Do Conditional Cash Transfers Increase School Enrollment? Evidence from Brazil

Jeanette L. Silvernale
University of San Francisco, jlsilvernale@usfca.edu

Follow this and additional works at: https://repository.usfca.edu/thes

Part of the Econometrics Commons, Growth and Development Commons, and the International Economics Commons

Recommended Citation
https://repository.usfca.edu/thes/1397

This Thesis is brought to you for free and open access by the Theses, Dissertations, Capstones and Projects at USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. It has been accepted for inclusion in Master's Theses by an authorized administrator of USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. For more information, please contact repository@usfca.edu.
Do Conditional Cash Transfers Increase School Enrollment? Evidence from Brazil

Nettie Silvernale
Department of Economics
University of San Francisco
2130 Fulton St.
San Francisco, CA 94117

Thesis Submission for the Master of Science Degree in International and Development Economics
e-mail: jlsilvernale@usfca.edu
6/7/2021

Abstract: Conditional cash transfer programs have been widely deployed across the globe. They seek to bolster human capital by providing benefits contingent upon a variety of actions such as school attendance and regular health checkups. Bolsa Familia is the most extensive conditional cash transfer program, serving 46 million Brazilian citizens. Despite its expansive size, there have been seldom large-scale micro evaluations of the program. Limited data has resulted in small sample sizes drawn from a single point in time. Our study diverges from the rest and utilizes a novel administrative dataset with an overall sample of nearly 60 million and a period of four years. We first examine the impact that Bolsa Familia has on school enrollment by establishing a baseline relationship via a fixed effect linear probability model. Then, we utilize fuzzy regression discontinuity approach that evaluates the school outcomes of families right above and below the income threshold. Initial linear probability regressions find that Bolsa Familia consistently reduces the risk of a child not being in school by around 11% for those currently enrolled. Our fuzzy regression discontinuity approach amplifies these results, suggesting that Bolsa Familia leads to a 53% reduction in dropout. Further, we find reductions in school non-enrollment rates amongst all ages, despite much of the literature only pointing to reductions for older children.
1. Introduction

Conditional cash transfer (CCT) programs are often heralded as the anti-poverty model of choice for many nations. Since the late 1990s, these programs have witnessed an exponential uptake, with much of Latin America and the Caribbean implementing some form of CCT. The program is laid out as such: money is transferred to beneficiaries conditional upon their compliance with a set criterion of human-capital investments such as sending their children to school and attending regular health checkups. CCT’s help boost both short-term welfare by providing a means of income smoothing, along with long-term macro-level human capital accumulation by requiring investments in education and health. Research also indicates that they have had positive effects in the realm of health (Rasella, et al., 2013; Shei, 2013; Paes-Sousa, et al., 2011), education (Glewwe, et al., 2012; De Brauw, et al., 2015; Schaffland, 2011), female empowerment (De Brauw, et al., 2014), and lowering crime (Chioda, et al., 2016; Loureiro, 2012; Alves, et al., 2019). This particular study aims to look at the educational effects of Bolsa Familia, the most extensive CCT program in the world.

Bolsa Familia is currently the largest CCT program in the world, delivering monetary support to nearly 11.1 million families (over 46 million Brazilian citizens) with a budget of US$11.2 Billion (Cirkovic, 2019). It was founded in 2003 and originated as the composition of four transfer programs that each had their own emphasis (“promoting schooling, health care, compensating for fewer price subsidies or promoting food consumption”) but similar target populations (Lindert, et al., 2007, p.13). In hopes to expand Brazil’s social safety net and create a more efficient system, President Lula combined the four programs into what is now known as Bolsa Familia. At its genesis, it was aimed at impoverished families with children in the household. As of 2014, eligibility was determined by having children under the age of 17 and a monthly income of less than R$154 (see Appendix A for complete benefit timeline and Appendix B for an example of benefits received by a fictitious family). The transfer of the cash is contingent upon school attendance by each child in the household, with 85% attendance expected for children between the ages of 6 and 15 and 75% attendance for children between 16 and 17.

Furthermore, the benefit amount goes up with age, as it is believed that older children have a higher opportunity cost of being in school. Data on compliance is obtained from schools reporting to their local municipalities who then disclose the information to the Ministry of Social Development (MDS). The funds for the program are distributed via electronic benefit cards that are preferentially given to a female head-of-household. In 2007, it was estimated that “93% of legally responsible beneficiaries were women” (Lindert et al.,
2007, p. 17). Our data reinforces this trend, finding that 94.21% of recipient households are female-headed. By providing cash transfers to a household contingent upon a minimum school attendance rate, it acts as an incentive to prioritize education. However, Bolsa Familia was not rolled out in a randomized fashion like Oportunidades (Mexico) and Bono de Desarrollo Humano (Ecuador), which makes drawing conclusions on causality challenging. Further, many of the studies available have not had access to large-scale microdata, but rather, only one round of the National Household Survey (PNAD) from 2006, which included Bolsa Familia questions. Our study employs a never-before-used administrative dataset that spans four years and includes demographic information on both Bolsa Familia recipients and those who did not qualify for the program. With nearly 18.5 million observations we are able to garner precise estimates of the program’s impact.

We first estimate the underlying relationship between Bolsa Familia and school enrollment by conducting a linear probability (LP) model with a vector of demographic controls and year and state fixed effects. We find that Bolsa Familia reduces a child’s risk of dropping out by nearly -11%, helping to keep approximately 60,627 children in school who would have otherwise dropped out. Once we establish a strong negative relationship between school dropout rate and Bolsa Familia, we utilize our next econometric model, a fuzzy regression discontinuity (RDD). This quasi-experimental approach assesses the differences in school enrollment for children belonging to families that receive Bolsa Familia benefits compared to children in families who do not. By limiting our average income window to those who fall within R$18 above and below the threshold, we can compare observations that should otherwise be statistically similar, with the only difference being Bolsa Familia enrollment. Therefore, if different school outcomes are observed between those right above the cutoff and those right below, we can confidently conclude the disparity is a result of the CCT program. The RDD results indicate that being currently enrolled in Bolsa Familia reduces the risk of dropping out by -53%. The estimates are significant at the 1% level and are robust across a variety of tests. We conclude that Bolsa Familia is a highly effective program and is achieving its goal of bolstering human capital.

Our study contributes to the existing literature in that it employs a novel dataset that has never been used for academic research. To our knowledge, this is the first study of this magnitude to assess how Bolsa Familia influences school enrollment. Not only do we assess the overall impact of Bolsa Familia but we also examine the heterogeneity in results across different age groups. Most of the literature has pointed to older children realizing the benefits of CCT programs, with little impact amongst primary schoolers. Our results diverge from
the literature in that we see meaningful reductions in dropout rates amongst all ages, with Bolsa Familia leading to nearly a -55% reduction for primary schoolers, and a -35% reduction for high schoolers. Witnessing meaningful reductions across all age groups reinforces the utilization of CCT’s as means in which to build up a nation’s human capital.

