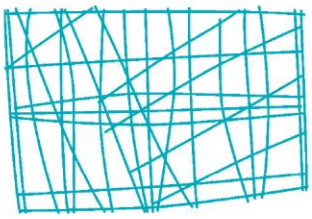


NEREUS
Núcleo de Economia Regional e Urbana
da Universidade de São Paulo
The University of São Paulo
Regional and Urban Economics Lab

LONG-TERM EFFECTS OF CONDITIONAL CASH TRANSFERS ON CHILDREN: THE BRAZILIAN CASE

Gabriel Lyrio de Oliveira
André Luis Squarize Chagas

TD Nereus 09-2020
São Paulo
2020

Long-Term Effects of Conditional Cash Transfers on Children: The Brazilian Case

Gabriel Lyrio de Oliveira* André Luis Squarize Chagas†

September 3, 2020

Abstract

In this paper, we present some long-term effects of the largest Conditional Cash Transfers program in the world, and one of the pioneers, the Bolsa Família Program (BFP). We focus on the effects on Schooling attained in early adulthood and Labor Market outcomes of individuals more or less exposed during their childhood. The estimates were enabled by linking identified data from Formal Labor Market, BFP Payment Records, and the Single Registry (SR). In this Natural Experiment, the main identification strategy relies on a rich set of control variables, and on the fact that the release of BFP resources for registered families is automatized and based on municipality poverty parameters estimated by the government. In an alternative identification strategy, we consider an instrumental variable, the observed proxy for the municipality effort to register vulnerable families. These strategies help to solve the potential selection bias of families to the SR, and consequently to the treatment. Nonetheless, since the program selects the most vulnerable families, the threats to the identification suggest that the estimates are lower bounds. Our main results show positive long-term effects on Schooling, and on the Formal Labor Market participation, while mixed results are observed for Earnings. Heterogeneity tests suggest that the effects are stronger for boys, for smaller cities, and for families with never formally employed parents.

Keywords: Conditional Cash Transfers, Long-term effects; Human Capital; Labor Market; *Bolsa Família*

JEL Codes: I25, I38, J24, O15

*Economics PhD Candidate - University of São Paulo - E-mail: **gabriel_ldo@hotmail.com**
†Department of Economics - University of São Paulo - E-mail: **achagas@usp.br**

1 Introduction

About twenty years ago, some of the Latin American developing countries implemented important social policies that later were revealed to be very powerful tools in alleviating misery and poverty. They are named, in general, as Conditional Cash Transfers programs (CCT). To receive the benefit the family must meet some requirements on income, education, and health. The requirements for education and health have the aim of improving the future of these families through the Human Capital formation of their children.

So far, the results on immediate poverty and inequality reduction are established consensus in the literature (Bastagli et al., 2016) , but the second aim of these programs began to be studied recently (Millán et al., 2019) and no evidence is known for the largest CCT of the world, the Bolsa Famíia Program (BFP).

The BFP covers about one-quarter of the Brazilian population[1]. The program began to operate in 2001 with the name *Bolsa Escola*. A few years later it was unified with other smaller programs and strongly expanded from 2004 to 2006, with its new name. Since its beginning, the program focuses on the poor and extremely poor people, and demands from the benefited families to present minimum school attendance of their children in schooling age, and health monitoring of the children and pregnant women.

The BFP short-term effects like poverty and inequality reduction (Soares et al., 2010, 2009), regional inequality reduction (Azzoni et al., 2009), increase of school attendance (Silveira-Neto and Duarte, 2010), and retention of migrants on their home municipality (De Oliveira and Chagas, 2020) were approached by the literature, nonetheless, there is a complete lack of evidence about its long-term effects. The children of some cohorts of the beginning of the program are not in schooling age anymore. Moreover, they are now in the initial years of their working age, which provides a good opportunity to explore these impacts of the program.

[1]Approximately 50 million people of the 210 million country residents are currently benefited by the program.

In this work, we focus on three outcomes: the Schooling attained by individuals in their early adulthood; their Formal Labor Market (FLM) Participation; and the Earnings they get on these employments. The relevance of Schooling for development is supported by Human Capital theory, dating back at least to Schultz (1961) and Becker (1962). Moreover, although early childhood investments in human capital have been shown as relatively more important to the individual development (Cunha and Heckman, 2007), we only observe in the data the program exposure beginning in later stages of childhood, and investments in these other stages are still relevant[2].

Considering the Labor Market outcomes, although the informal sector tends to be, as is the case for Brazil, very representative for developing countries, the lack of identified data avoids us to conduct an analysis for this sector. However, evidence for Brazil has shown that the formal sector concentrates better and higher productivity jobs (Meghir et al., 2015), increasing welfare.

To enable our research we linked three different identified datasets: the Single Registry for Social Programs (SR); the BFP Payment Records; and the Formal Labor Market administrative data (RAIS-MTE). The rich resulting data enables many future researches concerning the poor families of the country.

We rely on two alternative identification strategies. The first one takes advantage of the timing of the program, the municipality yearly quotas of benefits, and the rich identified data with many observable control variables. The database comprises individual socio-demographic characteristics, some family-background variables, including the schooling and labor market outcomes of the parents during individuals' childhood, and municipality socioeconomic indicators, also related to the individual's childhood period.

An alternative identification strategy considers the same control variables, but we introduce an instrumental variable approach[3]. We consider this strategy as a robustness check to

[2]In the coming years, it will be possible to focus on the younger cohorts of the BFP in the Labor Market and explore the effects of exposure during early childhood. Currently, the younger adults observed in the data were 6 years old at the beginning of the BFP Payment Records (2004).

[3]We applied this approach only for FLM Participation and Earnings outcomes. The Schooling outcome

the main results. We take advantage of the exogeneity, from the family perspective, of the municipality effort to include vulnerable families in the program registry. Once being registered is a condition to be benefited, this effort is strongly related to the children's exposure to the BFP.

In what follows, we present the literature review on the known effects of CCTs, the BFP design and evolution, the data and empirical methodology, and finally our results. In the last section, we summarize the paper and point to valuable future research questions related to this topic.

2 BFP design, and prior evidence

The Bolsa Família Program is a conditional cash transfer (CCT) that targets poor families with per capita income[4] currently bellow BRL 178,00, and extremely poor families, with income bellow BRL 89,00 (MC). The BFP emerged as a unification of four previous programs (*Bolsa Escola*, *Cartão Alimenteção*, *Bolsa Alimentação*, and *Auxílio Gás*) and went through a strong expansion since its creation in October 2003.

The selection process is implemented through the Single Registry (SR), a unified registry for federal social programs. For the family be listed in the SR, preferably the municipality social assistance team must visit it. Another way is the head of the family to go to the social assistance office and require registration. Then, after having the information collected, the family can be classified as eligible to the BFP or not.

To be benefited by the program the poor families must have a child, a teenager, or a pregnant woman on its composition, while for the extremely poor families only the income criteria are considered. According to the program regulation[5], beyond income and family composition criteria, the respective municipality must have an official estimate of poverty

is analyzed through an ordered probability model, the Ordered Logit, in which our instrumental variable specification faced convergence problems.

[4]The family income is computed disregarding the amounts of all official programs of cash transfers but the BPC, which is paid to the poor elderly or disabled individuals.

[5]*Portaria* 341 from October 7, Senarc - Ministry of Citizenship (2008).

incidence and the amount of new benefits must fit in the given federal budget established by the Budget Annual Law.

The rules give the ministry the right of assigning the families to some priority groups considering the social vulnerability, such as indigenous, *quilombolas*[6], children in a working situation, people released from slavery conditions, and families benefited by the BFP precursor programs. They also allow the ministry to identify if the municipality is in agreement with collaboration terms that promise its engagement in providing complimentary social policies.

After ordering by priority the vulnerable groups and families from municipalities with lower BFP coverage[7], if there are more eligible families than the assigned budget, the system selects the families with the lowest income and more children and teenagers on their composition.

All these rules are applied through an automatic and centralized system that classifies the family as beneficiary. The eligible families not selected to receive the benefit enter in a waiting line. The waiting line rolls with the federal budget assigned for the program in the next year or with other families getting out of the program.

Once selected to the benefits, the families can get out of the program either by voluntarily requiring it for the social assistance office or by the office requiring it, if some irregularity is discovered. Both are relatively rare events[8]. Moreover, the major part of them remains in the SR [9].

The family can be excluded from the SR under some conditions: all family members die; the family is not found for more than two years; the family denies to provide, omits, or lies in the SR information; the family requires its information to be excluded[10].

[6]Residents of settlements established by escaped slaves.

[7]Defined by the current number of beneficiaries divided by the estimated number o poor people in the municipality.

[8]Between 2003 and 2015, 3.1 million families got out of the program by its own will, and 3.0 million had the benefit canceled due to SR data inconsistencies (MC).

[9]Of all the heads of family benefited by the BFP in 2004, 82% appear in the SR data at least in one year after 2012.

[10]Major actions of exclusion of outdated registers happened at the end of 2013 and the beginning of 2015.

The types of benefits the families can receive varies with their income and composition: Basic Benefit of BRL 89.00 is given to extremely poor families; Variable Benefit of BRL 41.00 is given for each pregnant or breastfeeding woman, child and teenager of the family up to the limit of five per family; Variable Benefit for Young (established in 2008) is given for each teenager of 16 an 17 years old. The amount is BRL 48.00 and each family can receive at maximum two; Benefit to Overcome Extreme Poverty is given for the extremely poor families in the amount to set the family per capita income in the extreme poverty limit, i.e., BRL 89.00; The most recent one, established in 2019 is the Christmas Bonus, which is constituted by the same amount of the installment paid for the family in December.

Beyond income, important education and health conditionalities to receive all of these benefits are: keep at least 85% of school attendance for children and teenagers from 0 to 15 years old and 75% for 16 and 17 years old; keep children 0 to 7 years old vaccination card updated; pregnant women must attend the schedule of monitoring exams; breastfeeding women must participate in educational activities; women aged 14 to 44 must regularly check their health conditions.

On education conditions, the public schools, municipalities, states, and the Ministry of Education work together to register the children's attendance in the official systems. Considering health, the Ministry of Health and the municipalities' Health Secretary are in charge of ensuring that the families are matching the requirements, and must provide information about it.

The procedures of checking and reporting if the families were matching the education and health conditionalities were effectively established by the end of 2006 (Soares and Sátyro, 2010), and sharply increased the number of families investigated until 2008. Once some irregularity is found, the consequent actions are divided into five steps. In the first incident, the family receives a warning, and after two months, if the family is still not attending the requirements, the benefit is blocked for one month, although the respective value can be

Therefore, in the SR data of 2012, most of the families excluded are still observed.

withdrawn in the following month. If the family continues not meeting the conditions, the benefit is suspended for one month but at this time the amount will not be available in the next period. In the fourth round, the benefit is suspended for two months. In the last round, the family has the benefit canceled and is removed of the program[11].

Considering the income criteria, some rounds of matching data from the formal labor market with the SR were done by the government to find the non-eligible families that were receiving the benefit. As a result, the Ministry classified the families' information in a range from probably inconsistent to strongly inconsistent. In the latter, the benefit is immediately blocked, but for both cases, the Ministry sends the list of theses families for further investigation by the local BFP manager.

Having presented the rules and the program evolution, we report now the most relevant known short-term effects of the BFP and other CCTs, and what the new long-term effects literature is finding for the other countries' CCT programs.

The effects of the BFP on poverty and inequality reduction are established consensus (Soares et al., 2010), and in line with the findings for other countries, like the Mexican and Chilean programs (Skoufias and Di Maro, 2006; Soares et al., 2009).

Concerning the employment and labor force participation for the adults, the results seem to be mixed. For BFP some authors found no effect (De Brauw et al., 2015; Foguel and Barros, 2010), while others found positive effects on probability of studying and working at the same time (Chitolina et al., 2016) and on the duration of the formal employment (Santos et al., 2017). The same is true for other programs. For instance, Parker and Skoufias (2000) and Skoufias and Di Maro (2006) found no effect for the Mexican program, while Behrman and Parker (2013) found positive effects.

Side effects like regional inequality reduction (Silveira Neto and Azzoni, 2011), and reduction of intended migration influenced by differential of BFP coverage between municipalities (De Oliveira and Chagas, 2020) were also approached by the literature, but none with a

[11]From 2003 to 2009 all of the families received warnings, 36.6% got the benefit blocked, and only 4.5% got the benefit canceled.

long-term focus.

Positive effects on schooling are more robust than the labor market participation. For the BFP these effects were shown for schooling attainment in Silveira-Neto and Duarte (2010) and enrollment in (Chitolina et al., 2016), in line with the findings for other programs (Bastagli et al., 2016).

As some of the pioneer CCT programs are 15 to 20 years old, now we present some of the findings of the emerging literature of long-term effects, that can be seen in detail in Millán et al. (2019). In general, effects on schooling are mostly positive (Behrman et al., 2011; Barham et al., 2017; Millán et al., 2020). For cognitive development, the reported study for Nicaragua showed that exposed children presented better outcomes 10 years later (Barham et al., 2013).

Impacts on employment and labor market participation of the young adults are mixed. Some authors found no effect on the grown children participation in the Labor Market (Millán et al., 2020) while others found increase of participation (Parker and Vogl, 2018). The results on employment likely depend on the length of time observed after the individuals became adults. These mixed results for labor market may be expected if it is considered that individuals may be postponing their entrance in the labor market to study more. Therefore, effects on participation and earnings may appear only some years later.

Most of these mentioned studies take advantage of the random assignment of each program, either at the individual (Barrera-Osorio et al., 2019) or at a more aggregate level (Parker and Vogl, 2018). For the BFP, no random assignment was implemented, so that we had to build our Natural Experiment with a careful econometric identification strategy. Maybe, this explains why, to the best of our knowledge, there is no evidence for the long-term effects of the largest CCT in the world. The present research intends to fill this vacuum, and will hopefully be followed by others.

3 Methodology

This section presents our identification strategy, the econometric specification of each outcome of interest, and the data applied to estimate the long-term effects.

Our main strategy relies on a combination of properly selecting the sample, conditioning on many control variables, and exploring the timing and rules of implementation of the program. In robustness tests we apply an instrumental variable approach that corroborates our results.

In our sample, we select the individuals born between 1996 and 1998, from families that were first registered in the SR from 2004 to 2006, a strong expansion period. The cumbersome task of registering millions of vulnerable families in a narrow time window certainly responds for a random part of the observed variation in the timing that each family is incorporated on the records, and then potentially start receiving the BFP benefits. It is very unlikely that these families were beneficiaries of other programs before the BFP, since the other programs by this time already demanded them to be registered in the SR.

These cohorts of individuals are old enough to be beneficiary of the BFP during childhood in the first year of BFP Payment Records (2004). They are also teenagers in the first year of SR data (2012), which makes it possible to collect their family information before they eventually build their own family. Moreover, these individuals are also old enough to have finished high school and participate in the FLM in the most recent year of the data, i.e., more than 18 years old in 2017.

Unfortunately, we do not observe the real income of the family at the moment of registration, but even if we could, we would still observe a variation in the individuals' exposure to the program. The variation not related to the family eligibility rules emerges from three different sources.

The first source of variation can be seen as a measurement error. It comes from the difference between the average real income of the family during the individuals' childhood,

and the income that the family declares[12] to the social assistant at the moment of SR registration, or the moment of information update[13]. Beyond being a reasonable argument, it is supported by the evidence from the data. We observe individuals from families with the same parents' average formal income, but different BFP Exposure levels. Moreover, once the family turns its status to BFP beneficiary, the condition to cancel or suspend the benefit is the existence of strong evidence pointing to the family income above the eligibility. However, it seems unreasonable to expect that small income differences would be easily identified by the local BFP managers.

The second source of variability is also related to the family. It is related to the timing that the family gets registered in the SR. To put it simply, eligible families that are achieved by the SR earlier tend to become beneficiaries of the BFP before the families achieved later.

