-attendance at the municipality level, significant at the 1% significance level. To further account for this policy, we split our sample into just observations that occurred after 2007. However, we find no significance in this specification.

Additionally, we add to our heterogeneity analysis in this section by including Log GDP as a proxy for municipality wealth. From our model, we assume that the schooling decision s_2 in wealthier municipalities is more stable, as the wage is not as impacted by fluctuations in rainfall. Additionally, the inputs in the human capital accumulation function e_3 may be weighted differently. However, log GDP is not significant in any specification.

Long drought is not significant in either model, but it matches the sign of drought intensity and has a relatively similar magnitude. This matches the results in our main specification. However, the lag of long drought is significant and and negative, indicating that a long drought in the previous year leads to a large reduction in non-attendance. This matches our main specification in that long droughts may have a greater effect on crop success and therefore the household’s schooling decision.

Table 5: Heterogeneity Analysis with Long Drought

		Abandonment Rate	Overage Rate(in t + 1)	Passing Rate
Sample	Variables	(1)	(2)	(3)
Urban	Long drought	-1.426**	-5.115***	1.230*
		(0.644)	(1.631)	(0.720)
	Lag of Long Drought	-1.566***	-2.943**	2.012***
		(0.459)	(1.224)	(0.546)
	Observations	376,757	391,494	422,577
	R-squared	0.074	0.048	0.137
	Number of Schools	102,908	90,104	107,255
Rural	Long drought	-0.479	-2.040*	1.047
		(0.617)	(1.133)	(0.827)
	Lag of Long Drought	-1.329	-3.429**	1.453
		(0.947)	(1.455)	(1.208)
	Observations	317,318	240,821	365,278
	R-squared	0.114	0.098	0.309
	Number of Schools	110,502	97,477	113,881
Private	Long drought	0.520	0.510	-1.809**
		(0.665)	(1.814)	(0.784)
	Lag of Long Drought	-0.134	1.785	-0.174
		(0.440)	(1.668)	(1.087)
	Observations	115,393	120,865	129,392
	R-squared	0.049	0.189	0.038
	Number of Schools	34,247	28,548	36,447
Public	Long drought	-1.097**	-4.524***	1.184**
		(0.467)	(1.037)	(0.505)
	Lag of Long Drought	-1.462**	-4.538***	1.578**
		(0.662)	(1.056)	(0.788)
	Observations	578,682	511,450	658,463
	R-squared	0.119	0.096	0.299
	Number of Schools	175,842	156,193	181,246
Post-2007	Long drought	-0.249*	0.274	0.019
		(0.139)	(0.341)	(0.240)
	Lag of Long Drought	-0.001	-0.008	-0.000
		(0.010)	(0.020)	(0.013)
	Observations	184,271	136,831	196,415
	R-squared	0.000	0.000	0.000
	Number of Schools	184,271	136,831	196,415

Notes: Robust standard errors clustered at the municipality level. Same specificafion from the main model is used. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6: Heterogeneity Analysis using Municipal Level Data

	Non-attendance			
Variables	(1)	(2)	(3)	(4)
Drought Intensity	-0.121***	-0.125***		
	(0.00954)	(0.00968)		
Lag of Drought Intensity		-0.104***		
		(0.00951)		
Long drought			-0.148	-0.154
			(0.150)	(0.151)
Lag of Long Drought				-0.343*
				(0.178)
Observations	22.929	22.489	22.929	22.929
R-squared	0.228	0.234	0.219	0.219
Number of Schools	4.227	4.213	4.227	4.227

Notes: Robust standard errors clustered at the municipality level. Dependent variable is non-attendance kids in log form. We control for municipal GDP share of school with internet, water, and sewage, student teacher ratio, and and dummy post-2007. Significance: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

8 Conclusion

We estimate the effects of a series of negative rainfall shocks on three educational outcomes: school abandonment, passing rates and rate of overage students. Our analysis provides empirical evidence to a model of investment in human capital, namely the household's decision on the tradeoff between education and labor.

From the existing literature, there is no consensus on how households respond to a change in precipitation, which in rural areas acts as a proxy for household wage. Our main specification provides a preliminary idea of how these dynamics operate. Using a fixed effects specification at the school level, we find that negative variation in rainfall and the presence of a long-drought event decrease dropouts and rate of overage students, while increasing passing rates. This suggests that households facing a drought opt for increasing investments in their offspring's education, leading to fewer dropouts and more students passing the year. This supports the hypothesis presented in our theoretical model. Furthermore, we find that longer droughts lead to a larger magnitude of educational investment. We account for potential heterogeneity by dividing our school sample in categories: urban and rural and private and public. Additionally, we account for school level measurement error by estimating dropout rates at the municipal level.