The remainder of the paper is structured as follows: Section 2 gives an overview of the literature, Section 3 introduces the data and key variables, Section 4 highlights the econometric models used, Section 5 presents the empirical results, and Section 6 discusses the implications of this study and ideas for future research.

2 Literature Review

2.1 Macroeconomic ideas of education

To fully understand the theory backing CCT’s, we must revert to the foundational macroeconomic literature, which helps explain why it would be in the interest of a nation-state to invest in education. Within the field, there is a consensus that increased schooling has positive effects on sustainable economic development through various channels. Emerging in the late 1980s, endogenous growth theory posits that investment in human capital is an essential factor in development and improving the standard of living. As its name suggests, endogenous growth contends that economic growth occurs within the system rather than through exogenous forces, which was in direct contrast to previous neoclassical thought that considered external factors as a predictor for growth. For instance, neoclassical theory postulated that innovation was exogenous while, in endogenous growth theory, innovation was dependent upon the stock of knowledge present within the country, and the R&D required to bolster it (Romer, 1986).

Unlike the Solow model, endogenous growth theory posits that innovation is endogenous and dependent upon the stock of knowledge (Romer, 1986). Therefore, we can use policy intervention to increase investment in human capital and research to alter the long-run growth rates of a country. This model was undoubtedly in the minds of many policymakers and scientists responsible for the influx of CCT programs in the late 1990s and early 2000s, as it required massive investments in education and health, two primary components of human capital.

Another way in which education can lead to growth is through the spillover effects it ushers in (Nelson & Phelps, 1966). Endogenous growth theory also recognizes this channel and emphasizes how investment in human capital and R&D can affect a country’s ability to capitalize on positive externalities and spillover effects or its absorptive capacity (i.e., it must
have some baseline human capital stock to benefit from imitative behavior). More technology will be implemented and diffused into society as education increases and it is dependent upon the gap between the theoretical stock of technology and the actual level of technology being used in production (Nelson & Phelps, 1966). Not only will the individuals who receive the education benefit, but also those who do not will reap the externality of being able to imitate the techniques tested by others. Macro-level models not only highlight the individual benefits of schooling but also the societal advantages of having an educated populous.

2.2 Underinvestment in education

If education plays a role in bolstering country-level growth, and creating positive spillover effects, then why is there such underinvestment in school and a mismatch of the perceived value between households and the government? One possible explanation is that education takes on a quasi-public good form due to the externalities it produces. While education may not always be a perfect fit into the public good typology (non-rivalrous consumption, and non-excludability), general knowledge falls into this category, something that education produces (Stiglitz, 1999). Schooling directly benefits the individual receiving it via an increase in their lifetime earnings, better health, and marriage market outcomes (Oreopoulos & Salvanes, 2011). But it also can spillover to others through promoting a stable and democratic environment, increased human rights, more social cohesion, less crime, and the spread of new technology (Glaeser, et al., 2007; Grossman, 2006; Becker & Lewis, 1973; Dee, 2004). In taking the difference between these private and public benefits, it is easy to conceptualize that the cumulative positive externalities generated by schooling are vast and may exceed the private benefits, which firmly cement it into the realm of a quasi-public good. Unfortunately, a common feature among public goods is their under-provision due to the non-internalization of the positive externalities. A mismatch is created between the private and social optimum because the market does not allow the individual who received the schooling to realize the external benefits it reaped on others. This is especially true in communities where there may not be labor market opportunities that reward more schooling, making the returns to education very low, further exacerbating its under-investment.

This disparity between the private and social optima leads to a coordination failure, which is often thought of at a macro level. However, we can redefine coordination failures to be household-specific, where the price of school exceeds the utility derived from it. These coordination failures can result in one or more Nash equilibria that are not Pareto optimal, such as a family not sending their child to school and their neighbors also not doing so. The
primary problem is then how to move from this point to a Pareto-optimal Nash equilibrium in the face of coordination failure. It is, therefore, necessary that we have some sort of mechanism that corrects the private undervaluation of education.

2.3 Theory of the conditional cash transfer

The conditional cash transfer (CCT) is one such way of correcting for these market failures. Through a cash incentive, households are induced to adjust their behavior to the social optimum. Essentially, the transfer acts as a price effect on the action of schooling. Rather than the typical notion of a price effect, which assesses how market prices influence demand for a product, the cash transfer encourages an increase in the supply of an action (De Janvry & Sadoulet, 2005). The increase in the price for the action is a mechanism that helps reward the household for the externalities of education that are not captured by market forces due to its quasi-public-good structure. CCT’s should be viewed as a contract between households conditional upon the delivery of an action (school), which bridges the social and private optimum.

This opportunity cost of attending school can also help explain why we witness such disparities in school enrollment for older versus younger children. If the fallback option is outside work or production within the household, then younger children would have a very low opportunity cost because the child’s potential income stream would be minimal. Therefore, delayed enrollment would be non-optimal. Older children would have a higher opportunity cost because their outside options increase (De Janvry & Sadoulet, 2005). Further, even if their labor market alternatives are seldom, they are often kept home to supervise younger siblings who are not of school age. This helps explain why we often see an increase in transfer size as children get older.

There is no doubt that education plays a crucial role in sustainable economic development by enhancing human capital accumulation. Policymakers are well aware of the benefits that education reaps in on individuals and the general population. However, despite its positive spillovers, we see widespread underinvestment in education. The reasoning for this is that it takes on a quasi-public good nature and does not allow individuals to internalize the positive externalities it produces (Anttila-Hughes, 2020). This creates a mismatch between the private and public valuation of school and explains why we see such educational disparities around the globe. With this in mind, many policymakers have cultivated innovative solutions to bridge the gap in valuation. One such method has been CCTs. The transfer acts as a price effect on the action of attending school and raises the individual utility
garnered from education. This mechanism corrects for the misaligned social and private optima. Not only has its effects been discussed in much of the theory literature, but it has also been evaluated in rigorous empirical studies.

2.4 Applied research on CCT’s and education in Latin America

When studying Bolsa Familia, it is important to observe the educational outcomes found in neighboring nations with similar initiatives. To begin, arguably, one of the most widely examined CCT programs is Progresa (currently Prospera), which was rolled out in Mexico in 1997. Progresa is often the blueprint used for other CCT’s implemented across Latin America. Progresa is considered a CCT program that offers families below a specific income threshold grants contingent upon them keeping their children in school for 85% of the school year. Like many other CCT’s the grant size varies based on particular child characteristics, as the opportunity cost of keeping children in school differs with their age and gender. What makes this governmental program unique is its implementation. To ensure accurate impact evaluations, the program was randomly enacted in 314 of the 495 localities during the first two years, with the remaining locations receiving the program in the third year. This allowed evaluations of the program to take a quasi-experimental approach. Taking advantage of this exogeneity, Schultz (2004) conducted a study aimed at evaluating if school enrollment rates were statistically different between control villages and treatment villages. The authors found a 0.66 difference or about two-thirds of a year increase in children who were recipients of the grant than children who had not yet received funding. While the short-term impacts of Progresa seem to be effective, other studies have attempted to see if the results are upheld over a longer span of time. One study found that after five-and-a-half years of receiving benefits, there is still a positive effect on school attainment (Behrman, Parker, & Todd 2011). Mexico’s success with Progresa undoubtedly encouraged other countries to deploy similar programs in an attempt to boost their own stock of human capital.