Finally, the government centralized procedure of release of BFP benefits tries to balance the estimated coverage of the program across municipalities. The program coverage is computed by the current benefited families divided by the estimated number of poor families. Then, the third source of variation occurs between municipalities. For instance, families from municipalities that included many people in the registry and currently have higher BFP coverage could be assigned to a waiting list. This would occur if the annual budget of the Federal Government destined to the BFP was not enough to release the benefit for all eligible families at that moment.

3.1 Schooling

The variable chosen to represent the education attained by the child when it turns into an adult is the schooling level when the the individual is exactly 18 years old. There are two reasons for this age choice. First, individuals are already legally responsible for themselves, and in regular conditions would have just completed the secondary education[14]. This educa-

[12] We do not observe the family declared income in the SR previous to 2012.

[13] It is expected that the program local manager keeps the vulnerable families' information always updated in less than two years.

[14] In Brazil, a regular student completes secondary education (high-school) when it is 17 years old.

tion level is considered mandatory for any citizen and the government, at least in theory[15], is obliged to provide it free of additional costs for the entire population. The second reason is related to the program rule. The education follow-up of these early adults was important for the family to get the BFP benefits until one year ago, therefore the education variables observed are expected to contain more reliable and updated information than for individuals of later ages.

Through the information of literacy and grades attendance, it was possible to transform all the variables of education in the Single Registry into one variable of Schooling, divided into six levels with a natural increasing order. It starts at level 0, that corresponds to individuals that never went to school and do not know how to read or write, and stops at level 5, which comprises the individuals that finished tertiary education[16]. The Schooling level 1 represents individuals that finished pre-school or learned how to read and write. The level 2 comprises people that concluded elementary education[17]. Level 3 is defined by individuals that concluded the equivalent to middle school[18]. Finally, Schooling level 4 represents individuals that finished high-school, that in regular conditions is expected to happen at the age of 17 years old.

Considering the natural ordered characteristic of this outcome variable, we apply an Ordered Discrete Choice model, the Ordered Logit[19], a particular case of Ordinal Regression models that were first considered by McCullagh (1980). This model allows us to retrieve how the probability of achieving some level of education varies along with the program exposure levels. Besides the estimated probability, the sign of the parameters also gives the direction of the effects of the program exposure on schooling. Following the textbook of Wooldridge (2010), the model can be derived from a latent variable equation represented in Equation 1.

[15]Act. n. 9394 from December 20, 1996

[16]As was shown in the data description section, there are relatively few individuals with Schooling level 5, which is expected if the assumed age of reference (18 years old) is considered.

[17]In the Brazilian system, for a regular student, this is expected to happen at the age of 10 years old, in the conclusion of the fourth grade, currently renamed to fifth year

[18]In Brazil, this is expected to happen at the age of 14, with the conclusion of the 8^{th} grade, currently renamed to 9^{th} year.

[19]The error distribution assumptions, usually Normal or Logistic, pointed to the same results.

$$Schooling_i^* = Exposure_i\beta_1 + X_i\beta_x + \epsilon_i \tag{1}$$

The $\epsilon_i|Exposure_i, X_i$ is assumed to follow a Logistic distribution, $Exposure_i$ is the vector with the four interest dummy variables representing each level exposure to the BFP during childhood, from 1 to 4, and β_1 is the vector with the respective parameters of interest. The level of BFP Exposure 0 is omitted, which means that all results reported using these measures of program exposure must be interpreted in differential to the non-exposed group. The X_i comprises the vector with all the individual, family, and municipality control variables, without the constant term. β_x is the vector with the respective parameters to be estimated.

Let $\alpha_1 < \alpha_2 < \alpha_3 < \alpha_4 < \alpha_5$ be different unknown intercepts for each Schooling level and define the following variable that we observe, we have Equation 2.

$$Schooling_i = \begin{cases} 0, \text{ if } Schooling_i^* \leq \alpha_1 \\ 1, \text{ if } \alpha_1 < Schooling_i^* \leq \alpha_2 \\ 2, \text{ if } \alpha_2 < Schooling_i^* \leq \alpha_3 \\ 3, \text{ if } \alpha_3 < Schooling_i^* \leq \alpha_4 \\ 4, \text{ if } \alpha_4 < Schooling_i^* \leq \alpha_5 \\ 5, \text{ if } \alpha_5 < Schooling_i^* \end{cases} \tag{2}$$

Given the distribution assumed for ϵ_i, we retrieve the set of parameters finding the ones that maximize the likelihood of observing the data we have, i.e., we apply the method of Maximum Likelihood[20]. The log-likelihood function of the individual i is given by Equation 3.

[20]This model is estimated through the open-source software, R, with the function *polr* of the library *MASS*.

$$
\begin{aligned}
& l_i(\alpha, \beta_1, \beta_x) = \\
& 1[\text{Schooling}_i = 0]\ log[F(\alpha_1 - Exposure_i\beta_1 - X_i\beta_x)]\ + \\
& 1[\text{Schooling}_i = 1]\ log[F(\alpha_2 - \beta_1 Exposure_i - X_i\beta_x) - F(\alpha_1 - Exposure_i\beta_1 - X_i\beta_x)] + \\
& \ldots + \\
& 1[\text{Schooling}_i = 5]\ log[1 - F(\alpha_5 - Exposure_i\beta_1 - X_i\beta_x)]
\end{aligned} \tag{3}
$$

Where $F(.)$ is the logit function[21], the terms $1[Schooling_i = x]$ represent indicator variables equal to one, if the schooling level of the individual i is equal to x, and zero otherwise. In other words, for each individual, only one of the terms is different from zero. Then, the function to be maximized is the sum of the log-likelihood functions of all individuals in the sample.

$$
LL = \sum_{i=1}^{N} l_i(\alpha, \beta_1, \beta_x) \tag{4}
$$

Maximizing the function given by Equation 4 gives us the parameters of interest $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_x)$. Then, it is possible to estimate the probability of an individual to exhibit each of the Schooling levels conditioned on the level *Exposure* to BFP during its childhood, and also conditioned on the other observed characteristics X_i. In the results section, we report the estimated probabilities over the BFP Exposure levels, keeping the control variables fixed on their means[22].

The predicted probability of an individual with a BFP exposure of $Exposure_i$ and a set of characteristics X_i, to present the level of Schooling j, is then given by Equation 5, and the standard errors of this predicted probability can be obtained through Delta Method.

[21] $F(x) = e^x/(1 + e^x)$

[22] For the discrete control variables, we mention explicitly the assumed values.

$$P(Schooling_i = j|Exposure_i, X_i; \hat{\alpha}, \hat{\beta}_1, \hat{\beta}_x) = \\ F(\hat{\alpha}_j - Exposure_i\hat{\beta}_1 - X_i\hat{\beta}_x) \ - \ F(\hat{\alpha_{j-1}} - Exposure_i\hat{\beta}_1 - X_i\hat{\beta}_x) \tag{5}$$

3.2 Formal Labor Market Participation

To represent the Formal Labor Market participation and analyze how it varies with the BFP exposure, we define a binary variable, FLM_i, that indicates if the individual i appears at least one time in the formal labor market data until the most recent available year, 2017.

Following the textbook of Wooldridge (2010), we apply one of the most common probability model, the Logit. As is the case for the ordered Logit, presented in the previous section, the logit can also be derived from a latent variable model.

$$FLM_i^* = \beta_0 + Exposure_i\beta_1 + X_i\beta_x + \epsilon_i \tag{6}$$

This means that the probability of being observed in the formal labor market ($FLM_i = 1$) is given by Equation 7

$$P[FLM_i = 1|Exposure_i, X_i] = P[FLM_i^* > 0|Exposure_i, X_i] = \\ \text{P}[\ \epsilon_i > -\beta_0 - Exposure_i\beta_1 - X_i\beta_x] = \\ \text{F}(\ \beta_0 + Exposure_i\beta_1 + X_i\beta_x) \tag{7}$$

The last equality holds because the distribution of ϵ_i is symmetric around zero. The log-likelihood function for each individual is then given by Equation 8.

$$l_i(\beta_0, \beta_1, \beta_x) = \\ \text{FLM}_i log[F(\beta_0 + Exposure_i\beta_1 + X_i\beta_x)] + \\ (1 - \text{FLM}_i) log[1 - F(\beta_0 + Exposure_i\beta_1 + X_i\beta_x)] \tag{8}$$

The $F(.)$ function is the accumulated density function evaluated on a set of parameters

$(\beta_0, \beta_1, \beta_x)$ and the values observed for the BFP Exposure and the control variables X_i. Of course, only one of these terms will be different from zero for each individual. The function to be maximized by the solution set of parameters is the sum of the individuals log-likelihood functions:

$$LL = \sum_{i=1}^{N} l_i(\beta_0, \beta_1, \beta_x) \tag{9}$$

As was done for the Schooling, for the FLM outcome we also report the predicted probability, given the solution set of parameters $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_x)$, of an individual with an observed level of $Exposure_i$ and observed characteristics X_i to appear in the Formal Labor Market. This is computed through Equation 10. The standard errors of this predicted probability are computed through the Delta Method.

$$P[FLM_i = 1|Exposure_i, X_i; \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_x] = F(\hat{\beta}_0 + Exposure_i\hat{\beta}_1 + X_i\hat{\beta}_x) \tag{10}$$

The alternative identification strategy relies on the exogeneity of the municipality effort to include vulnerable families in the Single Registry. We measure the effort as the number of registered families in 2006 divided by the number of poor families in 2000[23].

On the implementation of this strategy for the FLM outcome, we apply the Instrumental Variable Probit[24], and changed the interest variable from Levels of Exposure (0 to 4) to the Years of Exposure (0 to 10) and its squared term, to capture the non-linearity observed in the estimates of other specifications.

Equations of this approach are stated in 11, 12, and 13.

[23]In 2006, this was the most recent year for which there was comparable poverty survey data for all municipalities. The Demographic Census of 2000 was also an official source of poverty measures considered by the government

[24]The Multivariate Normal Distribution of the errors makes the Maximum Likelihood estimates more straightforward in this case than a Multivariate Logistic assumption.

$$FLM_i^* = \beta_0 + Exposure_i\ \beta_1 + Exposure_i^2 \beta_2 + X_i \beta_x + \epsilon_i \tag{11}$$

$$Exposure_i = \gamma_0 + +SR\_Coverage_i\ \gamma_1 + SR\_Coverage_i^2\ \gamma_2 + X_i \gamma_x + u_i \tag{12}$$

$$Exposure_i^2 = \delta_0 + SR\_Coverage_i\ \delta_1 + SR\_Coverage_i^2\ \delta_2 + X_i\ \delta_x + v_i \tag{13}$$

Equation 11 is known as the structural equation of this system, and Equations 12 and 13 are known as the first stage regressions, that work as exogenous source of variation to the interest variables, the Exposure and its squared term. Because we are interested in the probability of participation in the formal labor market and want to respect the nature of the observed binary variable FLM_i, we follow the structural approach presented in (Cameron and Trivedi, 2009) and assume that the error terms of the three equations $(\epsilon_i, u_i, v_i) \sim N(\mathbf{0}, \Sigma)$. The estimation procedure is the Maximum Likelihood.

3.3 Earnings in the Formal Labor Market

To estimate the long-term effects of the BFP on the early adults' earnings, we chose to apply the dependent variable defined by the nominal hourly wages in the logarithmic scale. This is the most used specification of empirical earnings equations (Heckman et al., 2006) in the economic literature, but here we do not control for schooling or experience as is usually the case. We had to keep schooling and experience out, because these may be also outcomes of the treatment exposure, and as potential mechanisms driving the earnings, can take us to misleading interpretations. In other words, schooling and experience may act as bad controls that spoil the estimation of the effect of interest: the differential effect of BFP exposure on Earnings conditional on having at least one formal employment.

The wage regressions of the first identification strategy are estimated through the Ordinary Least Squares method. The model specification is given by equation 14.

$$Wage_{it} = \beta_0 + Exposure_i\beta_1 + X_i\beta_2 + \phi_t + \epsilon_{it} \tag{14}$$

The $Wage_{it}$ is measured as the logarithmic of the hourly average nominal wage of individual i reported by its firm at the year t. The k control variables are represented by vector X_i. Therefore, parameter β_2 belongs to the vector space of dimension k. The term ϕ_t represents the dummy variable for the year t, and ϵ_{it} represents the non-observed random shock component of the wage.

We chose to apply the panel data for the Earnings outcome instead of the cross-section for one reason. In the panel, we do not need to create another measure of Earnings making more assumptions to consider, for example, inflation. Moreover, we can incorporate dummy variables of year and individuals' cohort. These variables are expected to capture macroeconomic cycles and eventual correlations between these cycles with the BFP Exposure that emerge from the relative presence in the labor market of each cohort in each year.

For the Earnings outcome, we also apply the alternative identification strategy considering the Instrumental Variable approach. We take advantage of the observation of the municipality effort to include families in the Single Registry in the period of great expansion of the BFP. From the family point of view, it is reasonable to argue that, conditioned on variables like pre-program municipality poverty, density, population size, the effort of the social assistance team is not correlated to the long-term outcome variable by other channels than the BFP participation.

The drawback of this approach is that the variation of the instrument is at the aggregate level, while the endogenous variable, BFP exposure, is at the individual level. Moreover, the interest variable in this specification is the number of years that the individual was exposed to the program during childhood, and not the 5 levels measure as in the other models. This change was necessary because of the existence of only one instrument, and the previous measure would demand one instrument for each level of treatment exposure.

To tackle the eventual remaining endogeneity concerning the family exposure, we apply the Two-Stage Least Squares approach to retrieve the parameters of Equation 15.

$$Wage_{it} = \beta_0 + Exposure_i\,\beta_1 + Exposure_i^2\,\beta_2 + X_i\,\beta_3 + \phi_t + \epsilon_{it} \tag{15}$$

The first stage regressions that allow us to keep the variation on the exposure variables not related to the family unobservables are represented by equations 16 and 17.

$$Exposure_i = \gamma_0 + SR\_Coverage_i\gamma_1 + SR\_Coverage_i^2\gamma_2 + X_i\gamma_3 + \phi_t + u_{it} \tag{16}$$

$$Exposure_i^2 = \delta_0 + SR\_Coverage_i\delta_1 + SR\_Coverage_i^2\delta_2 + X_i\delta_3 + \phi_t + v_{it} \tag{17}$$

The data structure in this instrumental variable approach remains the same. In other words, although the BFP exposure does not vary along the time dimension for a given individual, it is repeated for every wage observation that it has. This explains the error terms u_{it} and v_{it} dimensions, and the year dummy variable ϕ_t on both first-stage equations.

With the predicted $Exposure$ and $Exposure^2$ replacing the original variables, the second stage regression is stated in Equation 18.

$$Wage_{it} = \beta_0 + \hat{Exposure}_i\,\beta_1 + +\hat{Exposure}_i^2\,\beta_2 + X_i\,\beta_3 + \phi_t + \epsilon_{it} \tag{18}$$

This approach could be very limited if the country had only a few municipalities. Fortunately, this is not the case of this work. Brazil has more than 5.5 thousand municipalities, which provides a variation that enables the identification of the desired parameters. However, this strategy has the cost of throwing away two important parts of the BFP exposure variation: the one related to the timing that each family got registered in the SR; and the one related to the difference between the family income at the moment of SR registration/update and the average income.

3.4 Data

To estimate the presented models we combined three main primary data[25] to obtain the characteristics of the individuals and their families. They are:

Single Registry for Social Programs[26] (SR) from the Ministry of Social Development[27] (MDS): This data comprises the individual and its family identification and characteristics. We also observe the family BFP condition, beneficiary, or not. Moreover, it contains the date of the family registration, and information like zip code, income per member and the family's expenses. The available years are 2012 to 2017.

BFP Payment Records - Ministry of Social Development: This data is used to estimate the treatment of BFP on children from 2004 to 2011 since the SR is not available for this period. It has the identification document number of the family responsible (head), and the total value of BFP received per year. Knowing the family of the benefited individual in the SR allows us to build the variable of program exposure for the selected cohorts, composed of individuals born between 1996 and 1998.

Annual Report of Social Information[28] (RAIS) from the Ministry of Labor and Employment[29] (MTE): This dataset comes from a form that every firm in the country may fill every year about their employees. It contains the nominal wage, dates of admission and lay-off, in addition to other personal and firm characteristics. It is available from 2002 to 2017.