Our results are relevant to policymakers in Brazil and other developing countries facing increasingly dry climates in the context of intensifying global warming. Overall, it is important to understand the role that droughts play in human capital investment, especially for households whose incomes are highly dependent on positive levels of rainfall. We provide an overview of potential responses to rainfall variations that can be employed in development policy.

References

Agamile, P. & Lawson, D. (2021), 'Rainfall shocks and children's school attendance: evidence from uganda', *Oxford Development Studies* **49**(3), 291–309.

Alem, Y. & Colmer, J. (2022), 'Blame it on the rain: Rainfall variability, consumption smoothing, and subjective well-being in rural ethiopia', *American Journal of Agricultural Economics* **104**(3), 905–920.

Branco, D. & Féres, J. (2021), 'Weather shocks and labor allocation: Evidence from rural brazil', *American Journal of Agricultural Economics* **103**(4), 1359–1377.

Branco, D. K. S. (2018), 'Drought shocks and its economic impacts on northeastern brazil'.

Brito, S. S. B., Cunha, A. P. M., Cunningham, C., Alvalá, R. C., Marengo, J. A. & Carvalho, M. A. (2018), 'Frequency, duration and severity of drought in the semiarid northeast brazil region', *International Journal of Climatology* **38**(2), 517–529.

Chacón-Montalván, E. A., Taylor, B. M., Cunha, M. G., Davies, G., Orellana, J. D. & Parry, L. (2021), 'Rainfall variability and adverse birth outcomes in amazonia', *Nature Sustainability* **4**(7), 583–594.

Cunha, F. & Heckman, J. (2007), 'The technology of skill formation', *American economic review* **97**(2), 31–47.

Haddad, F. (2008), *O Plano de Desenvolvimento da Educação: razões, princípios e programas*, MEC–Ministério da Educação, INEP–Instituto Nacional de Estudos e Pesquisas ....

Harris, I., Osborn, T. J., Jones, P. & Lister, D. (2020), 'Version 4 of the cru ts monthly high-resolution gridded multivariate climate dataset', *Scientific data* **7**(1), 1–18.

Holmes, G. M. (2002), 'Effect of extreme weather events on student test performance', *Natural Hazards Review* **3**(3), 82–91.

Huberts, A. & Machado, F. (2017), 'Brazil education panel database'.

Janzen, S. A. & Carter, M. R. (2013), 'The impact of microinsurance on consumption smoothing and asset protection: Evidence from a drought in kenya'.

Marchetta, F., Sahn, D. E. & Tiberti, L. (2018), 'School or work?: The role of weather shocks in madagascar'.

McKee, T. B. (1995), Drought monitoring with multiple time scales, *in* 'Proceedings of 9th Conference on Applied Climatology, Boston, 1995'.

McKee, T. B., Doesken, N. J., Kleist, J. et al. (1993), The relationship of drought frequency and duration to time scales, *in* 'Proceedings of the 8th Conference on Applied Climatology', Vol. 17, California, pp. 179–183.

Piya, L., Maharjan, K. L. & Joshi, N. P. (2019), 'Climate change and rural livelihoods in developing countries', *Socio-Economic Issues of Climate Change* pp. 11–33.

Rocha, R. & Soares, R. R. (2015), 'Water scarcity and birth outcomes in the brazilian semiarid', *Journal of Development Economics* **112**, 72–91.

Shah, M. & Steinberg, B. M. (2017), 'Drought of opportunities: Contemporaneous and long-term impacts of rainfall shocks on human capital', *Journal of Political Economy* **125**(2), 527–561.

Soares, J. F., Alves, M. T. G. & Xavier, F. P. (2016), 'Effects of brazilian schools on student learning', *Assessment in Education: Principles, Policy & Practice* **23**(1), 75–97.

Wik, M. (1999), 'Coping with risk in agriculture: Income-and consumption-smoothing strategies in ldcs'.

Zimmermann, L. (2020), ‘Remember when it rained–schooling responses to shocks in india’, *World Development* **126**, 104705.