Brazil has also witnessed successful educational outcomes for families enrolled in the short-run and long-run. A paper by Schaffland (2011) utilized data from the 2006 PNAD (the National Household Survey) and employed a propensity score methodology which matches two individuals, one that is enrolled, and the other not enrolled, but both with the same probability of being in the program, conditional on certain observable characteristics (income variables, characteristics of the child, characteristics of the household head, etc.). By using this approach, the authors determine children predicted to be enrolled in BFP had an increase in school enrollment of 4-5%. However, this result appears to become weaker as time goes on, with school enrollment rates falling by 1.18% over time. Though, the overall impact of
Bolsa Familia on schooling remains positive throughout time, reinforcing the literature that posits the long-term success of CCT’s.

Another heavily studied CCT, *Familias en Acción* (FA), was enacted in 2001 in Colombia. To improve child nutrition, school enrollment, and household consumption, eligibility was determined by falling beneath a certain income threshold and contingent upon keeping children in school for 80% of the school year and attending regular growth and development check-ups. Like many other programs, including Bolsa Familia, the grant given to families increased as children got older. To examine the causal effect on school attendance, Attanasio et al. (2005) collected data from household surveys conducted across municipalities that received FA funding and municipalities that did not. In order to perform a difference-in-difference approach, the authors selected municipalities that were extremely similar in socio-demographic, economic, population, and urbanization realms but only differed in having access to a bank (the reason for FA not being present). The results indicate that FA did not significantly change enrollment rates for children in grades 8-11 but had a significant difference in enrollment for children aged 12-17. Confirming their hypothesis that the cash transfer can help increase the utility derived from schooling, correcting for the private undervaluation of education.

Bolsa Familia has also shown similar results in the program being more effective for older children. De Brauw et al. (2015) assessed 11,000 households, with some enrolled in Bolsa Familia and others not. The authors implement a propensity-score weighted regression to assess differences in schooling for urban vs. rural and older vs. younger children. Similar to the prior study, they uncover that the largest positive impacts on schooling are seen by older children aged 15-17, perpetuating the theory that opportunity cost increases with age. Thus, the cash transfer will largely change the behavior of older children who would have dropped out of school without the transfer, while younger children would remain in school with or without it.

Many Latin American countries have experienced increased human capital as a result of their CCT’s. Even though the educational outcomes may show a greater level of significance in the short-run than the long-run and for older rather than younger children, there is certainly an overarching consensus of the positive effects that CCT’s exhibit.

3. Data and Measurement of Variables

To apply for Bolsa Familia, families must be registered within the Cadastro Unico system (CADUNICO), a central database that determines eligibility for nearly twenty-five
transfer programs in addition to Bolsa Familia. Thus, the registry contains information on those who qualified for Bolsa Familia and those who were just above the income threshold. The formulation of the registry was initially constructed to streamline targeting of beneficiaries, prevent duplication of benefits, and reduce administrative costs. Each of the 5,564 municipalities is responsible for registering families into the database and conducting interviews that garner information on household composition, income level, and other demographic features (Lindert et al., 2007). All beneficiaries are required to complete the survey every two years in order to continue to receive transfers. Our study utilizes the administrative data from CADUNICO for its analysis and takes advantage of its panel data structure. This dataset has never been used for academic research and its purpose was solely for administrative tracking. The magnitude of this dataset is unprecedented regarding public information available on Bolsa Familia, making this study one of a kind. Overall, the dataset contains nearly 1.2 billion observations that include every single Bolsa Familia payment made. However, for our analysis, we collapse the payment data to a measure of yearly enrollment and limit our period of interest to 2014-2017. We end up with a sample of approximately 60,452,388 observations. It is comprised of demographic information and household composition features that were gathered at the time of the survey. Table 1 provides an overview of demographic characteristics categorized by Bolsa Familia enrollment.

Table 1: Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable (%)</th>
<th>Total Sample</th>
<th>BFP Recipients</th>
<th>Non-BFP Recipients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average income</td>
<td>146.25</td>
<td>70.47</td>
<td>285.21</td>
</tr>
<tr>
<td>Female</td>
<td>57.84</td>
<td>60.79</td>
<td>39.21</td>
</tr>
<tr>
<td>Age</td>
<td>27.66</td>
<td>25.25</td>
<td>30.77</td>
</tr>
<tr>
<td>Literate</td>
<td>81.64%</td>
<td>81.59%</td>
<td>81.95%</td>
</tr>
<tr>
<td>Female household head</td>
<td>92.69%</td>
<td>94.21%</td>
<td>89.59%</td>
</tr>
<tr>
<td>Household size</td>
<td>3.13</td>
<td>3.38</td>
<td>2.68</td>
</tr>
<tr>
<td>Household head literate</td>
<td>89.03%</td>
<td>89.67%</td>
<td>87.70%</td>
</tr>
</tbody>
</table>

The educational outcomes captured in the dataset include a question asked of whether each member of the household “goes to school currently,” with the responses ranging from “yes,” “no, but has gone in the past” and “never went to school.” We define our measurement of not in school as children over the age of 7 and under the age of 17 who have reported “never went to school” and “no but has gone in the past.” In Brazil, children must be enrolled in school by the age of seven, however, guardians are encouraged to enroll their children
starting at six years old (UNESCO, 2020). This is why we chose to include children seven years and older in our sample. ‘Not in school’ is the primary outcome variable utilized in the empirical models. Table 2 provides an overview of school enrollment rates conditional on particular qualities.

Table 2: School enrollment

<table>
<thead>
<tr>
<th>Variable</th>
<th>Not in school</th>
<th>95% Confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Percent</td>
<td>Lower bound</td>
</tr>
<tr>
<td>Total not in school</td>
<td>3.63</td>
<td>3.62</td>
</tr>
<tr>
<td>Not in school by age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary School Age</td>
<td>6.90</td>
<td>6.88</td>
</tr>
<tr>
<td>Lower secondary school age</td>
<td>1.14</td>
<td>1.13</td>
</tr>
<tr>
<td>Upper secondary school age</td>
<td>3.37</td>
<td>3.35</td>
</tr>
<tr>
<td>Not in school by gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3.70</td>
<td>3.69</td>
</tr>
<tr>
<td>Male</td>
<td>3.55</td>
<td>3.54</td>
</tr>
<tr>
<td>Not in school by BFP participation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not enrolled</td>
<td>3.35</td>
<td>3.34</td>
</tr>
<tr>
<td>Currently enrolled</td>
<td>2.97</td>
<td>2.96</td>
</tr>
<tr>
<td>Currently and previously enrolled</td>
<td>2.50</td>
<td>2.49</td>
</tr>
<tr>
<td>Under 10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not enrolled</td>
<td>5.61</td>
<td>5.57</td>
</tr>
<tr>
<td>Currently enrolled</td>
<td>4.53</td>
<td>4.51</td>
</tr>
<tr>
<td>Currently and previously enrolled</td>
<td>3.94</td>
<td>3.92</td>
</tr>
<tr>
<td>Between 10-13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not enrolled</td>
<td>1.16</td>
<td>1.14</td>
</tr>
<tr>
<td>Currently enrolled</td>
<td>0.71</td>
<td>0.70</td>
</tr>
<tr>
<td>Currently and previously enrolled</td>
<td>0.59</td>
<td>0.59</td>
</tr>
<tr>
<td>Over 13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not enrolled</td>
<td>3.47</td>
<td>3.45</td>
</tr>
<tr>
<td>Currently enrolled</td>
<td>3.46</td>
<td>3.45</td>
</tr>
<tr>
<td>Currently and previously enrolled</td>
<td>2.86</td>
<td>2.85</td>
</tr>
</tbody>
</table>