Combining these three sources with municipalities indicators from the Demographic Census of 2000, the Ministry of Education, and Ministry of Social Development, produced two final datasets. The first is a panel where we observe the wages and employment characteristics of individuals of the selected cohorts. The first year is 2012, which is the first year when

[25] The procedures adopted to sample, merge and prepare the data are the same presented in De Oliveira (2020).

[26] *Cadastro Único para Programas Sociais*. This identified data was obtained under special authorization linked to the SAGI/MDS Process number 71000.018838/2018-35.

[27] *Ministério do Desenvolvimento Social e Agrário*, currently restructured and renamed to *Ministério da Cidadania* (Ministry of Citizenship).

[28] *Relatório Anual de Informações Sociais*.

[29] *Ministério do Trabalho e Emprego*. This Ministry was dissolved in 2019, and its activities are currently shared between the Ministry of Economy and the Ministry of Citizenship.

the older cohort (1996) could be formally employed, and the last is 2017.

Even though we observe in this data the wages over the adulthood years, we only observe one fixed BFP exposure measure for each individual during its childhood. This fact is related to the nature of the problem and prevented the consideration of individual-fixed-effects models. However, considering the set of control variables observed, disregarding individual fixed-effect seem to pose limited damage to the estimates.

We chose to consider this panel data in the baseline for the Earnings outcomes because it allows to control for cohort and year fixed effects. The fixed effects clean the estimates from macroeconomic effects, and individual experience and maturity effects on the wages. An alternative approach to the Earnings outcome models would be to apply a national inflation index and take the average wage, transforming the panel in a cross-section. However, we believe that this procedure could introduce additional noise in the estimates[30].

Considering the wages of the individuals' parents during their childhood, we had to apply the inflation index. The years in which we observe the interest cohorts in the labor market are not the same as the years of their parents during their childhood. The same is true for the parents schooling, for which we took averages for the values observed in the years of SR before the individual become an adult (18 y.o.).

The second dataset is a cross-section derived from the panel. In this data, are observed the Schooling attained when individuals are 18 years old and their Formal Labor Market Participation. The data also include information on the individuals' parents restricted to their childhood period, like the average schooling attained, Formal Labor Market participation, and average Earnings. All variables are time-fixed, which justifies why we keep only one observation per individual.

[30]Estimates considering the cross-section data and the averaged real wages as dependent variable also presented mixed results for the Earnings outcome. In this specification, the nominal wages were corrected with one of the official inflation indexes, the INPC (IBGE), before calculating the averages.

4 Results

4.1 Descriptive Statistics

In this section, we describe the primary data, approaching the total number of individuals observed in the SR, the BFP Payment Records, and the Formal Labor Market (RAIS) over the years. Then, we turn to the merging statistics, showing the number of individuals from each dataset found in the other ones. We also generate the same information selecting specific periods of each dataset. This procedure allows us to show that the problem of attrition in this sample seems to be very small.

The strong expansion in the first years of the Single Registry and the BFP benefits constitute the most important feature of the program for the purpose of this work. We take advantage of the timing of this expansion to identify the program's effects. These characteristics are observable in the next two tables.

Table 1 shows the number of members of families registered in the Single Registry from 2001 to 2017. The second column reports the total of individuals that, in their first year of appearance in the SR data, belonged to a family registered in the corresponding year of the table. With the data of this column, we computed the share of individuals whose first family were registered between 2004 and 2006. This group corresponds to 17.8% of all registered individuals, and 24.85% of the individuals from the SR that were born between 1996 and 1998. We focus the analysis on this last group, composed of 1,895,925 individuals.

Table 1: Registered Individuals in the SR

Year	Total (Accumul.)	Current Year	% of Total	Cohorts 96-98 (Accumul.)	Current Year	% of Total
2001	732,834	732,834	0.64	58,970	58,970	0.77
2002	16,620,703	15,887,869	13.83	1,523,592	1,464,622	19.20
2003	31,212,495	14,591,792	12.70	2,770,157	1,246,565	16.34
2004	38,017,285	6,804,790	5.92	3,452,296	682,139	8.94
2005	41,535,749	3,518,464	3.06	3,786,641	334,345	4.38

Table 1: Registered Individuals in the SR

Year	Total (Accumul.)	Current Year	% of Total	Cohorts 96-98 (Accumul.)	Current Year	% of Total
2006	51,665,672	10,129,923	8.82	4,666,082	879,441	11.53
2007	58,350,552	6,684,880	5.82	5,160,863	494,781	6.49
2008	63,764,198	5,413,646	4.71	5,518,378	357,515	4.69
2009	69,963,859	6,199,661	5.40	5,894,279	375,901	4.93
2010	77,031,813	7,067,954	6.15	6,258,651	364,372	4.78
2011	82,526,440	5,494,627	4.78	6,471,542	212,891	2.79
2012	91,317,252	8,790,812	7.65	6,762,387	290,845	3.81
2013	97,004,927	5,687,675	4.95	6,949,148	186,761	2.45
2014	102,458,254	5,453,327	4.75	7,137,037	187,889	2.46
2015	106,259,807	3,801,553	3.31	7,273,476	136,439	1.79
2016	110,159,950	3,900,143	3.39	7,436,330	162,854	2.13
2017	114,906,578	4,746,628	4.13	7,628,572	192,242	2.52

Note: This table reports the yearly evolution of the SR. The second column presents the accumulated number of registered individuals until that given year. The third column reports the total of individuals registered on each year. The fourth computes the value of the second column divided by the final number of registered individuals, in 2017. The following three columns report the same measures as the previous three but considering only individuals that were born between 1996 and 1998. Source: Author's calculations with the data from MDS.

With the data from the National Demographic Census of 2010 of the Brazilian Institute of Geography and Statistics[31] (IBGE), we estimate the Brazilian population born between 1996 and 1998 in a total of 10,258,680. This means that around three-quarters of the individuals born in these years entered in the Single Registry at some point of their lives. In other words, they were members of an economically vulnerable family. Moreover, this emphasizes that the sample represents a big part of the country population within these birth years.

Table 2 shows the granting of BFP benefits from 2004 to 2017. The second column reports the total of different families benefited by BFP accumulated until each year. For instance,

[31] *Instituto Brasileiro de Geografia e Estatísica.*.

as of 2017, 31 million families were benefited by the program at least one month from 2004 until 2017. In the third column, we present the number of families that received the benefit at least in one month of each year. These columns show that the program achieved its peak around 2013, which contrasts with the increasing official indexes of poverty in Brazil after this year.

We do not observe the total of families that were benefited by BFP previous to 2004, but with the total of benefited families in January of 2004, it is reasonable to assume that the BFP expansion was strong in this year.

Table 2: BFP Benefited Families over the years

Year	Families Accumulated	Currently Benefited	% of Total
January 2004	3,598,484		
2004	6,681,581	6,681,581	21.53
2005	8,989,951	8,969,428	7.44
2006	12,634,727	12,285,626	11.74
2007	13,851,885	12,326,107	3.92
2008	14,816,545	12,280,090	3.11
2009	17,448,327	13,914,796	8.48
2010	19,238,621	14,327,731	5.77
2011	20,984,811	14,592,121	5.63
2012	22,274,294	14,891,010	4.15
2013	24,478,012	16,347,545	7.10
2014	25,956,655	15,680,465	4.76
2015	27,379,479	16,075,319	4.58
2016	29,096,630	15,931,219	5.53
2017	31,038,939	16,020,743	6.26

Table 2: BFP Benefited Families over the years

Year	Families Accumulated	Currently Benefited	% of Total

Note: The table reports the BFP number of benefited families over the years. The second column presents the accumulated number of families at least once benefited by BFP in previous years. The third column presents the number of current benefited families each year. The last column divides the current number by the final (2017) accumulated number of benefited families. As the beneficiary families previous to 2004 are not observed, we report the number of January of that year as a proxy for the families benefited in 2003. Source: Author's calculations with the data from MDS.

Table 3 concerns the third data source of this work, where we report the numbers of the Formal Labor Market from 2002 to 2017. We observe a total of 107,251,994 distinct individuals from 2002 until 2017. Two economic phenomenons that happened in Brazil during these years come to attention. The economic growth and the formalization of the labor market are represented in this table by the sharp increase in the number of formally-employed individuals from 2002 until 2014. Afterward, the effects of the economic crisis that began in that year are also represented in this table by the decrease in the number of formally-employed individuals.

From all of the formal employees, the share that appears in the Single Registry ranges from 22% to 33% over the years. Concerning the cohorts of interest of this work, there is an underestimate of 4,839,089[32] distinct individuals that were born between 1996 and 1998. It corresponds to approximately half of the 10 million residents of Brazil that were born in the same years, participating in the formal labor market at least once in this time range. Interestingly, also around half of these young individuals formally employed appear at some point in families registered from 2001 to 2017 in the Single Registry.

Furthermore, in the columns corresponding to the selected cohorts this table shows a

[32]This is an underestimate because we do not observe the individuals' birth-year in the RAIS data for the years of 2011, 2012, and 2013.

sharp increase from 2012 to 2017. This is expected if it is considered that the cohorts of 1996, 1997, and 1998 just turned 16 years old, the legal age to be formally employed, respectively in the years of 2012, 2013, and 2014.

Table 3: Total of Individuals observed in the Formal Labor Market (RAIS)

Year	N	N in the SR	N Cohorts 1996 to 1998	N Cohorts 1996 to 1998 in the SR
2002	34,950,398	7,796,335	0	0
2003	35,925,326	8,008,151	0	0
2004	37,856,735	8,631,268	0	0
2005	40,179,150	9,201,687	0	0
2006	42,486,868	10,123,408	0	0
2007	45,227,446	10,967,784	0	0
2008	48,573,811	12,282,379	0	0
2009	50,219,948	12,947,561	0	0
2010	53,771,613	14,397,004	0	0
2011	56,641,168	16,418,381	N/A	0
2012	58,738,850	17,586,211	N/A	137,712
2013	60,450,823	19,091,211	N/A	382,494
2014	61,492,767	20,014,778	1,662,430	821,619
2015	59,856,891	19,567,103	2,437,596	1,228,197
2016	56,906,493	18,590,272	2,980,462	1,492,256
2017	55,561,692	18,713,269	3,498,993	1,822,894

Note: The table presents the evolution of the number of observed individuals in the Formal Labor Market. The second column presents the total distinct individuals observed each year. The third column presents the number of individuals from the second column that are observed at least one year in the Single Registry. The following two columns present the same numbers considering only individuals that were born between 1996 and 1998. Source: Author's calculations with the data from MDS and MTE

In the previous paragraphs and descriptive statistics, it was demonstrated the evolution over time of the aggregate numbers concerning the three identified data, the Formal Labor Market, the Single Registry, and the BFP benefits payment. We believe that it is clear that the expansions of the SR and the BFP in the initial years were very strong, which enables us to assume that a phenomenon like this is certainly charged with a lot of randomnesses. Moreover, it should be clear that the data comprises a large part of the Brazilian population.

The last descriptive statistics about aggregate numbers are reported in Table 4. We put apart the interest cohorts for a moment to present some aggregate numbers of the linking procedure applied to the three datasets. We concentrate on individuals' identification

documents and the specific time ranges that they appear on each data source.

Table 4: Aggregate Statistics of the Data Linking Process

	SR 2012-2017	BFP 2004-2017	BFP 2004	FLM 2002-2017	FLM 2012-2017
SR 2012-2017	**114,906,578**				
BFP 2004-2017	28,194,877	**31,038,939**			
BFP 2004	5,444,694	6,681,581	**6,681,581**		
FLM 2002-2017	35,537,005	11,285,157	1,705,291	**107,251,994**	
FLM 2012-2017	30,313,216	9,131,623	1,341,976	86,010,351	**86,010,351**

Note: This table presents statistics of the linking process of the Single Registry (SR) Formal Labor Market (FLM) and BFP Payment Records (BFP). It presents the number of individuals observed in both data sets represented in the names of the lines, and names of the columns. Each unit of the BFP data represents a family since only the head of the family appears in the payment records. For the SR and FLM, the unit is the individual. Source: Author's calculations with the data from MDS, and MTE.

In Table 4 the Single Registry (SR) numbers consider each individual of each family registered. The BFP records only comprise the head of each family, which is the person that receives directly the benefits. The Formal Labor Market data consider each different individual that appears in the labor market during the indicated periods.

From the 6.68 million heads of family that appear in BFP Payment Records of 2004, 1.24 million (18.51%) do not appear in the SR data. Of this 1.24 million, only 80.4 thousand individuals appeared in the formal labor market after 2012. We can use this number to get an estimate of the attrition problem in this sample.

Assume for these missing heads-of-family the formal employment rate of 30%, which is a magnitude close to the observed for the heads-of-family that are not missing in the SR. Then, the estimate of total missing individuals is 268 thousand (equivalent to 80.4 divided by 30%). This represents only 4% of the individuals that received at least once the BFP benefit from 2004 to 2017.

If we extrapolate the heads-of-family's SR disappearance rate to all family members, only 4% of the individuals that were registered before 2012 would have been excluded from the SR data. This is a relatively small attrition rate. Moreover, although we have no means to state

what happened to these missing individuals, the most probable reason is that they improved a lot their economic condition so that they became non-eligible to the BFP and non-eligible to the SR, which means they were not vulnerable anymore. Under this condition, they may have been excluded by ordering it, or by getting caught by the local manager of the program. Both reasons pose a small threat to the identification of the effects of BFP exposure, but the biases in these cases are toward the underestimation of positive effects.

From the 7,628,572 individuals from selected cohorts in the SR, 1,895,925 were from families registered from 2004 to 2006. After sampling 10% of them and applying the cleaning procedures to keep only reliable information, 149,023 individuals remain. This illustrates the quality of the three identified data. The amount of individuals lost because they were not children of the head of the family in the first SR appearance, or because of inconsistencies in the document number or other fixed characteristic, responds for only 21.4%.

Table 5 shows descriptive statistics for all of the explanatory and outcome variables for the older selected cohort, formed by those that were born in the year of 1996. The respective tables[33] for the cohorts of 1997 and 1998 have equivalent results and could replace this one without any damage to the interpretations and arguments provided.

Table 5: Descriptive statistics across BFP Exposure - Cohort 1996

		BFP Exposure Level			
Variable	Measure	0	1	2	3
Individuals	N	3,062	9,416	20,421	15,584
Schooling attained at 18 y.o.	Mean	1.96	2.25	2.36	2.47
	Std. Dev.	0.94	0.87	0.81	0.77
	Min	0.00	0.00	0.00	0.00
	Max	5.00	5.00	5.00	5.00
Formal Employed	Mean	0.45	0.47	0.42	0.42
	Std. Dev.	0.50	0.50	0.49	0.49
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
Avg. Real Hourly Wage	Mean	7.41	7.37	7.30	7.32
	Std. Dev.	2.16	2.79	2.47	2.53
	Min	2.03	1.55	1.94	1.97

[33]The tables with descriptive statistics for the other cohorts are reported in the Appendix.

Table 5: Descriptive statistics across BFP Exposure - Cohort 1996

		BFP Exposure Level			
Variable	Measure	0	1	2	3
	Max	30.68	50.33	47.44	48.71
Years BFP Exp.	Mean	0.00	2.37	5.48	7.58
	Std. Dev.	0.00	0.76	0.74	0.49
	Min	0.00	1.00	4.00	7.00
	Max	0.00	3.00	6.00	8.00
Age at First Employment	Mean	18.32	18.32	18.55	18.62
	Std. Dev.	1.47	1.50	1.46	1.43
	Min	16.00	16.00	16.00	16.00
	Max	21.00	21.00	21.00	21.00
Female	Mean	0.48	0.49	0.48	0.48
	Std. Dev.	0.50	0.50	0.50	0.50
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
White	Mean	0.38	0.36	0.29	0.29
	Std. Dev.	0.49	0.48	0.45	0.45
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
Migrant	Mean	0.43	0.44	0.41	0.38
	Std. Dev.	0.50	0.50	0.49	0.49
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
Formal Head	Mean	0.40	0.41	0.30	0.25
	Std. Dev.	0.49	0.49	0.46	0.43
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
Co-head	Mean	0.58	0.54	0.54	0.53
	Std. Dev.	0.49	0.50	0.50	0.50
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
Formal Co-head	Mean	0.33	0.31	0.26	0.24
	Std. Dev.	0.47	0.46	0.44	0.43
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
Co-head Avg Real Hourly Wage	Mean	8.40	7.88	7.46	6.96
	Std. Dev.	4.82	4.04	4.24	3.77
	Min	2.08	1.50	1.49	1.84
	Max	55.83	54.91	56.15	54.82
Head Avg Real Hourly Wage	Mean	6.37	6.04	5.91	5.74
	Std. Dev.	3.50	2.89	2.99	3.13
	Min	1.85	1.54	1.45	1.43

Table 5: Descriptive statistics across BFP Exposure - Cohort 1996

		BFP Exposure Level			
Variable	Measure	0	1	2	3
	Max	43.59	48.66	51.08	56.66
Head Schooling	Mean	2.20	2.12	1.90	1.88
	Std. Dev.	1.20	1.14	1.13	1.08
	Min	0.00	0.00	0.00	0.00
	Max	5.00	5.00	5.00	5.00
Co-head Schooling	Mean	1.96	1.86	1.65	1.60
	Std. Dev.	1.13	1.11	1.12	1.09
	Min	0.00	0.00	0.00	0.00
	Max	5.00	5.00	5.00	5.00
Family Size	Mean	3.94	4.12	4.42	4.53
	Std. Dev.	1.35	1.41	1.58	1.58
	Min	1.00	1.00	1.00	1.00
	Max	13.00	14.00	16.00	14.00
Munic. Density 2000	Mean	951.58	919.94	1,042.92	1,187.86
	Std. Dev.	2,046.99	1,943.48	2,148.06	2,199.27
	Min	0.22	0.19	0.13	0.16
	Max	11,924.99	11,924.99	11,924.99	11,924.99
Munic. Families 2000	Mean	229.07	188.43	204.68	249.94
	Std. Dev.	652.93	505.20	606.92	637.36
	Min	0.53	0.35	0.30	0.27
	Max	3,173.50	3,173.50	3,173.50	3,173.50
Munic. Poor Fam 2000	Mean	0.47	0.46	0.52	0.52
	Std. Dev.	0.20	0.19	0.21	0.21
	Min	0.11	0.10	0.07	0.10
	Max	0.96	0.93	0.96	0.93
Munic. Education Quality 2007	Mean	4.13	4.13	3.87	3.87
	Std. Dev.	0.74	0.76	0.80	0.74
	Min	1.80	1.80	0.90	1.80
	Max	6.90	7.60	7.70	7.50
Munic. SR Coverage 2006	Mean	0.79	0.77	0.71	0.71
	Std. Dev.	0.25	0.22	0.23	0.24
	Min	0.20	0.20	0.20	0.10
	Max	3.50	3.50	3.50	3.50

Table 5: Descriptive statistics across BFP Exposure - Cohort 1996

		BFP Exposure Level			
Variable	Measure	0	1	2	3

Note: This table reports descriptive statistics for each variable of the liked data for the individuals that were born in 1996, whose family was registered in the SR between 2004 and 2006. The mean of the dummy variables (Female, Migrant, Formal Head, Co-Head, and Formal Co-Head) must be interpreted as the proportion of that category in the data. The number of municipality resident families are measured in thousands. Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

Table 5 shows some interesting features of the BFP. Considering the poverty alleviation focus of the program, many of these statistics should be expected. Even so, this table brings many new information for the public. The most important message of this table is that the BFP seems to have been focused on the most disadvantaged families. As a consequence, the program also focused on the less developed municipalities, which tend to have a higher concentration of these families.

The average individual Schooling at the age of 18 shows a substantial and monotonic increase from the non-exposed group (1.96) to the more exposed one (2.47). Considering the poverty focus of the program and that these are unconditional averages, this is a very surprising result.

On the other hand, the table shows that the higher the level of exposure to the BFP the lower the rate of formally employed and the lower the average real hourly wage on these employments.

Since the years of BFP Exposure are split into levels, in the variable Years of BFP Exposure it can be seeing the range and average of time of program exposure within each of the defined levels. As a consequence of the years observed, the maximum possible exposure during childhood for the cohort of 1996 is Level 3, which ranges from 7 to 8 years of benefits in this case.

The variable Female indicates that there are no substantial differences across exposure groups. The average indicates an expected result since it assumes value 1 if the individual

is a female and 0 otherwise. The dummy variable that indicates white individuals shows a consistent decreasing pattern as the exposure level increases, which is consistent with the country's history of slavery and discrimination.

The last variable concerning the individual shows initially a non-expected result, with a slight increase in the proportion of migrants from the exposure level zero to one, but afterward it decreases until the maximum exposure level, which is consistent with the known positive selection of migrants.

The average family characteristics of the more exposed individuals show remarkable differences to the non-exposed group. The formal labor market participation of the Head and the Co-Head decreases from the non-exposed group to the most exposed. Another expected result is the decrease in the parents' average wages in the formal sector as the BFP exposure level increases. The same is true for the parents' schooling level. The fact that the family income considered to the BFP eligibility is per capita may explain an important part of the higher average family size for the more exposed groups.

Considering variables of the environment in which these individuals probably grew up, the municipality characteristics across the exposure groups also show interesting patterns. The more exposed individuals tend to live in municipalities that in the year 2000 had a higher incidence of poverty, and higher demographic density. Moreover, the average municipality of residence of the more exposed also presented worse education quality in 2007.

The SR Coverage[34] in 2006 tend to be lower in the more exposed groups (2 and 3) than in the less exposed ones (0 and 1). Considering that in this sample there is only individuals that were registered in 2006, this may represent a potential competition for the benefits since the yearly budget maybe were not sufficient for all eligible families. In other words, given that an individual is registered, its chance to get the BFP benefit is higher if its municipality registered less eligible individuals.

To summarize the most important results of these descriptive statistics, we can state that

[34]Defined by the ratio between the number of registered individuals in 2006 and an estimate of poor individuals with the Demographic Census of 2000.

the BFP program was strongly focused on the poor families, and maybe as a consequence, focused on less developed municipalities.

Fortunately, for this work, the most important potential source of endogeneity pushes the results to being considered as lower bounds. The total income of the families at the moment of SR registration is not observed, but the Formal Labor Market income of the parents during the individuals' childhood suggests a negative relation between real income and program exposure.

4.2 Results on early adults Schooling and Labor Outcomes

In this section, we report separately the results obtained for the effects of BFP Exposure on the Schooling, the FLM participation, and on the Earnings outcomes. For the former two outcomes, the data applied is the cross-section derived from the panel data from 2002 to 2017. For the last outcome, the applied data comprises the years 2012 to 2017. It covers the first year in which each of the selected cohorts has their formal wages observed.

Although the treatment is a fixed childhood characteristic of these young adults, it is worth to take advantage of several years of labor market data, to avoid to introduce errors in the earnings measure. Since the composition of the cohorts by BFP exposure groups may not be proportionally represented over the years[35], macroeconomic fluctuations could be correlated to this unbalance, which justifies the inclusion of year and cohort fixed effects. Differences in the cost of living in different places are assumed to be controlled by the municipalities' initial conditions.

Each subsection begins reporting the table with the main results. Then, we report the estimates of heterogeneous effects for specific groups selected within the complete sample. Finally, we present the robustness checks, changing the rule of sample selection, and changing the identification strategy to the instrumental variable approach, for the cases of the outcomes of FLM participation and Earnings.

[35]For instance, the more exposed may postpone their labor market participation if they choose to stay more years studying.

4.2.1 Schooling

In this section, we report the results of the estimates of the effects of BFP Exposure during childhood on the Schooling level, through the Ordered Logit model. The dependent variable is given by the Schooling observed when the individual is 18 years old. we computed the Schooling Level from 0 to 5, following procedures described in the previous section. The results of the estimates of equation 1 are presented in Table 6.

Table 6: BFP Exposure effects on the Schooling Level

	Regression		
	(1)	(2)	(3)
BFP Exp 1	0.664***	0.642***	0.698***
	(0.034)	(0.034)	(0.035)
BFP Exp 2	0.836***	0.829***	0.913***
	(0.032)	(0.032)	(0.033)
BFP Exp 3	1.094***	0.975***	1.075***
	(0.031)	(0.032)	(0.033)
BFP Exp 4	1.182***	0.868***	0.997***
	(0.038)	(0.039)	(0.040)
Cohort 1997		0.370***	0.330***
		(0.014)	(0.014)
Cohort 1998		0.522***	0.469***
		(0.015)	(0.015)
Female		0.525***	0.535***
		(0.011)	(0.011)
White		0.085***	0.004
		(0.012)	(0.014)
Migrant		0.086***	0.036***
		(0.011)	(0.012)
Co-Head			0.118***
			(0.019)
Formal Co-Head			0.213***
			(0.054)
Formal Head			0.202***
			(0.055)
Head Avg Wage			−0.203***

Table 6: BFP Exposure effects on the Schooling Level

	Regression		
	(1)	(2)	(3)
			(0.031)
Co-head Avg Wage			-0.112^{***}
			(0.028)
Head Schooling			0.230^{***}
			(0.005)
Co-head Schooling			0.067^{***}
			(0.007)
Family Size			0.051^{***}
			(0.015)
Sq. Family Size			-0.009^{***}
			(0.001)
Mun. Density 2000			0.016^{***}
			(0.004)
Mun. Families 2000			-0.033^{***}
			(0.006)
Mun. % Poor Fam. 2000			0.497^{***}
			(0.026)
Mun. Educ. Quality			1.273^{***}
2007			
			(0.042)
Alpha 1	-4.0842^{***}	-3.6426^{***}	-2.1034^{***}
	(0.0451)	(0.0460)	(0.0901)
Alpha 2	-1.3513^{***}	-0.8983^{***}	0.6438^{***}
	(0.0306)	(0.0320)	(0.0835)
Alpha 3	0.8388^{***}	1.3367^{***}	2.9518^{***}
	(0.0303)	(0.0319)	(0.0838)
Alpha 4	3.5968^{***}	4.1457^{***}	5.8336
	(0.0322)	(0.0341)	(0.0852)
Alpha 5	7.8442^{***}	8.40^{***}	10.1068^{***}
	(0.0962)	(0.0968)	(0.1247)
Observations	118,285	118,221	116,876
Log-likelihood	$-136{,}011.4$	$-134{,}102.1$	$-129{,}939.4$

Statistical Significance: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Table 6: BFP Exposure effects on the Schooling Level

	Regression		
	(1)	(2)	(3)

Note: This table presents the Maximum Likelihood estimates of the coefficients of the Ordered Logit model for the Schooling level attained in early adulthood. The interest variables are the BFP Exposure level dummies. The BFP level of exposure zero is the reference category. From Column (1) to Column (2) the covariates of the individuals' characteristics are introduced. From Column (2) to Column (3) we add covariates with characteristics of family background, and characteristics of the residence municipality of the time of the individual's childhood. The Alpha coefficients represent the α's of equation (2) of the methodology section. Source: Author's calculations with the data from MDS, MTE, IBGE, and INEP.

Interestingly, the BFP Exposure coefficients do not change much from the regression without control variables (1) to the regression controlled for individual characteristics (2), and the regression with individual, family and municipality characteristics (3). The latter is the baseline model and shows that the coefficients of the BFP Exposure levels are positive, and increase from level of exposure 1 until level 3. The coefficient of the exposure level 4 is only smaller than level 3. This shows a potential maximum of the BFP effects on schooling occurring around 7 to 9 years of program exposure during childhood.

Regression (3) shows that things appear to be improving, as the younger cohorts have higher schooling levels than the older one, formed by those that were born in 1996. Females appear with higher schooling levels than males. Furthermore, the migrant positive selection, famous in the literature, is also present with the positive coefficients estimated in regressions (2) and (3).

The family background variables show another interesting set of results. The individual schooling seem to be higher if the Co-head is present in its family composition. The Co-head and Head schooling level also appear as positive influences on the children, while the latter seems to matter less than the former. Finally, to be born in a bigger family seems to be good to increase schooling attainment, but when it gets too big, the effect may be negative, as is shown by the coefficient of the variable with the square of the individual's family size.

External effects, captured by the municipality variables, are also shown to be important. As expected, the municipality Index of Education Quality of 2007 has a positive influence on the individual schooling attained at the age of 18 years old. Moreover, to be a resident of a municipality with a higher population density and smaller population, has also positive effects. Surprisingly, individuals from municipalities that had higher proportion of poverty in 2000 have a higher probability of attaining higher schooling levels.

Considering the baseline model, Regression (3) with the full set of control variables, we predicted the probability density of BFP Exposure within each Schooling level, through Equation 5. We set the control variables to the values reported in Table 19 in the Appendix. The values chosen for the continuous variables were simply the sample means. For the discrete variables, we made arbitrary choices. However, the direction of the results presented remains unchanged whichever values one defines, and the magnitude changes only slightly in the neighborhood of them.

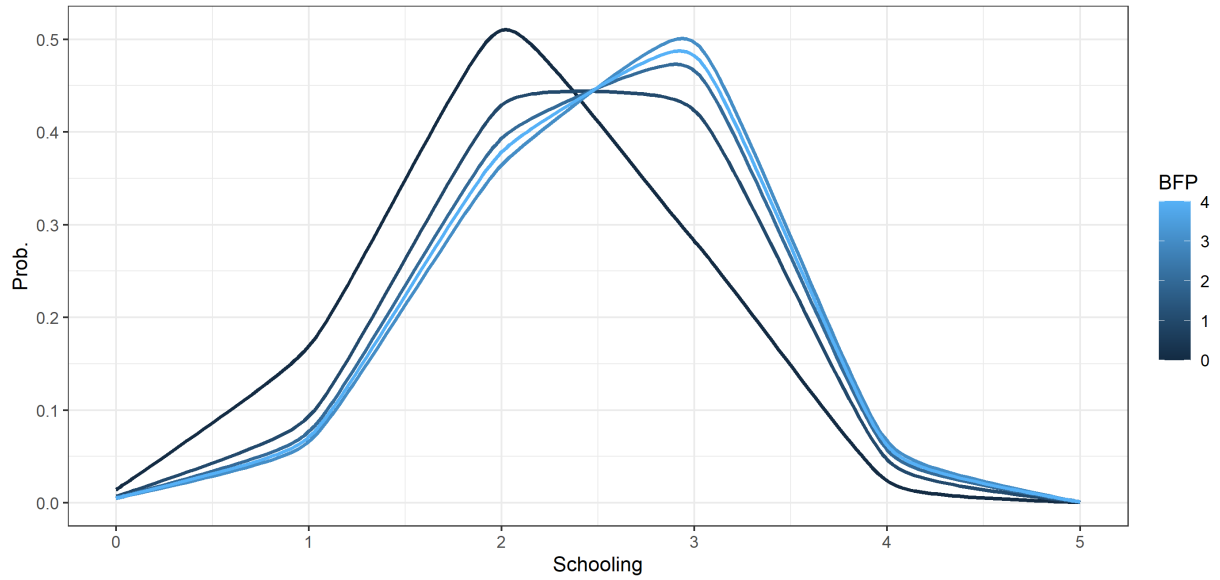
The resulting probabilities are reported in Table 20 of the Appendix, but Figure 1 summarises it. It shows the right shift in Schooling distribution for higher levels of exposure. Another important remark is that tertiary education studies (level 5) is relatively rare for the low-income population, which concentrates probability on medium and lower schooling levels (2 and 3).

The next results reported concern some heterogeneous effects investigation and robustness tests, to compare with the baseline specification, the regression 3 of Table 6. First, Table 7 reports the regressions for distinct samples, considering some specific groups of individuals by demographics, family background, or municipality characteristics[36].