At first glance, we do see a difference in school rates for recipients versus non-recipients. Children who do not belong to families that receive benefits are out of school at slightly higher rates (3.35%) than children belonging to beneficiary families (2.97%). Further, we investigate whether being enrolled for two consecutive terms makes a difference. By restricting our sample to include those who were enrolled in the last period of observation, we see that children belonging to this category have the lowest rate of not being in school (2.50%). This could be attributed a learning mechanism, where families that have been in the program in a prior period are more aware of the program stipulations or have successfully altered their behavior more so than households that have only been enrolled for one period.
Table 2 also reveals that the highest rates of school non-enrollment are amongst primary-aged children (6.90%). We speculate that this is because there is a lag in the time it takes families to enroll their young children into the school system. Lower secondary school age is when most children are enrolled in school, but once they reach upper secondary school, the unenrollment rates begin to climb again. Figure 1 summarizes the average number of children reported as not being in school, plotted against age. It is divided by program participation highlighting the differentials between children belonging to families who received Bolsa Familia transfers and families who did not.

**Figure 1: Not in school by age**

Initially, we do see that individuals who belong to families that are not enrolled in Bolsa Familia are out of school at the highest rates. Further, we witness that those who are currently enrolled in Bolsa Familia and were enrolled in the last unit of observation have the highest school enrollment averages amongst all ages. Prior to running any econometric models, the raw data does point to an interesting pattern.

Next, our treatment variable is enrollment in Bolsa Familia. The measure was constructed by observing monthly payment data. We created a binary indicator variable that was 1 if the individual received a transfer. If the individual never received a transfer, it was 0.
There was often the case where the transfer would stop and later resume. When this occurred, we continued to mark that individual as enrolled. In our LP and RDD model, we use a contemporaneous measure of Bolsa Familia enrollment. What this means is that the individual was reported as enrolled at the time we observe their school outcomes. Our primary model utilizes contemporaneous enrollment, as the threshold is much more pronounced and exhibited a stronger first-stage result than using the lagged version (discussed in greater detail in section 4.2). Table 3 highlights basic Bolsa Familia enrollment statistics.

<table>
<thead>
<tr>
<th>Enrolled in BFP</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrolled in BFP</td>
<td>40,852,018</td>
<td>67.58</td>
</tr>
<tr>
<td>Not enrolled in BFP</td>
<td>19,600,370</td>
<td>32.42</td>
</tr>
<tr>
<td>Currently and previously enrolled</td>
<td>34,468,188</td>
<td>57.02</td>
</tr>
</tbody>
</table>

The average income variable has been calculated by the Brazilian government based on the income composition features reported in the CADUNICO filing such as income from work, unemployment benefits, child support, donations, etc. The final calculated number is what is used to determine if an individual is eligible for the program. We utilize the average income variable in our RDD approach to help predict enrollment. The maximum per capita monthly household income threshold has changed numerous times throughout the years. In 2014, it increased from R$140 to R$154, and in 2016 it increased to R$170 (see Appendix A for complete benefit timeline and Appendix C for a graphical representation of the changing threshold). We also construct a binary measurement that captures whether the individual’s average income fell above or below the threshold in the year in which they applied for the program. This is an additional variable that we employ to predict Bolsa Familia enrollment in the first stage of our RDD.

4. Empirical Model and Hypotheses

4.1 Linear Probability Model

Before moving into the primary identification strategy of an RDD, it is important to establish the relationships between our variables of interest. Firstly, we restrict the sample to include only the conditional portion of Bolsa Familia. Being a linear probability (LP) model, the outcome variable is a binary indicator of whether a child is currently not in school and the independent variable of interest is Bolsa Familia enrollment. The LP model allows
us to estimate the probability of $Y_k$ (not being in school) occurring, given, $X_k$ (being enrolled in Bolsa Familia and our control variables). Thus, the two outcomes represent the event occurring $P_k = Y_k = 1$, or not occurring $(1 - P_k) = Y_k = 0$. The model takes the form of:

$$E(Y_k | X_k) = P_k = \alpha + \sum \beta_k X_k$$

where $P$ is the probability that the child will not be in school.

Additional control variables incorporated in the model include age, gender (1=female, 0=male), and average income in logarithmic form. These aspects are incorporated into the model because much research points to the increasing opportunity cost of schooling for older children and females (De Brauw et al., 2015, De Janvry & Sadoulet, 2015; Emerson & Souza, 2007; Duryea, et al., 2007). By including these in our model, we will be able to determine if it is an influencing factor. Finally, average income is utilized as a control because it influences both Bolsa Familia enrollment and school outcomes; thus, omitting it would create biased results.

The model is as follows

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 \text{age} + \beta_3 \text{gender} + \beta_4 \text{average income} + \alpha_s + \delta_t + \epsilon_i$$

(1)

The $Y$ variable comprises of the following measure:

- **Not in school**: children between the ages of 7 to 17, who reported “never went to school” and “no but have gone in the past”. It is coded as a “1” if they are not in school, and a “0” if they are in school.

The $X_1$ variable comprises of the following measures:

- **Currently enrolled in Bolsa Familia**: the contemporaneous measure of enrollment. Meaning, the unit was enrolled at the time we are checking the ‘not in school’ measure.

We also include a full set of state, $\alpha_s$, and year, $\delta_t$, fixed effects in all models. The reasoning for including fixed effects is to eradicate any unobserved cross-sectional heterogeneity. By taking the average of each individual over time and subtracting it from the level each period, the omitted variation across groups has been ‘fixed,’ which removes the unobserved heterogeneity between cross-sections, leaving us with within-state and year variation, allowing for better identification of causal relationships. In total, we control for a total of 26 states and four years.
A fixed-effect method was chosen over random effects because the assumptions are a bit laxer, allowing for covariance between the time-invariant portion of the error and our regressors. To confirm this choice, we employ a Hausman test to ensure we are using the most efficient estimator. We reject the null, meaning the difference in coefficients is systematic, and the covariance is not equal to zero. The use of random effects would be biased, making fixed effects the most appropriate model.

Finally, we cluster our standard errors, $\epsilon_i$, by state. This adjustment helps correct if school dropout rates are not uniformly distributed within each state (i.e.: observations are correlated to each other within each group). We believe this is a likely scenario as factors such as education policy are decided at the state level in Brazil.

4.2 Fuzzy Regression Discontinuity

A fuzzy regression discontinuity design (RDD) is employed to uncover the causal relationship between Bolsa Familia participation and educational outcomes. This methodology requires us to have exact knowledge of the conditions that lead to treatment and a large sample size that includes participants and non-participants. The clear income threshold for Bolsa Familia and our extensive dataset allows us to observe the outcomes of those gathered right above and below the income cutoff. Our sample is restricted to the conditional portion of the program and households with children and adolescents under the age of seventeen. By analyzing households that fall within a very narrow window of monthly income, we assume that they are statistically similar, with the only significant difference being one receiving treatment (enrollment in Bolsa Familia) and the other being left untreated (not enrolled in Bolsa Familia). To ensure the statistical similarity, we determine an optimal window of analysis by generating a variety of data-driven bandwidth estimators and select the one that gives us the largest range. We settle on a symmetric bandwidth of $[-18, 18]$.

An RDD design is an excellent methodology when you have a sharp cutoff between treated and untreated, and there is total compliance on each side. However, because the demographic information obtained is self-reported, there are likely households in the untreated category receiving benefits and eligible families not receiving benefits. This could be due to erroneous reporting on the civilian or governmental side. The mix around the threshold requires us to implement an identification strategy that accounts for endogeneity issues of partial compliance. An RDD allows for a cutoff point that is probabilistic rather than deterministic. There still needs to be a discontinuity at the threshold, but it does not need to
show perfect compliance. Figure 2 depicts the discontinuity between whether a family received a transfer and their income relative to the qualification criteria in the year in which they applied.

**Figure 2: Program Discontinuity**

As suspected, Figure 2 indicates a jump in the probability of receiving Bolsa Familia payments at the income threshold. Those that fall above it are much more likely to be non-recipients, and those that fall below it appear more likely to receive benefits. Thus, we implement an instrumental variable technique that isolates the treatment effect on the compliers, or those who properly qualify and receive Bolsa Familia. Enrollment is instrumented by average income and being below the threshold. Average income is a metric calculated by the Brazilian government and includes various components reported by families in their CADUNICO filing. It is an extremely complex measure to recreate, and all of our attempts to backward engineer it failed. Because of this, we believe that it would be very difficult for an applicant to know how to manipulate their answers to achieve an average income that falls right below the threshold. Furthermore, the constantly changing income criteria and extension rules would make it difficult for an applicant to anticipate future cutoff values (see Appendix A for a complete timeline of threshold values). Therefore, we argue that average income is a good instrument for enrollment. We utilize contemporaneous
income to predict enrollment, meaning the most recently reported average income, as this produced the largest discontinuity or best prediction of actually being enrolled. We did explore the option of using a lagged version of income, meaning the unit’s average income that was reported in the previous period. However, when establishing the initial discontinuity, it was much weaker than the contemporaneous measure. Figure 3 depicts the lagged versus contemporaneous discontinuities.

**Figure 3: Lagged vs. Current Enrollment Discontinuity**

Figure 3 illustrates that the contemporaneous measure is much better at predicting enrollment, justifying our use of it in the model.

We use a fuzzy RDD design in which enrollment is instrumented by a family’s calculated monthly average income relative to the threshed in the year they applied, a binary measure of being above or below the cutoff, and an interaction term. The average income is centered on zero, with a bandwidth of [-18, 18]. We also include year and state fixed effects in all models, helping to eradicate any unobserved cross-sectional heterogeneity. The model is as follows:
First Stage:
\[ T_i = \alpha_0 + \lambda_1 D_i + X_i + \zeta_i \]  
(2)

\[ T_i = \begin{cases} 1 & \text{if enrolled in BFP} \\ 0 & \text{not enrolled in BFP} \end{cases} \]

Second Stage:
\[ Y_i = \beta_0 + \beta_1 \hat{T}_i + \delta_1 X_i + \epsilon_i \]  
(3)

The \( T_i \) variable comprises of the following measure:
- \textit{Enrollment}: if a family was enrolled between 2014 through 2017. Coded as “1” if yes, and coded as “0” if no.

The \( D_i \) variable comprises of the following measures:
- \textit{Centered average income}: Average income is centered on the threshold in the year in which they filed. It is restricted to 18 above the threshold and 18 below the threshold.
- \textit{Below Threshold}: a binary variable that indicates whether the family’s average income fell below the threshold in the year in which they applied. It is coded as a “1” if they are below the threshold, and a “0” if they are above.
- \textit{Interaction}: An interaction term between the centered average income and the binary indicator of being below the threshold.

The \( Y_i \) variable comprises of the following measure:
- \textit{Not in school}: children between the ages of 7 to 17, who reported “never went to school” and “no but have gone in the past”. It is coded as a “1” if they are not in school, and a “0” if they are in school.

The \( X_i \) covariate is the year and state fixed effects.
- The year fixed effects include 2014, 2015, 2016, and 2017. We control for 26 states.

In the first stage, we define \( D_i \) as an instrument for \( T_i \). The variable, \( D_i \), indicates whether the household is eligible by using their centered monthly per capita income in the year of application, a binary indicator of being below the threshold, and an interaction term between
the two. Once we obtain our first stage estimate, we run the second stage which utilizes our estimated $\tilde{\pi}_i$ that predicts whether a household has been treated. The parameter of interest is $\beta_1$ and captures the effect of the Bolsa Familia on educational outcomes. We also utilize triangular kernel weighting. This simply applies higher weights to the observations that are closer to the income threshold, eliminating bias that stems from using data further away from the cutoff point.

The effect uncovered will be the local average treatment effect (LATE). It isolates the average impact of the treatment on the compliers or the average educational effect on the individuals who qualified and received Bolsa Familia. It is derived by dividing the jump in the regression of not being in school on the covariate to the jump in the regression of enrollment on the covariate:

$$
\tau = \frac{\lim_{x \uparrow c} E [Y_i = 1 | D_i = x] - \lim_{x \downarrow c} E [Y_i = 1 | D_i = x]}{\lim_{x \uparrow c} E [T_i = 1 | D_i = x] - \lim_{x \downarrow c} E [T_i = 1 | D_i = x]} \quad (4)
$$

Where the numerator is the difference in school outcomes ($Y_i$) for those with incomes right below and right above the threshold ($D_i$), and the denominator is the difference in the enrollment ($T_i$) for those with incomes right below and right above the threshold ($D_i$).