Regression (1) of Table 7 considers only the women of the sample. This sample allows to check if, for example, the fertility decision can produce different effects of the BFP on the Schooling. On the other hand, regression (2) considers only men. Comparing both samples' results, we can state that the effects of the program exposure on the long-term schooling

[36]The predicted probability considering a hypothetical average individual for each heterogeneous sample is presented in Tables 21, 22, 23, 24, 25, and 26 of the Appendix.

Figure 1: Predicted Distribution of the Schooling within BFP Exposure Level



Note: The figure represents the distribution of the Schooling level within each BFP Exposure Level. Higher levels of exposure are represented in lighter shades. Although the Schooling levels are discrete, we chose to represent the results in continuous curves for easier visualization.

level are stronger for males than females. Nonetheless, both present positive and increasing coefficients until the BFP exposure level 3.

Table 7: Heterogeneous Effects on Schooling Level at 18 y.o.

	Regression				
	(1)	(2)	(3)	(4)	(5)
BFP Exp 1	0.594***	0.795***	0.544***	0.743***	0.288***
	(0.050)	(0.048)	(0.088)	(0.043)	(0.081)
BFP Exp 2	0.823***	0.995***	0.685***	0.971***	0.480***
	(0.047)	(0.046)	(0.084)	(0.040)	(0.079)
BFP Exp 3	0.970***	1.174***	0.891***	1.098***	0.734***
	(0.047)	(0.046)	(0.085)	(0.040)	(0.078)
BFP Exp 4	0.936***	1.057***	0.690***	1.070***	0.528***
	(0.059)	(0.056)	(0.118)	(0.050)	(0.095)
Observations	55,649	61,227	10,662	78,238	20,358

Statistical Significance: *p<0.1; **p<0.05; ***p<0.01

Table 7: Heterogeneous Effects on Schooling Level at 18 y.o.

	Regression				
	(1)	(2)	(3)	(4)	(5)

Note: This table presents the Maximum Likelihood estimates of the Ordered Logit mode for the Schooling level attained by young adults, with the interest dummy variables representing the levels of BFP Exposure, for different samples. The sample of Column (1) considers only the girls, and Column (2) only the boys. The sample of the regression of Column (3) selects only the children of parents with at least one formal employment during individuals childhood. Column (4) considers individuals from small cities (<200,000 families in 2000), while Column (5) considers individuals from large cities (>200,000 families in 2000). Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

Regression (3) selects only individuals whose parents appeared at least once in the Formal Labor Market during their childhood. It is interesting to note that coefficients have a slightly smaller magnitude at all levels. This result makes sense if one thinks that the program may lose importance the better the family economic conditions are. Although many poor families of the country are excluded, with this sample we can control much better for income selection to the program since we observe the average wage of all parents during individuals childhood.

The regressions (4) and (5) consider samples of families that live, respectively, in small and large municipalities. This is an important heterogeneity supported by Urban and Regional Economics literature since the dynamics, quality of schools, and opportunities in the labor market may considerably vary with the city size. We defined small municipalities as those with less than 50,000 resident families in the Demographic Census of 2000. On the other hand, large municipalities are defined by those with more than 200,000 resident families.

From the sample with small to the sample with large municipalities we observe a sharp drop in the coefficients of all the BFP exposure levels. However, the positive and increasing effects (until level 3) remains unchanged. This result may suggest that larger cities, which are denser and richer on average, can potentially provide more public schools, with better quality and closer to each family, so that the BFP makes less difference on the school attendance opportunity costs.

In Table 8 we turn to robustness analyses and report the results of the estimates for

alternative approaches. We consider a different rule to define the sample of interest and apply alternative measures of the interest variables of program exposure.

Table 8: Robustness tests of the Effects on Schooling Level

	Regression		
	(1)	(2)	(3)
Years of BFP	0.209***		
	(0.009)		
Sq. Years of BFP	−0.012***		
	(0.001)		
BFP Exp 1		0.642***	0.316***
		(0.063)	(0.064)
BFP Exp 2		0.665***	0.353***
		(0.061)	(0.061)
BFP Exp 3		1.097***	0.416***
		(0.058)	(0.060)
BFP Exp 4		0.930***	0.389***
		(0.067)	(0.067)
Observations	116,876	41,477	66,924

Statistical Significance: *p<0.1; **p<0.05; ***p<0.01
Note: This table presents the coefficients of the robustness tests, estimating the Ordered Logit changing the interest variable definition to continuous, and introducing its squared term. In Column (2) we select the individuals from families registered in 2004. Column (3) estimates use only the individuals that still receive the BFP, whether from its initial family or on their own new family. Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

Regression (1) considers alternative measures of BFP Exposure, changing from the defined levels to the Years of Exposure and its squared term. In this alternative interest variable definition, we observe both features of the baseline specification. We have a positive effect and increasing until 8.7 years of exposure, when the parabola achieves its maximum. The effect would be positive until 17.4 years of exposure, which is completely out of the sample, which has a maximum possible exposure of 10 years during childhood.

The regression 2 throws away an important part of the BFP Exposure variation, by

considering only individuals of families registered in 2004. This procedure guarantee that all families got registered in the SR of the municipality in the same year. Thus, the remaining variation conditioned on all control variables occurs from one municipality to the other. From those municipalities with the same poverty level, in the ones that more families were registered, the chance of each family to be granted with the BFP was lower. This argument is supported by the program rule that prioritizes municipalities with lower current benefit coverage.

Regression (3) of Table 8 considers a potential source of bias in the estimates of the baseline model. If the more exposed families tend to update more the SR information, the schooling level of the more exposed would naturally tend to be higher since the Schooling does not reduce over time. This threat occurs because updating the SR information each two years is a condition to keep the family receiving the benefits. However, the positive and increasing effects of BFP exposure until level 3 remain, but the magnitude of the coefficients suffers a substantial reduction.

The baseline, heterogeneity, and robustness results altogether lead us to the same conclusion. Some groups or specifications may point to stronger of softer results, but all point to positive effects of the BFP time of exposure during childhood on the observed early adulthood schooling level.

4.2.2 Formal Labor Market Participation

This section presents the results for the first Labor Market outcome, the probability of participating in the Formal Labor Market. We present the baseline specification results and perform the prediction of the variation in the probability with the program exposure measure for a representative individual. Then, we proceed to the investigation of heterogeneous effects and present some robustness checks, including the alternative identification strategy through an instrumental variable approach.

In Table 9 we report the estimated coefficients and standard errors, of the Logit regres-

sions given by equation 6 of the methodology section. Each column represents a different regression, with a different set of control variables.

Table 9: Formal Labor Market Participation - Logit

	Regression		
	(1)	(2)	(3)
BFP Exp. 1	0.053*	0.076***	0.078**
	(0.028)	(0.029)	(0.031)
BFP Exp. 2	−0.049*	−0.019	0.140***
	(0.026)	(0.027)	(0.029)
BFP Exp. 3	−0.345***	−0.156***	0.175***
	(0.026)	(0.027)	(0.029)
BFP Exp 4	−0.826***	−0.189***	0.183***
	(0.037)	(0.039)	(0.041)
Cohort 1997		−0.391***	−0.491***
		(0.014)	(0.014)
Cohort 1998		−0.913***	−1.084***
		(0.015)	(0.016)
Female		−0.304***	−0.342***
		(0.011)	(0.012)
White		0.593***	0.138***
		(0.012)	(0.014)
Migrant		0.015	0.079***
		(0.012)	(0.013)
Co-head			−0.135***
			(0.022)
Formal Co-head			0.182***
			(0.055)
Formal Head			0.438***
			(0.053)
Head Avg Wage			−0.065**
			(0.030)
Co-head Avg Wage			0.063**
			(0.027)
Head Schooling			0.094***
			(0.006)

Table 9: Formal Labor Market Participation - Logit

	Regression		
	(1)	(2)	(3)
Co-head Schooling			0.033***
			(0.008)
Family Size			0.057***
			(0.018)
Sq. Family Size			−0.013***
			(0.002)
Mun. Density 2000			0.029***
			(0.004)
Mun. Pop. 2000			−0.062***
			(0.006)
Mun. % Poor Fam.			−0.736***
2000			
			(0.027)
Mun. Educ. Quality			1.306***
2007			
			(0.048)
Intercept	−0.511***	−0.281***	−2.480***
	(0.024)	(0.027)	(0.093)
Observations	149,023	148,934	145,273
Log-likelihood	−93,743.46	−90,360.64	−82,436.49

Statistical Significance: *p<0.1; **p<0.05; ***p<0.01
Note: This table reports the Maximum Likelihood estimates of the Logit model for the Formal Labor Market Participation probability. The interest variables are the BFP Exposure level dummies. The BFP level of exposure zero is the reference category. From Column (1) to Column (2) the covariates with the individuals' characteristics are introduced. From Column (2) to Column (3) we add covariates with characteristics of family background, and characteristics of the residence municipality of the time of the individual's childhood. Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

Contrasting with the results for the schooling outcome, controlling the regressions for the covariates seem to play an important role in the Formal Labor Market Participation. From regression (1) to regression (3) all the coefficients of the dummy variables for the program

exposure levels increase. The baseline model, regression (3), shows the increase at decreasing rate effect of the program exposure on FLM Participation. Interestingly, this non-linearity was also present in the results for the Schooling outcome.

The individual characteristics show that the youngest tend to be less present in the formal labor market. The same is true considering female, non-white, and non-migrant individuals. Considering the family background variables, we observe a positive influence of having a Co-head in the family, as long as it participated in the formal labor market. The schooling of the Head and Co-head are also positive influences on the individual FLM participation. Moreover, the same effect present in the schooling results appears here, with the family size being a positive influence but at a decreasing rate.

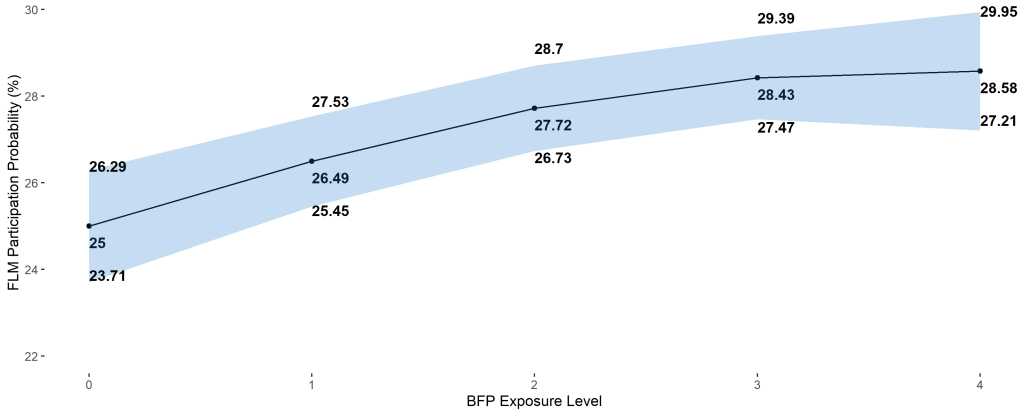
Individuals that reside in municipalities with higher population density and a smaller population in the year 2000 have a higher probability of participating in the FLM. Municipalities with a smaller proportion of poor families in 2000 and higher Index of Education Quality in 2007 also affect positively the probability of its residents getting a formal job.

Assuming the coefficients of regression (3) reported in Table 9 we computed the conditional probability of an individual, with a given set of observed characteristics, to appear in the Formal Labor Market at least once until 2017. The formula of this probabilities is given by Equation 7.

For simplicity purposes, we applied the same values for the control variables when predicting the Formal Labor Market participation probability across the BFP Exposure levels that we used in the Schooling outcome section. These values are presented in Table 19 of the Appendix.

The Figure 2 shows that an individual with the given set of characteristics is expected to have a probability of 28.6% to get a formal job if it was exposed to the program during 10 years, i.e., 3.6 percentage points higher than if it was not exposed to the BFP during its childhood (25%). The effect increases in a decreasing rate, as was also the case in the Schooling outcome.

Figure 2: Predicted Probability of Participation in the Formal Labor Market by BFP Exposure



This is a remarkable result and goes against many thoughts and comments made by the public common sense. With the formal sector being more productive than the informal one, this constitutes a channel in addition to the schooling increase to foster productivity and development for the most disadvantaged groups of Brazilian society.

The next groups of results concentrate on possible heterogeneous effects[37] concerning individual, family and municipality characteristics, and robustness analyses, changing the BFP exposure measure, the sample selection rule, and the econometric identification strategy.

Table 10: BFP Exposure effects on Formal Labor Market Participation
Heterogeneous Effects

	Regression				
	(1)	(2)	(3)	(4)	(5)
BFP Exp. 1	0.032	0.123***	−0.017	0.109***	−0.064
	(0.044)	(0.042)	(0.072)	(0.039)	(0.073)
BFP Exp. 2	0.080*	0.199***	0.075	0.174***	0.036
	(0.042)	(0.040)	(0.069)	(0.037)	(0.071)
BFP Exp. 3	0.116***	0.232***	0.050	0.204***	−0.0003
	(0.042)	(0.039)	(0.070)	(0.036)	(0.070)
BFP Exp 4	0.122**	0.242***	0.018	0.235***	−0.031
	(0.061)	(0.055)	(0.111)	(0.053)	(0.092)
Observations	69,385	75,888	14,805	95,093	26,044

[37]The predicted probability considering a hypothetical average individual for each heterogeneous sample of Table 10 is presented in Table 27 of the Appendix.

Table 10: BFP Exposure effects on Formal Labor Market Participation
Heterogeneous Effects

	Regression				
	(1)	(2)	(3)	(4)	(5)

Statistical Significance: *p<0.1; **p<0.05; ***p<0.01
Note: This table presents the coefficients of the BFP Exposure levels for maximum likelihood estimates of Formal Labor Market participation probability, with the full set of control variables, selecting subsamples. Regression (1) considers only the girls, Regression (2) only the boys. Regression (3) selects only the children of parents with at least one formal employment during individuals childhood. Regression (4) considers individuals from small cities (<200,000 families in 2000), while Regression (5) considers individuals from large cities (>200,000 families in 2000). Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

Regression (1) selects only Females. This is justified to investigate potential differential effects of the fertility decision, and the potential effects of discrimination in the Labor Market. With the same justification, in Regression (2) are reported the results considering only the males of the sample. Comparing both, the results for males are stronger than for females.

Regression (3) selects individuals whose parents (Head and Co-Head) appeared at least once in the formal labor market during their childhood. In this sample, we control much better the program selection since we observe the formal income of these parents. The drawback is that we select only the individuals with a better family background, for which the program is shown to have no effect.

The last two regressions of the heterogeneous effects concern the size of the residence municipality of each individual. The sample used in regression (4) selects only Small Municipalities, arbitrarily defined by those with less than 50,000 families in the Demographic Census of 2000. On the other hand, Regression (5) selects only big municipalities, defined by those with more than 200,000 resident families in the Demographic Census of 2000.

These are important sources of heterogeneous effects and are supported by Urban and Regional Economics literature. Since the dynamics, quality of schools, and opportunities in the labor market may vary considerably with the city size, the effects of the BFP on the individual Formal Labor Market participation may also vary with that.

In the next set of results of this section, we concentrate the robustness analyses in three additional regressions, which results are reported in Table 11.

Table 11: BFP Exposure effects on Formal Labor Market Participation
Robustness Tests

	Regression		
	(1)	(2)	(3)
Years of BFP	0.025***	0.7234***	
	(0.009)	(0.008)	
Squared Years of BFP	−0.001	−0.0778***	
	(0.001)	(0.008)	
BFP Exp. 1			0.058
			(0.055)
BFP Exp. 2			0.080
			(0.053)
BFP Exp. 3			0.156***
			(0.050)
BFP Exp 4			0.127**
			(0.061)
Observations	145,273	145,273	51,706

Statistical Significance: *p<0.1; **p<0.05; ***p<0.01
Note: This table presents the coefficients of robustness tests, always considering the full set of control variables. Column (1) reports the coefficients of the Logit model for the Formal Labor Market Participation probability, considering the Years of Exposure and its squared term as the interest variables. Column (2) presents the Instrumental Variable Probit estimates considering both interest variables as endogenous and instrumenting them by the logarithm of municipality SR Coverage in 2006, and its respective squared term. Column (3) presents the coefficients estimated considering the sample of individuals from families registered in the SR in 2004. Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

In Regression (1) are reported the coefficients of the Logit estimated with the interest variable measure redefined to the Years of Exposure instead of levels. To consider the apparent non-linear effects we incorporate the quadratic term as well. We observe that the positive effect of the BFP exposure on the Formal Labor Market participation probability remains. However, the quadratic term appears as non-significant in this specification.

Regression (2) reports the Instrumental Variable Probit, considering the municipality effort to get the families registered in the SR and the respective squared term as instruments for both the Years of BFP and Squared Years of BFP exposure. In this specification, the positive and decreasing effect of the program exposure remain. The coefficients suggest that the maximum occurs around 4.65 years of BFP Exposure, while the effect remains positive until the second root of the quadratic equation, which occurs when Years of Exposure is equal to 9.3. Considering that the maximum possible exposure in the sample is 10 years, it is reasonable to state that this specification suggests that the program had positive effects on the outcome for almost the entire range of years of exposure.

The estimates of the first stage of the Instrumental Variable method are reported in Table 12. Considering that in the sample we have only the registered individuals whose family was registered until 2006, the coefficients are negative as expected, considering the competition effects. In other words, the higher the municipality registration of other families the lower the expected exposure of the already registered individuals of this municipality.

Table 12: BFP Exposure effects on Formal Labor Market Participation
First Stage Coefficients

	Endogenous Variable	
Instrument	BFP Years of Exposure	Sq. Years of Exposure
SR Coverage 2006	-1.2269^{***}	-9.0980^{***}
	(0.0412)	(0.4400)
Squared SR Coverage 2006	0.3960^{***}	6.7398^{***}
	(0.0383)	(0.4090)

Statistical Significance: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$
Note: This table presents the coefficients of the endogenous variables of BFP Exposure and its squared term, and their instruments, the logarithm of the observed SR Coverage in 2006 in the municipality, and its squared term. Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

In the last robustness test, Regression (3), only individuals of families registered in 2004 are considered. Although an important part of the variation in the program exposure mea-

sure is lost, this guarantees that all the families achieved or were achieved by the BFP manager of the municipality in the same year. This means that, given individual and family characteristics, the exposure varies from one municipality to the other. From those municipalities with the same poverty level, in the ones that more families were registered, the chance of each family to be granted with the BFP is lower. This logic makes sense considering the program's rule of balancing the BFP coverage between municipalities. This is done by prioritizing the municipalities with lower benefit coverage at the moment of releasing decision.

Note that, for the FLM participation and Earnings outcome, there is not much sense in making the last robustness test[38] that was done for Schooling. Formal Labor Market Participation and Earnings are not self-reported information, and do not suffer from the possibility of being outdated, as is the case for the SR information. These outcomes are usually reported by the human resources department of each company.

4.2.3 Earnings in the Formal Labor Market

Now we turn to the last long-term outcome of interest of this work, the Earnings that these young adults get in their formal jobs. We measure the earnings as the yearly average of their nominal wage per hour. In all specifications, we consider the wages on the logarithmic scale, as usual in the literature.

Table 13: BFP exposure effects on Earnings

	Regressions		
	(1)	(2)	(3)
BFP Exp. 1	-0.018^{***}	-0.016^{***}	-0.013^{***}
	(0.004)	(0.004)	(0.004)
BFP Exp. 2	-0.022^{***}	-0.019^{***}	-0.013^{***}
	(0.003)	(0.003)	(0.004)
BFP Exp. 3	-0.040^{***}	-0.028^{***}	-0.012^{***}
	(0.003)	(0.003)	(0.004)

[38]The test of Regression (3) of Table 8 that considers the sample with only individuals that are still BFP beneficiaries.

Table 13: BFP exposure effects on Earnings

	Regressions		
	(1)	(2)	(3)
BFP Exp. 4	-0.085^{***}	-0.032^{***}	-0.015^{**}
	(0.006)	(0.006)	(0.006)
Cohort 1997		-0.034^{***}	-0.039^{***}
		(0.002)	(0.002)
Cohort 1998		-0.069^{***}	-0.079^{***}
		(0.002)	(0.002)
Female		-0.021^{***}	-0.025^{***}
		(0.002)	(0.002)
White		0.054^{***}	0.025^{***}
		(0.002)	(0.002)
Migrant		0.012^{***}	0.016^{***}
		(0.002)	(0.002)
Co-Head			0.012^{***}
			(0.003)
Formal Co-Head			-0.119^{***}
			(0.007)
Formal Head			-0.097^{***}
			(0.007)
Avg Wage Head			0.054^{***}
			(0.004)
Avg Wage Co-Head			0.056^{***}
			(0.004)
Head Schooling			0.001
			(0.001)
Co-head Schooling			-0.003^{***}
			(0.001)
Fam. Size			-0.003
			(0.003)
Sq. Fam. Size			0.0002
			(0.0002)
Mun. Density 2000			0.001^{**}
			(0.001)
Mun. Pop 2000			-0.011^{***}

Table 13: BFP exposure effects on Earnings

	Regressions		
	(1)	(2)	(3)
			(0.001)
Mun. % Poor Fam. 2000			-0.111^{***}
			(0.003)
Mun. Educ. Quality 2007			-0.006
			(0.007)
Intercept	1.493^{***}	1.466^{***}	1.467^{***}
	(0.007)	(0.007)	(0.014)
Observations	116,685	116,604	113,162
Adjusted R^2	0.219	0.233	0.251

Statistical Significance: *p<0.1; **p<0.05; ***p<0.01
Note: This table reports the estimates of Ordinary Least Squares coefficients of the log-earnings regressions on BFP Exposure level dummy variables and covariates. The BFP level of exposure zero is the reference category. From Column (1) to Column (2) the covariates of the individuals' characteristics are introduced. From Column (2) to Column (3) we add covariates with characteristics of family background, and characteristics of the residence municipality of the time of the individual's childhood. Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

In Table 13 the presented estimates consider a different set of covariates, which are incorporated in the model by groups. In the first column, is reported the regression estimates with no control variables. In the second column, Regression (2), the individual's control variables are added. In the last column, Regression (3), are introduced the control variables of family background and municipality characteristics.

Table 14 presents the results of the inquiry on some heterogeneous effects. On each column, there is a different sample, selected based on the individual's demographics or the characteristics of its municipality.

Table 14: BFP Exposure effects on Earnings - Heterogeneous Effects

	Regressions				
	(1)	(2)	(3)	(4)	(5)
BFP Exp. 1	−0.017***	−0.010*	−0.014*	−0.014***	−0.026***
	(0.005)	(0.005)	(0.008)	(0.005)	(0.009)
BFP Exp. 2	−0.019***	−0.008*	−0.012	−0.017***	−0.005
	(0.005)	(0.005)	(0.008)	(0.005)	(0.008)
BFP Exp. 3	−0.017***	−0.008*	−0.010	−0.015***	−0.007
	(0.005)	(0.005)	(0.008)	(0.005)	(0.008)
BFP Exp. 4	−0.022**	−0.009	−0.042***	−0.014*	−0.012
	(0.009)	(0.008)	(0.015)	(0.008)	(0.014)
Observations	48,548	64,614	17,061	64,879	23,363
Adjusted R^2	0.236	0.263	0.284	0.246	0.242

Statistical Significance: *p<0.1; **p<0.05; ***p<0.01
Note: This table presents the coefficients of the BFP Exposure levels for ordinary least square regressions of log-earnings, with the full set of control variables, selecting sub-samples. Regression (1) considers only the girls, Regression (2) only the boys. Regression (3) selects only the children of parents with at least one formal employment during individuals childhood. Regression (4) considers individuals from small cities (<200,000 families in 2000), while Regression (5) considers individuals from large cities (>200,000 families in 2000). Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

The first column of coefficients represents the results for the sample restricted to Females. On the other hand, Regression (2) concerns the sample restricted to men. As commented in the other outcomes, this distinction is important if it is considered that fertility decisions may affect more the females. It is observed that although with a small magnitude, the negative effects for females are stronger and significant for all exposure levels. As the dependent variable is in the logarithmic scale, the interpretation is that a female exposed to the BFP during her childhood is expected to earn 2.2% less than a female in the control group.

Regression (3) considers only individuals from families in which both the Head and Co-head appear in the Formal Labor Market at least once during the individual childhood. This test shows us differences in the effects when we consider individuals from families with a better background. Moreover, with this sample, we control much better for the program income selection problem.

Regressions (4) and (5) consider potential heterogeneities between different municipality's population sizes. This investigation is motivated by the results from traditional Regional and Urban Economics literature, which states the existence of agglomeration economies. The dynamics and quality of schools, and opportunities in the labor market, may vary considerably with the urban population size.

The last set of results of the section comprises the robustness checks on the previous specifications. In Table 15 we present the results obtained when we change the BFP Exposure measure and the identification strategy to the Instrumental Variable approach.

Table 15: BFP Exposure effects on Earnings - Robustness Tests

	Regressions		
	(1)	(2)	(3)
Years of BFP	-0.002^{*}	0.002	
	(0.001)	(0.018)	
Squared Years of BFP	0.0002	0.0003	
	(0.0001)	(0.0019)	
BFP Exp. 1			0.001
			(0.006)
BFP Exp. 2			-0.001
			(0.006)
BFP Exp. 3			0.005
			(0.006)
BFP Exp. 4			0.008
			(0.008)
Observations	113,162	113,162	43,278
Adjusted R^2	0.251	0.249	0.255

Statistical Significance: *p<0.1; **p<0.05; ***p<0.01

Table 15: BFP Exposure effects on Earnings - Robustness Tests

	Regressions		
	(1)	(2)	(3)

Note: This table presents the coefficients of robustness tests, always considering the full set of control variables. Column (1) reports the coefficients of Ordinary Least Squares of log-earnings, considering the Years of Exposure and its squared term as the interest variables. Column (2) presents the Two-Stage Least Squares estimates considering both interest variables as endogenous and instrumenting them by the logarithm of municipality SR Coverage in 2006, and its respective squared term. Column (3) presents the coefficients estimated considering the sample of individuals from families registered in the SR in 2004. Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

In Regression (1) we assume different measures of Exposure. The time of BFP Exposure (in years) and the respective squared term.

In Regression (2) we apply an Instrumental Variable approach, considering the municipality effort to register vulnerable families in the SR (in logarithmic scale), and the respective squared term. The squared term is justified by all the results of other specifications, which suggest non-linear effects of the BFP duration of exposure.

For the Instrumental Variable approach, we had to use the years of exposure because of the existence of only two instruments. The municipality effort to include families in the Single Register is represented by the SR Coverage observed in 2006, i.e., the number of families registered in 2006 divided by the number of poor families in the municipality, estimated with the 2000 Demographic Census data.

The first stage coefficients are reported in Table 16.

Table 16: Wages - First Stage Coefficients

	Endogenous Variable	
Instrument	BFP Years of Exposure	Sq. BFP Years of Exp
SR Coverage 2006	-1.3516^{***}	-10.21^{***}
	(0.052)	(0.46)
Squared SR Coverage 2006	0.3464^{***}	5.59^{***}
	(0.0442)	(0.4157)

Table 16: Wages - First Stage Coefficients

	Endogenous Variable	
Instrument	BFP Years of Exposure	Sq. BFP Years of Exp

Statistical Significance: *p<0.1; **p<0.05; ***p<0.01
Note: This table presents the first stage coefficients of the Two-Stage Least Squares method applied to the Earnings equation, whose coefficients are reported in Table 15. Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

The results of the first stage are in line with the expected. For each family, the more its municipality registers other families the lower its chance to be benefited, since the sample is composed only by registered families.

Regression (3) of Table 15 comprises only individuals of families registered in 2004. A lot of variation is lost, but this guarantees that each family was achieved by the municipality in the same year. The release of benefits will vary from one municipality to the other. From those municipalities with the same poverty level, in the ones that more families were registered, the chance of each family to be granted with the BFP is lower due to the program rule of prioritizing the municipalities with lower current benefit coverage.

5 Final Remarks

This paper presented the first long-term effects estimates of the largest Conditional Cash Transfer program in the world, the Bolsa Família Program (BFP). We focused on the effects of BFP exposure during individuals' childhood and adolescence on the Schooling level attained in early adulthood, the probability to participate in the Formal Labor Market (FLM), and the Earnings (Wages) obtained on these formal employments.

To enable these estimates we linked identified data from the Single Registry for Social Programs (SR), with the BFP payment registers, and the administrative records of the FLM, which resulted in a dataset with more than one-quarter of the Brazilian families, and practically the entire poor population of the country. Because of the work purposes and the timing of the program, we focused the analysis on the individuals born between 1996 to

1998, which represent 7.6 million people.

The program implementation was not a randomized trial, so we had to elaborate a strategy to select the proper sample and construct a Natural Experiment. The selected individuals individuals have enough age to be a child during the first program payment year in which the data is available (2004). In addition, they are likely under the legal liability of a family head in the first year of the Single Registry data, i.e., they are less than 18 years old in 2012. To avoid families that were benefited in previous years in the sample, we focused on families registered in the SR from 2004 to 2006. This was a period of strong expansion when roughly one-quarter of the families observed were identified.

Summarizing the main identification strategy, we rely on the inclusion of many observable characteristics of the individual and its family as control variables, and on the rules of release of the benefits that assign priority to families from municipalities with lower current BFP coverage. This means that the same family in another municipality could be selected for the program just because the other municipality covered fewer families in the SR at a given year. Moreover, the yearly budget reserved for the program can also leave a family that was incorporated later in the SR in a waiting list.

The three outcomes of interest are variables of different mathematical natures and therefore demanded different econometric models. The Schooling measure was constructed from primary data in the SR, and divided into six different and increasing levels. The econometric approach for this outcome is the Ordered Logit. The Formal Labor Market (FLM) participation has a binary nature, and for that reason, we applied one of the most traditional probabilistic models, the Logit. Both Schooling and FLM participation equations were estimated through traditional Maximum Likelihood methods. For the last outcome, Earnings, we estimated a log-linear model through Ordinary Least Squares. Concerning the alternative identification strategy, we applied the Instrumental Variable Probit for the FLM participation and the method of Two-Stage Least Squares for the Wages.

The descriptive statistics showed that the program had a strong focus on the most vul-

nerable families and the poorest municipalities. These numbers reinforced the need to add these variables in the estimates of the baseline econometric specifications.

The major results can be summarized as follows: the higher the time of exposure to the BFP during childhood and adolescence, the higher the level of Schooling the individual present when it is 18 years old, and the higher the probability of participating in the Formal Labor Market. Results for wages vary from slightly negative to non-significant when we include more control variables or change the identification strategy to the instrumental variable approach.

Selecting the sample groups to test for heterogeneous effects showed that the stronger schooling and formal labor market participation effects occurred for males, for children of never formally employed parents, and for smaller cities. This last result is consistent with the theories of Urban Economics since bigger urban centers may have strong economic dynamics and the relative importance of the CCTs may be much smaller. A similar argument can be constructed for children of formally employed parents. These children potentially have a much better family background, and their parents would invest in them regardless of the BFP conditionalities and benefits.

One of the robustness tests was to assume an alternative identification strategy. In this approach, we took advantage of the exogeneity, from the family perspective, of the municipality effort to get vulnerable families into the Single Registry, which is a necessary condition to have the benefit released. This instrumental variable approach showed the same positive effects for the probability of Formal Labor Market participation, while presented not statistically significant results for the wage equation.

In general, the results showed that the BFP had also the desired long-term role of improving the second generation future in Brazil. Although we did not find positive effects for Earnings, it is remarkable that the more exposed individuals from poor families achieved higher schooling level and are more prone to get a formal job, which is related to better quality and higher productivity activities.

Immediate and challenging extensions of this topic would be to investigate what happened decades later with the parents of these children, and how the local economies were affected in the long-term by the exposure to this program.