5. Empirical Results

5.1 Linear Probability Model

Table 4, column (1) shows the LP regression results for Bolsa Familia’s impact on school enrollment. It illustrates how contemporaneous enrollment (being currently enrolled) in Bolsa Familia influences a child’s risk of not being in school. The initial regression results show a clear negative relationship between school dropout and being enrolled in Bolsa Familia. In the model, the coefficient of our measures of enrollment was significantly different from 0 at the 1% level. They moved in the opposite direction of our school outcome variable, meaning that children in recipient families are more likely to be in school than children belonging to non-enrolled families.
Table 4: Linear probability regression with controls and state and year fixed effects

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Currently enrolled in BFP</td>
<td>-0.0035***</td>
<td>-0.0035***</td>
<td>-0.0034***</td>
<td>-0.0034***</td>
</tr>
<tr>
<td></td>
<td>(0.0005)</td>
<td>(0.0005)</td>
<td>(0.0006)</td>
<td>(0.0005)</td>
</tr>
<tr>
<td>Female</td>
<td>0.0018***</td>
<td>0.0018 ***</td>
<td>0.0014***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0004)</td>
<td>(0.0004)</td>
<td>(0.0004)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.0007</td>
<td>0.0006</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0009)</td>
<td>(0.0009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log(Average income)</td>
<td></td>
<td></td>
<td>0.0004</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.0003)</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>0.0537***</td>
<td>0.0528***</td>
<td>0.0445***</td>
<td>0.0419***</td>
</tr>
<tr>
<td></td>
<td>(0.0024)</td>
<td>(0.0024)</td>
<td>(0.0118)</td>
<td>(0.0116)</td>
</tr>
<tr>
<td>Observations</td>
<td>17,322,182</td>
<td>17,322,182</td>
<td>17,322,182</td>
<td>15,854,703</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0031</td>
<td>0.0031</td>
<td>0.0032</td>
<td>0.0032</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td>0.0508</td>
<td>0.0508</td>
<td>0.0508</td>
<td>0.0503</td>
</tr>
</tbody>
</table>

(Clustered Standard Error in Parentheses)
***p<0.01, **p<0.05, *p<0.1

While the magnitude of the coefficients seems small, they do make a sizable difference due to our sample size of nearly 17 million and the low baseline percentage of children out of school. For instance, in Table 4, we see that being currently enrolled in Bolsa Familia decreases a child’s risk of dropping out by -0.35 percentage points. The baseline number of children not in school for this sample is 3.08 percent, so a -0.35 percentage point difference translates to nearly an -11% decrease. This means that being currently enrolled in Bolsa Familia is helping to keep nearly 60,627 children in school who would have otherwise dropped out.

Once we incrementally add controls, the results hold up and remain consistently negative and extremely similar in magnitude. For instance, in column (4) on Table 4, the probability of a child dropping out of school holding all other factors constant is -0.34 percentage points lower for current enrollees. The consistency and statistical significance of our estimates indicate the validity of our initial LP results.

An important aspect in the literature of CCT’s is the impact that age has on school outcomes. To investigate this further, we run all fixed effect regressions with age categories, divided by primary age, lower secondary, and upper secondary. Tables 5 highlights these outcomes. The estimate that is the largest in magnitude and significance is for primary-aged children. For instance, a child under 10 belonging to a household currently receiving benefits is -0.86 percentage points less likely to not be in school. According to this estimate, current
enrollment in Bolsa Familia is helping to keep about 37,000 children under the age of 10 in school. For lower secondary school, Bolsa Familia decreases the likelihood of dropping out by -0.45. However, for high schoolers, we see differing results. The percentage point impact diminishes significantly to -0.08 percentage points, meaning that Bolsa Familia is only changing the school behavior of around 6,579 children over the age of 13, despite having a total of 285,345 children out of school within this age range. These results are surprising in that they indicate that Bolsa Familia encourages the largest enrollment shifts among the youngest children. Much of the literature has contended that we witness the biggest gains in school attendance for older children when it comes CCTs. Our basic LP model reveals a different story. However, it is important to note that despite including several controls and fixed effects, we still face issues of endogeneity and bias. There are likely important confounding and unobserved features within the data. To eradicate this, we move to our next identification strategy, a fuzzy regression discontinuity.

**Table 5**: Linear probability regression by age category with state and year fixed effects

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1)</th>
<th>(5)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Under 10 years</td>
<td>10-13 years old</td>
<td>Over 13 years old</td>
</tr>
<tr>
<td>Currently enrolled in BFP</td>
<td>-0.0086***</td>
<td>-0.0045***</td>
<td>-0.0008</td>
</tr>
<tr>
<td></td>
<td>(0.0010)</td>
<td>(0.0003)</td>
<td>(0.0007)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.1665***</td>
<td>0.0225***</td>
<td>0.0164***</td>
</tr>
<tr>
<td></td>
<td>(0.0086)</td>
<td>(0.0012)</td>
<td>(0.0011)</td>
</tr>
<tr>
<td>Observations</td>
<td>4,314,971</td>
<td>4,784,032</td>
<td>8,223,179</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.0455</td>
<td>0.0035</td>
<td>0.0053</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td>0.0484</td>
<td>0.0084</td>
<td>0.0347</td>
</tr>
</tbody>
</table>

(Clustered standard errors in parentheses)

***p<0.01, **p<0.05, *p<0.1

5.3 Fuzzy Regression Discontinuity Results

Table 6 shows the fuzzy RDD results where contemporaneous Bolsa Familia enrollment is instrumented. The sample observed is 7-17-year-old children, that belong to households whose monthly income falls between R$18 above or below the threshold. While our basic OLS models had around 17 million observations, this bandwidth narrows down our observations to just above 950,000. The first stage results are extremely strong, with our binary indicator of being below the threshold highly predictive of Bolsa Familia enrollment, almost a 30-percentage point increase. *Income centered on zero* is a measure where the Bolsa
Familia threshold is centered on zero, being below zero means you have an income that qualifies enrollment, and being above zero means your income is too large and you do not qualify. Therefore, our negative coefficient on income centered on zero aligns with what we would expect, individuals whose average monthly income falls below the threshold are more likely to be enrolled. Our first stage f-statistic is 14,830. Rule of thumb contends that a valid first stage should have an f-statistic above 10. Figure 2 also visually reinforces the validity of a clear program discontinuity, predicted by falling above or below the threshold. These indicators leave us confident about our first stage and subsequent results.

The second stage outcome does indicate that current Bolsa Familia enrollment significantly decreases the probability of a child not being in school. Our coefficient indicates that there is about a -1.7 percentage point decrease in the likelihood of dropping out. While the coefficient may appear small in magnitude, it translates to nearly a -53% decrease in the risk of not being enrolled in school.