References

Azzoni, C. R., J. J. Guilhoto, E. A. Haddad, G. J. Hewings, M. A. Laes, and G. R. Moreira (2009). Social policies, personal and regional income inequality in brazil: an io analysis. In *Brazil under Lula*, pp. 243–261. Springer. 1

Barham, T., K. Macours, and J. A. Maluccio (2013). Boys' cognitive skill formation and physical growth: Long-term experimental evidence on critical ages for early childhood interventions. *American Economic Review: Papers and Proceedings 103*(3), 467–71. 7

Barham, T., K. Macours, and J. A. Maluccio (2017). Are conditional cash transfers fulfilling their promise? schooling, learning, and earnings after 10 years. *CEPR Discussion Paper No. DP11937*. 7

Barrera-Osorio, F., L. L. Linden, and J. E. Saavedra (2019). Medium-and long-term educational consequences of alternative conditional cash transfer designs: Experimental evidence from colombia. *American Economic Journal: Applied Economics 11*(3), 54–91. 7

Bastagli, F., J. Hagen-Zanker, L. Harman, V. Barca, G. Sturge, T. Schmidt, and L. Pellerano (2016). Cash transfers: what does the evidence say. *A rigorous review of programme impact and the role of design and implementation features. London: ODI*. 1, 7

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of political economy 70*(5, Part 2), 9–49. 2

Behrman, J. R. and S. W. Parker (2013). Is health of the aging improved by conditional cash transfer programs? evidence from mexico. *Demography 50*(4), 1363–1386. 6

Behrman, J. R., S. W. Parker, and P. E. Todd (2011). Do conditional cash transfers for schooling generate lasting benefits? a five-year followup of progresa/oportunidades. *Journal of Human Resources 46*(1), 93–122. 7

Cameron, A. C. and P. K. Trivedi (2009). *Microeconometrics using stata*, Volume 5. Stata press College Station, TX. 15

Chitolina, L., M. N. Foguel, and N. A. Menezes-Filho (2016). The impact of the expansion of the bolsa familia program on the time allocation of youths and their parents. *Revista Brasileira de Economia 70*(2), 183–202. 6, 7

Cunha, F. and J. Heckman (2007). The technology of skill formation. *American Economic Review 97*(2), 31–47. 2

De Brauw, A., D. O. Gilligan, J. Hoddinott, and S. Roy (2015). Bolsa família and household labor supply. *Economic Development and Cultural Change 63*(3), 423–457. 6

De Oliveira, G. L. (2020). Long-term effects of conditional cash transfers on children: The brazilian case. *Ph.D Dissertation - University of São Paulo*, 1–107. 18

De Oliveira, G. L. and A. L. S. Chagas (2020). Effects of a cash transfer programme on origin–destination migration flows. *Regional Science Policy & Practice 12*(1), 83–104. 1, 6

Foguel, M. N. and R. P. d. Barros (2010). The effects of conditional cash transfer programmes on adult labour supply: an empirical analysis using a time-series-cross-section sample of brazilian municipalities. *Estudos Econômicos (São Paulo) 40*(2), 259–293. 6

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education 1*, 307–458. 15

MC. Ministry of citizenship. *available at ¡ http://www. mds. gov. br/bolsafamilia¿, Retrieved on March 11, 2020*. 3

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological) 42*(2), 109–127. 10

Meghir, C., R. Narita, and J.-M. Robin (2015). Wages and informality in developing countries. *American Economic Review 105*(4), 1509–46. 2

Millán, T. M., T. Barham, K. Macours, J. A. Maluccio, and M. Stampini (2019). Long-term impacts of conditional cash transfers: Review of the evidence. *The World Bank Research Observer 34*(1), 119–159. 1, 7

Millán, T. M., K. Macours, J. A. Maluccio, and L. Tejerina (2020). Experimental long-term effects of early-childhood and school-age exposure to a conditional cash transfer program. *Journal of Development Economics 143*, 102385. 7

Parker, S. and E. Skoufias (2000). The impact of progresa on work, leisure and time allocation. *Washington, DC: International Food Policy Research Institute*. 6

Parker, S. W. and T. Vogl (2018). Do conditional cash transfers improve economic outcomes in the next generation? evidence from mexico. *NBER Working Paper No. w24303*. 7

Santos, D. B., A. R. Leichsenring, N. Menezes Filho, and W. Mendes-Da-Silva (2017). Os efeitos do programa bolsa família sobre a duração do emprego formal dos indivíduos de baixa renda. *Revista de Administração Pública-RAP 51*(5), 708–733. 6

Schultz, T. W. (1961). Investment in human capital. *The American economic review*, 1–17. 2

Senarc - Ministry of Citizenship (2008). *Portaria* n. 341 from october 7, 2008. 3

Silveira Neto, R. D. M. and C. R. Azzoni (2011). Non-spatial government policies and regional income inequality in brazil. *Regional Studies 45*(4), 453–461. 6

Silveira-Neto, R. d. M. and G. B. Duarte (2010). Impacto do programa bolsa família sobre a frequência escolar: o caso da agricultura familiar no nordeste do brasil. *Revista de Economia e Sociologia Rural 48*(3), 635–657. 1, 7

Skoufias, E. and V. Di Maro (2006). *Conditional cash transfers, adult work incentives, and poverty*. The World Bank. 6

Soares, S., P. H. G. F. De Souza, R. G. Osório, and F. G. Silveira (2010). Os impactos do benefício do programa bolsa família sobre a desigualdade e pobreza. *Bolsa Família 2003-2010: Avanços e Desafios*, 27–52. 1, 6

Soares, S., R. Guerreiro Osorio, F. Veras Soares, M. Medeiros, and E. Zepeda (2009). Conditional cash transfers in brazil, chile and mexico: impacts upon inequality. *Estudios económicos*. 1, 6

Soares, S. and N. Sátyro (2010). O programa bolsa família: desenho institucional e possibilidades futuras. *Bolsa Família 2003-2010: avanços e desafios, chapter 1*, 27. 5

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press. 10, 13

Appendix A

Tables with Descriptive Statistics for Cohorts 1997 and 1998.

Table 17: Descriptive statistics across BFP Exposure - Cohort 1997

		BFP Exposure Level			
Variable	Measure	0	1	2	3
Individuals	N	2,242	7,155	13,016	28,073
Schooling attained at 18 y.o.	Mean	2.30	2.47	2.52	2.58
	Std. Dev.	0.95	0.77	0.75	0.72
	Min	0.00	0.00	0.00	0.00
	Max	5.00	5.00	5.00	5.00
Formal Employed	Mean	0.35	0.38	0.36	0.31
	Std. Dev.	0.48	0.49	0.48	0.46
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
Avg. Real hourly Wage	Mean	7.44	7.27	7.25	7.17
	Std. Dev.	2.41	2.23	2.17	2.31
	Min	2.79	2.01	1.76	2.04
	Max	31.24	42.16	50.05	55.07
Years of BFP Exposure	Mean	0.00	2.28	4.94	7.85
	Std. Dev.	0.00	0.72	0.86	0.87
	Min	0.00	1.00	4.00	7.00
	Max	0.00	3.00	6.00	9.00
Age at First Employment	Mean	18.07	18.16	18.13	18.38
	Std. Dev.	1.31	1.34	1.33	1.27
	Min	16.00	16.00	16.00	16.00
	Max	20.00	20.00	20.00	20.00
Female	Mean	0.50	0.49	0.48	0.47
	Std. Dev.	0.50	0.50	0.50	0.50
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
White	Mean	0.42	0.39	0.35	0.26
	Std. Dev.	0.49	0.49	0.48	0.44
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
Migrant	Mean	0.45	0.41	0.40	0.38
	Std. Dev.	0.50	0.49	0.49	0.48
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
Formal Head	Mean	0.44	0.46	0.42	0.27
	Std. Dev.	0.50	0.50	0.49	0.44
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
Co-head	Mean	0.61	0.56	0.53	0.54
	Std. Dev.	0.49	0.50	0.50	0.50
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
Formal Co-head	Mean	0.37	0.34	0.30	0.25
	Std. Dev.	0.48	0.47	0.46	0.43
	Min	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00
Co-head Avg Real Hourly Wage	Mean	8.91	8.30	7.73	7.07
	Std. Dev.	4.92	4.40	4.00	3.83
	Min	2.74	1.56	2.02	1.57

Table 17: Descriptive statistics across BFP Exposure - Cohort 1997

		BFP Exposure Level			
Variable	Measure	0	1	2	3
	Max	46.05	54.15	55.74	55.07
Head Avg Real	Mean	6.60	6.35	6.23	5.85
Hourly Wage	Std. Dev.	3.47	3.39	3.21	2.59
	Min	2.25	1.47	1.51	1.49
	Max	43.75	52.53	49.33	43.84
Head Schooling	Mean	2.28	2.16	2.07	1.87
	Std. Dev.	1.17	1.14	1.12	1.10
	Min	0.00	0.00	0.00	0.00
	Max	5.00	5.00	5.00	5.00
Co-head Schooling	Mean	1.99	1.94	1.82	1.56
	Std. Dev.	1.17	1.13	1.11	1.09
	Min	0.00	0.00	0.00	0.00
	Max	5.00	5.00	5.00	5.00
Family Size	Mean	3.80	4.07	4.21	4.54
	Std. Dev.	1.25	1.39	1.48	1.60
	Min	1.00	1.00	1.00	1.00
	Max	10.00	14.00	16.00	14.00
Munic.	Mean	840.86	888.12	1,155.01	1,029.14
Density 2000	Std. Dev.	1,880.25	1,894.83	2,215.74	2,093.90
	Min	0.28	0.13	0.13	0.13
	Max	11,924.99	11,924.99	11,924.99	11,924.99
Munic.	Mean	184.45	204.04	218.45	197.33
Families 2000	Std. Dev.	575.01	518.42	613.40	564.02
	Min	0.37	0.34	0.33	0.26
	Max	3,173.50	3,173.50	3,173.50	3,173.50
Munic. Poor	Mean	0.46	0.45	0.47	0.55
Families 2000	Std. Dev.	0.19	0.20	0.20	0.21
	Min	0.10	0.11	0.07	0.10
	Max	0.92	0.96	0.94	0.96
Munic. Education	Mean	4.16	4.20	4.06	3.78
Quality 2007	Std. Dev.	0.75	0.76	0.76	0.76
	Min	2.20	1.80	1.80	0.90
	Max	7.70	7.50	7.70	7.60
Munic. SR	Mean	0.82	0.77	0.72	0.72
Coverage 2006	Std. Dev.	0.26	0.23	0.23	0.23
	Min	0.22	0.13	0.19	0.10
	Max	3.50	3.50	3.50	3.50

Note: This table reports descriptive statistics for each variable of the liked data for the individuals that were born in 1997, whose family was registered in the SR between 2004 and 2006. The mean of the dummy variables (Female, Migrant, Formal Head, Co-Head, and Formal Co-Head) must be interpreted as the proportion of that category in the data. The number of municipality resident families are measured in thousands. Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

Table 18: Descriptive statistics across BFP Exposure - Cohort 1998

		BFP Exposure Level				
Variable	Measure	0	1	2	3	4
Individuals	N	1,890	5,829	11,218	22,924	8,193
Schooling	Mean	2.33	2.59	2.62	2.62	2.60
attained at 18 y.o.	Std. Dev.	1.02	0.72	0.71	0.70	0.71
	Min	0.00	0.00	0.00	0.00	0.00
	Max	5.00	5.00	5.00	5.00	5.00
Formal Employed	Mean	0.27	0.27	0.27	0.20	0.21
	Std. Dev.	0.44	0.44	0.44	0.40	0.41
	Min	0.00	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00
Avg. Real	Mean	7.38	7.10	7.21	7.14	7.05
Hourly Wage	Std. Dev.	2.42	2.10	2.33	2.50	2.14
	Min	3.89	2.25	1.67	1.74	2.00
	Max	37.08	42.15	47.16	55.10	34.61
Years of BFP	Mean	0.00	2.41	4.98	8.08	10.00
Exposure	Std. Dev.	0.00	0.67	0.73	0.67	0.00
	Min	0.00	1.00	4.00	7.00	10.00
	Max	0.00	3.00	6.00	9.00	10.00
Age at First	Mean	17.71	17.70	17.81	17.92	17.90
Employment	Std. Dev.	1.14	1.16	1.14	1.10	1.11
	Min	16.00	16.00	16.00	16.00	16.00
	Max	19.00	19.00	19.00	19.00	19.00
Female	Mean	0.49	0.48	0.48	0.47	0.47
	Std. Dev.	0.50	0.50	0.50	0.50	0.50
	Min	0.00	1.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00
White	Mean	0.46	0.40	0.37	0.25	0.26
	Std. Dev.	0.50	0.49	0.48	0.44	0.44
	Min	0.00	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00
Migrant	Mean	0.42	0.39	0.39	0.38	0.36
	Std. Dev.	0.49	0.49	0.49	0.48	0.48
	Min	0.00	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00
Formal Head	Mean	0.52	0.50	0.46	0.31	0.30
	Std. Dev.	0.50	0.50	0.50	0.46	0.46
	Min	0.00	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00
Co-head	Mean	0.59	0.56	0.52	0.55	0.53
	Std. Dev.	0.49	0.50	0.50	0.50	0.50
	Min	0.00	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00
Formal Co-head	Mean	0.40	0.37	0.33	0.27	0.26
	Std. Dev.	0.49	0.48	0.47	0.44	0.44
	Min	0.00	0.00	0.00	0.00	0.00
	Max	1.00	1.00	1.00	1.00	1.00
Co-head Avg	Mean	9.04	8.65	8.23	7.47	7.15
Real Hourly Wage	Std. Dev.	4.96	4.54	4.29	4.12	3.64
	Min	2.99	2.08	2.02	1.46	1.76
	Max	49.72	53.69	56.07	54.71	49.84
Head Avg Real	Mean	6.97	6.61	6.37	6.21	6.01
Hourly Wage	Std. Dev.	3.81	3.11	2.94	3.15	2.54
	Min	1.82	1.75	1.43	1.52	1.44
	Max	47.67	46.31	55.10	52.53	39.36
Head Schooling	Mean	2.46	2.22	2.15	1.91	1.87
	Std. Dev.	1.18	1.13	1.12	1.11	1.06

Table 18: Descriptive statistics across BFP Exposure - Cohort 1998

		BFP Exposure Level				
Variable	Measure	0	1	2	3	4
	Min	0.00	0.00	0.00	0.00	0.00
	Max	5.00	5.00	5.00	5.00	5.00
Co-head Schooling	Mean	2.18	2.00	1.87	1.58	1.57
	Std. Dev.	1.13	1.12	1.10	1.08	1.08
	Min	0.00	0.00	0.00	0.00	0.00
	Max	5.00	5.00	5.00	5.00	5.00
Family Size	Mean	3.66	4.09	4.15	4.51	4.58
	Std. Dev.	1.20	1.41	1.44	1.60	1.59
	Min	1.00	1.00	1.00	1.00	2.00
	Max	12.00	17.00	14.00	15.00	16.00
Munic.	Mean	977.04	852.17	1,077.25	879.72	1,193.52
Density 2000	Std. Dev.	2,144.63	1,855.81	2,122.21	1,983.41	2,135.14
	Min	0.23	0.16	0.13	0.13	0.23
	Max	11,924.99	11,924.99	11,924.99	11,924.99	11,924.99
Munic.	Mean	213.13	208.82	217.98	156.33	242.30
Families 2000	Std. Dev.	639.98	516.14	620.29	498.60	601.53
	Min	0.53	0.48	0.27	0.26	0.44
	Max	3,173.50	3,173.50	3,173.50	3,173.50	3,173.50
Munic. Poor	Mean	0.44	0.44	0.46	0.56	0.53
Families 2000	Std. Dev.	0.19	0.19	0.19	0.21	0.20
	Min	0.12	0.10	0.11	0.11	0.10
	Max	0.91	0.93	0.96	0.96	0.96
Munic. Education	Mean	4.22	4.23	4.10	3.76	3.81
Quality 2007	Std. Dev.	0.71	0.74	0.77	0.78	0.72
	Min	1.90	1.80	1.80	0.90	1.60
	Max	6.20	6.90	7.70	7.60	7.50
Munic. SR	Mean	0.81	0.76	0.74	0.73	0.71
Coverage 2006	Std. Dev.	0.23	0.22	0.23	0.23	0.24
	Min	0.27	0.21	0.21	0.10	0.17
	Max	1.88	3.50	3.50	3.50	2.00

Note: This table reports descriptive statistics for each variable of the liked data for the individuals that were born in 1998, whose family was registered in the SR between 2004 and 2006. The mean of the dummy variables (Female, Migrant, Formal Head, Co-Head, and Formal Co-Head) must be interpreted as the proportion of that category in the data. The number of municipality resident families are measured in thousands. Source: Author's calculation with data from IBGE, INEP, MDS, and MTE.