**Table 6:** Regression discontinuity with state and year fixed effects

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>First Stage: Contemporaneous enrollment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Below threshold</td>
<td>0.2992*** (0.0025)</td>
</tr>
<tr>
<td>Income centered on zero</td>
<td>-0.0042*** (0.0002)</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.0032*** (0.0003)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.4253*** (0.0024)</td>
</tr>
<tr>
<td>Observations</td>
<td>953,945</td>
</tr>
<tr>
<td>F-value</td>
<td>14830.5</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.1116</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Second stage: Not in school</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated BFP enrollment</td>
</tr>
<tr>
<td>Income centered on zero</td>
</tr>
<tr>
<td>Interaction</td>
</tr>
<tr>
<td>Constant</td>
</tr>
<tr>
<td>Dependent variable mean</td>
</tr>
</tbody>
</table>

(Robust standard errors in parentheses)

***p<0.01, **p<0.05, *p<0.1
Within our narrow bandwidth of R$18 above and below the threshold, we estimate that being currently enrolled in Bolsa Familia is responsible for keeping approximately 16,218 children in school who would have otherwise dropped out. While our results only measure local average treatment, if we were to apply the percentage decrease to the total sample, it would translate to nearly 294,478 children remaining in school. However, this calculation is likely an underestimate, as our RDD sample was restricted to Bolsa Familia recipients who had the highest income. As we move further away from the threshold, it is likely the transfers change behavior even more drastically, greatly reducing the risk of dropping out.

Graphically, we do see a discontinuity at the threshold of the percentage of children not in school. Figure 4 plots out the residual estimates of those not in school by Bolsa Familia enrollment using a linear fit. For those under the threshold, there appears to be a fairly flat line which jumps at the threshold and begins on an upward slope. Figure 5 utilizes the same data, only we employ a local polynomial fit.

Figure 4: Not in school discontinuity linear fit
In our basic LP model, we found an interesting pattern regarding age. It was observed that Bolsa Familia had a large influence on bolstering school enrollment for primary-aged children. With this in mind, we divided our RDD into the same age buckets as the LP model: under 10 years, 10 to 13 years, and 13 years and older. The output can be found in Table 7. Surprisingly, our results diverge from the LP model, in that Bolsa Familia enrollment seems to have quantitative and qualitative significance amongst all ages. We see meaningful reductions for primary, middle, and high school-aged children. For the youngest children, being currently enrolled in Bolsa Familia reduces the risk of not being in school by -2.69 percentage points or nearly -55%. For the oldest children, we see a reduction of -1.27 percentage points or about -35%. For middle-schoolers, the results are a bit more complex. Children between the ages of 10-13 are out of school at the lowest rates, so any gain from Bolsa Familia makes a sizable impact. Our findings are very interesting since much of the literature only points to seeing enrollment gains amongst high schoolers. However, we witness meaningful reductions in school non-enrollment for each age category.
To further investigate the effect of age, we reran our RDD model for each individual age, as seen in Table 8.

Table 8: Second stage fuzzy regression discontinuity results for each age

<table>
<thead>
<tr>
<th>Estimated BFP Enrollment</th>
<th>Current enrollment</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate (standard error)</td>
<td></td>
</tr>
<tr>
<td>Age 7</td>
<td>-0.0340 (0.0211)</td>
<td>63,865</td>
</tr>
<tr>
<td>Age 8</td>
<td>-0.0232 (0.0154)</td>
<td>70,258</td>
</tr>
<tr>
<td>Age 9</td>
<td>-0.0175** (0.0085)</td>
<td>74,799</td>
</tr>
<tr>
<td>Age 10</td>
<td>-0.0137* (0.0074)</td>
<td>80,006</td>
</tr>
<tr>
<td>Age 11</td>
<td>-0.0066 (0.0068)</td>
<td>86,587</td>
</tr>
<tr>
<td>Age 12</td>
<td>-0.0089 (0.0060)</td>
<td>89,473</td>
</tr>
<tr>
<td>Age 13</td>
<td>-0.0164*** (0.0046)</td>
<td>92,201</td>
</tr>
<tr>
<td>Age 14</td>
<td>-0.0150*** (0.0047)</td>
<td>96,286</td>
</tr>
<tr>
<td>Age 15</td>
<td>-0.0165*** (0.0075)</td>
<td>99,051</td>
</tr>
<tr>
<td>Age 16</td>
<td>-0.0304+ (0.0105)</td>
<td>101,083</td>
</tr>
</tbody>
</table>

(Robust standard errors in parentheses)

***p<0.01, **p<0.05, *p<0.1
Table 8 illustrates the second stage of our RDD model. Ages 13, 14, 15, and 16 seem to have the most quantitative and qualitative significance. Meanwhile, Figure 6 plots the estimates found in the Table 8. Like the table, we see the largest reductions amongst primary and high school-aged children. However, it appears that the estimate is much nosier for primary-aged children, depicted by the wider confidence intervals. We see the smallest magnitude amongst middle-schoolers; however, this group is already out of school at the lowest rates. Our RDD results lead us to conclude that Bolsa Familia is effective for all ages, but especially children between the ages of 13 to 17.

**Figure 6:** Bolsa Familia’s impact on school dropout

![Bolsa Familia's impact on school dropout for those currently enrolled in BFP](image)

5.3 Robustness Checks

To ensure our results are valid, we conduct several robustness checks. We assume that as we minimize the neighborhood and get closer to the threshold our estimated program effects should stay stable. However, to ensure this is the case we need to check that there are no signs of manipulation such as a large jump in observations right below the income cutoff. If we witness a large difference in the density of observations, it would indicate that recipients are able to successfully alter their reported monthly income to
receive benefits. To test this, we construct Figure 7, which was generated by artificially creating a 1,000 different threshold and conducting a test statistic (on the x-axis) that indicates if there is a significant difference in density above and below said threshold. Once mapped out, it should take the form of a normal distribution centered on zero. If there was absolutely no significant difference in density above and below the threshold, we would expect to see our true cutoff point centered on zero. In this case, our true threshold falls slightly below zero at about -0.25, however, it still does not indicate a significant difference in density above and below the threshold. Coupling this with the difficulty that it takes to backward engineer and predict the income qualification, we believe that there are little to no signs of manipulation.

**Figure 7: McCrary Test**

Another common robustness check used in RDD analysis is estimating the model with different sized bandwidths. This checks the sensitivity of our estimates to alternative specifications. We rerun the total analysis with a bandwidth of [-24, 24] and [-12, 12]. Our original bandwidth of [-18, 18] indicated that Bolsa Familia reduces school dropout by -1.5 percentage points. The larger bandwidth estimates a -1.49 reduction, and our smaller bandwidth estimates a -1.02 percentage point reduction. These estimates are similar in magnitude and remain significant at the p<0.05 level. Table 9 illustrates these findings.
Table 9: Fuzzy regression discontinuity with differing bandwidths

<table>
<thead>
<tr>
<th>VARIABLES</th>
<th>(1) -12, 12</th>
<th>(2) -18, 18</th>
<th>(3) -24, 24</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First Stage: Contemporaneous enrollment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Below threshold</td>
<td>0.2856***</td>
<td>0.2992***</td>
<td>0.2947***</td>
</tr>
<tr>
<td>Income centered on zero</td>
<td>-0.0063***</td>
<td>-0.0042***</td>
<td>-0.0033***</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.0045***</td>
<td>0.0032***</td>
<td>0.0010***</td>
</tr>
<tr>
<td>Constant</td>
<td>0.4259***</td>
<td>0.4253***</td>
<td>0.4415***</td>
</tr>
<tr>
<td>Observations</td>
<td>650,301</td>
<td>953,945</td>
<td>1,325,409</td>
</tr>
<tr>
<td>F-value</td>
<td>7312.71</td>
<td>14830.5</td>
<td>21775.8</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.1100</td>
<td>0.1116</td>
<td>0.1203</td>
</tr>
<tr>
<td><strong>Second stage: Not in school</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimated BFP enrollment</td>
<td>-0.0102**</td>
<td>-0.0170***</td>
<td>-0.0149***</td>
</tr>
<tr>
<td>Income centered on zero</td>
<td>-0.0001</td>
<td>-0.0003**</td>
<td>-0.0002***</td>
</tr>
<tr>
<td>Interaction</td>
<td>0.0004*</td>
<td>0.0002</td>
<td>0.0002**</td>
</tr>
<tr>
<td>Constant</td>
<td>0.0358***</td>
<td>0.0350***</td>
<td>0.0380***</td>
</tr>
<tr>
<td>Dependent variable mean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Above threshold</td>
<td>0.0320</td>
<td>0.0319</td>
<td>0.0316</td>
</tr>
<tr>
<td>Below threshold</td>
<td>0.0298</td>
<td>0.0305</td>
<td>0.0295</td>
</tr>
</tbody>
</table>

Robust standard errors in parentheses

***p<0.01, **p<0.05, *p<0.1

Next, we implement a variety of fixed effects before settling on state and year. Originally, we ran all models with municipality fixed effects. The results from these estimations remained consistent once we changed to state-level, as this is the level at which educational policy is decided. Finally, we completed both our LP and RDD models with robust standard errors, and then again with clustered standard errors. Again, this examines how sensitive our estimate is to different econometric specifications. In each analysis, our standard errors and coefficients remain extremely similar, and we do not lose significance or magnitude.

6. Conclusion

These results suggest that Bolsa Familia effectively lowers school dropout rates for program beneficiaries. We first utilized a LP model to establish a relationship between Bolsa Familia enrollment and school outcomes. The model finds that being enrolled in Bolsa
Familia reduces a child’s risk of dropping out by -0.35 percentage points, or -11%. Next, we employ a fuzzy RDD model which is more rigorous quasi-experimental approach that helps eradicate bias. The fuzzy RDD amplified the probability LP results and finds that being enrolled in Bolsa Familia reduces the risk of not being in school by -1.7 percentage points or -53%. At its genesis, Bolsa Familia was created to boost human capital by incentivizing school enrollment through monetary transfers. Our study finds that the transfer is effective in helping children stay in school. We observe a narrow neighborhood of 18 above and 18 below the threshold and see a significant difference in school enrollment rates. By looking at individuals that make within R$36 of each other, we assume that they should be extremely comparable, with the only distinction being program enrollment. Therefore, the significant difference in school enrollment must be attributed to Bolsa Familia.

Prior studies have had limited access to enrollment data, specifically only one round of the National Household Survey (PNAD). Our novel administrative dataset was reconstructed into a panel format, making it conducive for a quasi-experimental approach that uncovers the causal effects of Bolsa Familia. We see meaningful reductions in school non-enrollment for all ages. Previous literature on CCT’s has pointed to the strongest results amongst high school-aged children. However, our RDD analysis found that primary-aged children belonging to recipient families are -2.69 percentage points or -55% less likely to not be in school than their non-recipient counterparts. For high-schoolers, Bolsa Familia reduces their risk of dropout by -1.27 percentage points or -35%. These results are a major contribution to the literature, in that they point to the broad effectiveness of CCT programs.

While this study examined the conditional threshold of Bolsa Familia, there is another extension of the program that remains unstudied, the unconditional portion. We determined that the conditions attached are effective in changing school attendance behavior, however, is the same true in the absence of conditions? Future areas of research that apply the same econometric approach but look at the extreme poverty line, where individuals with a monthly income below R$89 receive a fixed benefit, and those above receive a conditional benefit, would add an interesting dimension to the CCT literature. Given this novel administrative dataset and rigorous quasi-experimental approach, we confidently conclude that Bolsa Familia is extremely effective in lowering school dropout rates, and we provide a replicable model and data for future studies.
References


Appendix A: Bolsa Familia Conditional Benefit Timeline

Bolsa Familia is created
Income threshold is R$100
Maximum variable benefit is R$45
Variable benefit for children
0-14: R$15

Income threshold is R$120
Maximum variable benefit is R$54
Variable benefit for children
0-14: R$18

Income threshold is R$120
Maximum variable benefit is R$110
Variable benefit for children
0-15: R$20
Variable benefit for children
16-17: R$30

Income threshold is R$140
Maximum variable benefit is R$92
Variable benefit for children
0-15: R$22
Variable benefit for children
16-17: R$33

Variable benefit for pregnant women: R$32

Bolsa Familia is created
Income threshold is R$100
Maximum variable benefit is R$45
Variable benefit for children
0-14: R$15

Income threshold is R$120
Maximum variable benefit is R$54
Variable benefit for children
0-14: R$18

Income threshold is R$120
Maximum variable benefit is R$110
Variable benefit for children
0-15: R$20
Variable benefit for children
16-17: R$30

Income threshold is R$140
Maximum variable benefit is R$92
Variable benefit for children
0-15: R$22
Variable benefit for children
16-17: R$33

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$172
Variable benefit for children
0-15: R$32
Variable benefit for children
16-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32

Income threshold is R$140
Maximum variable benefit is R$236
Variable benefit for children
0-14: R$32
Variable benefit for children
15-17: R$38

Variable benefit for pregnant women: R$32
Appendix B: Sample Benefit Distributions

2003

Age 2  Age
R$15  R$15

Total Benefit: R$30

2007

Age 4  Age 6  Age 10
R$18  R$18  R$18

Total Benefit: R$54

2008

Age 5  Age 7  Age 11
R$20  R$20  R$20

Total Benefit: R$60

2009

Age 6  Age 8  Age 12
R$22  R$22  R$22

Total Benefit: R$66

2011

Age 8  Age 10  Age 14
R$32  R$32  R$32

Total Benefit: R$128

2013

Age 2  Age 10  Age 12  Age 16
R$32  R$32  R$32  R$38

Total Benefit: R$134

2014

Age 3  Age 11  Age 13  Age 17
R$35  R$35  R$35  R$42

Total Benefit: R$147

2018

Age 7  Age 15  Age 17  Age 21
R$41  R$41  R$48  R$0

Total Benefit: R$130
Appendix C: Changing income threshold

Bolsa Familia Monthly Income Threshold
Appendix D: Directed Acyclic Graph (DAG)

Bolsa Familia Enrollment Measurements ($X_1$)
- Enrolled in Bolsa Familia
- Current enrollment

School Outcome Measurements ($Y$)

Controls
- Age
- Gender

Controls
- Log average income