Predicted Probabilities and Heterogeneous Effects of BFP Exposure on Schooling and FLM Participation

Table 19: Conditional Probability Prediction
Values assumed for the Control Variables

Control Variable	Value
Cohort	1998
Female	0
White	0
Migrant	0
Co-head	1
Formally Employed Head	1
Formally Employed Co-head	1
Avg. Hourly Wage Head	6.14
Avg. Hourly Wage Co-head	7.65
Schooling Level Head	1.98
Schooling Level Co-head	1.70
Family Size at 17 y.o.	4.36
Municipality Density 2000	1,022.61
Munic. Families 2000	204,273.30
Munic. % Poor Families 2000	50.99
Munic. Educ. Quality Index 2007	3.93

Note: The table presents the values assumed for the covariates for the estimation of the Conditional Probability of attaining each level of Schooling within each BFP exposure level. Source: Author's calculations with the data from MDS, MTE, IBGE, and INEP.

Table 20: Schooling Predicted Conditional Distributions

BFP Exp. Level	Schooling Level	Prob. within BFP (%)	Lower 95% C.I.	Upper 95% C.I.	Prob. within Schooling (%)
0	0	1.42	1.26	1.58	38.13
0	1	16.95	15.73	18.18	35.46
0	2	50.98	50.33	51.64	24.57
0	3	28.22	26.54	29.91	13.13
0	4	2.38	2.17	2.59	9.29
0	5	0.03	0.03	0.04	8.98
1	0	0.71	0.65	0.78	19.10
1	1	9.36	8.89	9.83	19.58
1	2	42.89	42.00	43.78	20.67
1	3	42.30	41.18	43.41	19.68
1	4	4.67	4.41	4.93	18.22
1	5	0.07	0.06	0.08	18.04
2	0	0.58	0.53	0.62	15.43
2	1	7.71	7.35	8.07	16.13
2	2	39.31	38.44	40.18	18.95
2	3	46.59	45.63	47.56	21.67
2	4	5.72	5.44	6.01	22.33
2	5	0.09	0.07	0.10	22.36
3	0	0.49	0.45	0.53	13.13
3	1	6.64	6.35	6.94	13.90
3	2	36.45	35.60	37.29	17.57
3	3	49.66	48.80	50.51	23.10
3	4	6.66	6.36	6.97	25.99
3	5	0.10	0.08	0.12	26.30
4	0	0.53	0.48	0.58	14.20
4	1	7.14	6.75	7.53	14.94
4	2	37.84	36.81	38.88	18.24
4	3	48.20	47.10	49.30	22.42
4	4	6.19	5.84	6.54	24.16
4	5	0.09	0.08	0.11	24.32

Note: The table presents the predicted probability of attaining a given Schooling level conditional on the control variables assuming the values reported in Table 19, and the BFP Exposure level. For the covariates considered in the logarithm scale in the models (parents' wages and municipality characteristics), we applied the log transformation on the values reported in Table 19. The 95% Lower and Upper Confidence Intervals estimated through Delta Method are also reported. The last column reports the schooling conditional distribution over the BFP exposure, in the hypothetical scenario in which all BFP exposure levels have the same unconditional probability. Source: Author's calculations with the data from MDS, MTE, IBGE, and INEP.

Table 21: Schooling Level Predicted Probability - Females

BFP Exp. Level	Schooling Level	Prob. within BFP (%)	Lower 95% C.I.	Upper 95% C.I.
0	0	0.98	0.82	1.14
0	1	11.80	10.48	13.12
0	2	42.21	40.44	43.97
0	3	40.49	37.90	43.09
0	4	4.46	3.90	5.01
0	5	0.06	0.05	0.08
1	0	0.54	0.47	0.61
1	1	6.94	6.42	7.46
1	2	32.80	31.46	34.13
1	3	51.82	50.50	53.14
1	4	7.78	7.21	8.36
1	5	0.11	0.08	0.14
2	0	0.43	0.38	0.49
2	1	5.61	5.23	6.00
2	2	28.86	27.67	30.05
2	3	55.36	54.34	56.38
2	4	9.59	8.97	10.21
2	5	0.14	0.11	0.18
3	0	0.37	0.33	0.42
3	1	4.89	4.57	5.21
3	2	26.39	25.31	27.48
3	3	57.25	56.40	58.10
3	4	10.93	10.28	11.59
3	5	0.16	0.12	0.20
4	0	0.39	0.34	0.44
4	1	5.05	4.64	5.46
4	2	26.96	25.57	28.35
4	3	56.84	55.76	57.92
4	4	10.61	9.81	11.41
4	5	0.16	0.12	0.20

Note: The table presents results considering only the females of the sample. It reports the predicted probability of attaining each Schooling level for each BFP Exposure level, conditional on the control variables assuming the values reported on Table 19. For the covariates considered in logarithm scale in the models (parents' wages and municipality characteristics), we applied the log transformation on the values reported on Table 19. The 95% Lower and Upper Confidence Intervals estimated through Delta Method are also reported. Source: Author's calculations with the data from MDS, MTE, IBGE, and INEP.

Table 22: Schooling Level Predicted Probability - Males

BFP Exp. Level	Schooling Level	Prob. within BFP (%)	Lower 95% C.I.	Upper 95% C.I.
0	0	1.29	1.10	1.49
0	1	16.20	14.57	17.84
0	2	53.49	52.60	54.38
0	3	26.89	24.59	29.18
0	4	2.09	1.83	2.35
0	5	0.03	0.02	0.04
1	0	0.59	0.52	0.66
1	1	8.15	7.58	8.72
1	2	43.76	42.48	45.04
1	3	42.91	41.38	44.45
1	4	4.51	4.17	4.86
1	5	0.07	0.05	0.09
2	0	0.48	0.43	0.54
2	1	6.79	6.35	7.23
2	2	40.23	39.00	41.45
2	3	46.95	45.63	48.28
2	4	5.46	5.08	5.83
2	5	0.09	0.06	0.12
3	0	0.40	0.36	0.45
3	1	5.75	5.40	6.10
3	2	36.92	35.74	38.09
3	3	50.37	49.19	51.54
3	4	6.45	6.04	6.87
3	5	0.11	0.08	0.14
4	0	0.45	0.40	0.51
4	1	6.41	5.93	6.89
4	2	39.10	37.65	40.55
4	3	48.16	46.63	49.68
4	4	5.78	5.33	6.24
4	5	0.10	0.07	0.12

Note: The table presents results considering only the males of the sample. It reports the predicted probability of attaining each Schooling level for each BFP Exposure level, conditional on the control variables assuming the values reported on Table 19. For the covariates considered in logarithm scale in the models (parents' wages and municipality characteristics), we applied the log transformation on the values reported on Table 19. The 95% Lower and Upper Confidence Intervals estimated through Delta Method are also reported. Source: Author's calculations with the data from MDS, MTE, IBGE, and INEP.

Table 23: Schooling Level Predicted Probability - Parents Not Formal

BFP Exp. Level	Schooling Level	Prob. within BFP (%)	Lower 95% C.I.	Upper 95% C.I.
0	0	1.47	1.23	1.71
0	1	15.87	14.08	17.67
0	2	51.43	50.33	52.53
0	3	28.79	26.14	31.43
0	4	2.40	2.07	2.73
0	5	0.04	0.03	0.05
1	0	0.61	0.54	0.69
1	1	7.38	6.85	7.91
1	2	39.70	38.36	41.04
1	3	46.61	45.13	48.10
1	4	5.60	5.18	6.03
1	5	0.09	0.06	0.11
2	0	0.54	0.48	0.60
2	1	6.50	6.10	6.91
2	2	37.25	36.08	38.41
2	3	49.24	48.04	50.45
2	4	6.37	5.97	6.78
2	5	0.10	0.07	0.12
3	0	0.45	0.40	0.50
3	1	5.56	5.25	5.87
3	2	34.16	33.12	35.20
3	3	52.25	51.26	53.25
3	4	7.46	7.03	7.88
3	5	0.12	0.09	0.15
4	0	0.48	0.42	0.54
4	1	5.86	5.45	6.28
4	2	35.20	33.86	36.54
4	3	51.27	49.97	52.58
4	4	7.07	6.57	7.58
4	5	0.11	0.08	0.14

Note: The table presents results considering the sample with only individuals whose parents were never formally employed. It reports the predicted probability of attaining each Schooling level for each BFP Exposure level, conditional on the control variables assuming the values reported on Table 19. For the covariates considered in logarithm scale in the models (municipality characteristics), we applied the log transformation on the values reported on Table 19. The 95% Lower and Upper Confidence Intervals estimated through Delta Method are also reported. Source: Author's calculations with the data from MDS, MTE, IBGE, and INEP.

Table 24: Schooling Level Predicted Probability - Formal Parents

BFP Exp. Level	Schooling Level	Prob. within BFP (%)	Lower 95% C.I.	Upper 95% C.I.
0	0	0.78	0.51	1.06
0	1	13.76	11.12	16.39
0	2	45.47	42.73	48.20
0	3	36.33	31.78	40.88
0	4	3.61	2.80	4.41
0	5	0.05	0.02	0.09
1	0	0.46	0.32	0.59
1	1	8.53	7.37	9.69
1	2	37.57	35.10	40.03
1	3	47.30	44.52	50.08
1	4	6.05	5.18	6.93
1	5	0.09	0.04	0.15
2	0	0.40	0.28	0.51
2	1	7.50	6.55	8.45
2	2	35.17	32.86	37.49
2	3	49.91	47.48	52.35
2	4	6.90	6.00	7.81
2	5	0.11	0.05	0.17
3	0	0.32	0.23	0.42
3	1	6.20	5.48	6.93
3	2	31.60	29.47	33.72
3	3	53.40	51.36	55.43
3	4	8.34	7.36	9.33
3	5	0.13	0.06	0.21
4	0	0.39	0.27	0.52
4	1	7.47	6.27	8.67
4	2	35.09	32.10	38.08
4	3	50.00	46.89	53.11
4	4	6.93	5.78	8.09
4	5	0.11	0.04	0.17

Note: The table presents results considering the sample with only individuals whose parents were at least once formally employed. It reports the predicted probability of attaining each Schooling level for each BFP Exposure level, conditional on the control variables assuming the values reported on Table 19. For the covariates considered in logarithm scale in the models (parents' wages and municipality characteristics), we applied the log transformation on the values reported on Table 19. The 95% Lower and Upper Confidence Intervals estimated through Delta Method are also reported. Source: Author's calculations with the data from MDS, MTE, IBGE, and INEP.

Table 25: Schooling Level Predicted Probability - Small Cities

BFP Exp. Level	Schooling Level	Prob. within BFP (%)	Lower 95% C.I.	Upper 95% C.I.
0	0	1.50	1.31	1.69
0	1	18.99	17.47	20.52
0	2	51.81	51.26	52.36
0	3	25.56	23.73	27.39
0	4	2.11	1.89	2.33
0	5	0.03	0.02	0.04
1	0	0.69	0.62	0.76
1	1	9.79	9.21	10.37
1	2	43.76	42.73	44.78
1	3	41.17	39.86	42.49
1	4	4.53	4.24	4.82
1	5	0.06	0.05	0.08
2	0	0.55	0.49	0.60
2	1	7.97	7.54	8.41
2	2	40.02	39.00	41.03
2	3	45.76	44.62	46.89
2	4	5.63	5.30	5.95
2	5	0.08	0.06	0.10
3	0	0.47	0.43	0.52
3	1	6.97	6.60	7.33
3	2	37.43	36.44	38.43
3	3	48.58	47.54	49.61
3	4	6.46	6.11	6.81
3	5	0.09	0.08	0.11
4	0	0.50	0.44	0.55
4	1	7.28	6.82	7.75
4	2	38.29	37.07	39.51
4	3	47.67	46.36	48.97
4	4	6.17	5.76	6.58
4	5	0.09	0.07	0.11

Note: The table presents results considering the sample with only individuals from municipalities that had less than 200,000 resident families in 2000. It reports the predicted probability of attaining each Schooling level for each BFP Exposure level, conditional on the control variables assuming the values reported on Table 19. For the covariates considered in logarithm scale in the models (parents' wages and municipality characteristics), we applied the log transformation on the values reported on Table 19. The 95% Lower and Upper Confidence Intervals estimated through Delta Method are also reported. Source: Author's calculations with the data from MDS, MTE, IBGE, and INEP.

Table 26: Schooling Level Predicted Probability - Big Cities

BFP Exp. Level	Schooling Level	Prob. within BFP (%)	Lower 95% C.I.	Upper 95% C.I.
0	0	1.21	0.92	1.50
0	1	11.58	9.58	13.58
0	2	47.04	44.40	49.69
0	3	37.09	32.90	41.28
0	4	3.02	2.41	3.62
0	5	0.05	0.03	0.08
1	0	0.91	0.74	1.08
1	1	9.00	7.93	10.06
1	2	42.85	40.70	45.00
1	3	43.19	40.45	45.92
1	4	3.98	3.44	4.53
1	5	0.07	0.04	0.11
2	0	0.75	0.62	0.89
2	1	7.57	6.69	8.46
2	2	39.65	37.45	41.86
2	3	47.15	44.61	49.70
2	4	4.78	4.16	5.40
2	5	0.09	0.05	0.13
3	0	0.59	0.48	0.69
3	1	5.99	5.31	6.67
3	2	35.11	32.96	37.26
3	3	52.12	49.91	54.33
3	4	6.08	5.35	6.81
3	5	0.11	0.06	0.17
4	0	0.72	0.58	0.86
4	1	7.24	6.33	8.16
4	2	38.81	36.40	41.21
4	3	48.13	45.41	50.86
4	4	5.01	4.32	5.69
4	5	0.09	0.05	0.14

Note: The table presents results considering the sample with only individuals from municipalities that had more than 200,000 resident families in 2000. It reports the predicted probability of attaining each Schooling level for each BFP Exposure level, conditional on the control variables assuming the values reported on Table 19. For the covariates considered in logarithm scale in the models (parents' wages and municipality characteristics), we applied the log transformation on the values reported on Table 19. The 95% Lower and Upper Confidence Intervals estimated through Delta Method are also reported. Source: Author's calculations with the data from MDS, MTE, IBGE, and INEP.

Table 27: FLM Participation Predicted Probability - Heterogeneous Effects

Sample	BFP Exposure Level	Prob. within BFP (%)	Lower 95% C.I.	Upper 95% C.I.
Complete	0	25.00	23.71	26.29
	1	26.49	25.45	27.53
	2	27.72	26.73	28.70
	3	28.43	27.47	29.39
	4	28.58	27.21	29.95
Females	0	20.30	18.68	21.92
	1	20.82	19.53	22.12
	2	21.62	20.40	22.84
	3	22.24	21.05	23.43
	4	22.34	20.60	24.08
Males	0	23.58	21.91	25.25
	1	25.87	24.52	27.22
	2	27.36	26.07	28.65
	3	28.02	26.78	29.27
	4	28.22	26.43	30.00
Parents Never Formally Employed	0	14.69	13.28	16.10
	1	15.65	14.55	16.74
	2	16.69	15.67	17.70
	3	17.07	16.13	18.00
	4	17.83	16.41	19.25
Parents Formally Employed	0	27.54	24.32	30.77
	1	27.20	24.69	29.70
	2	29.06	26.66	31.45
	3	28.55	26.27	30.83
	4	27.90	24.12	31.69
Small Cities	0	27.81	26.21	29.40
	1	30.04	28.71	31.37
	2	30.99	29.73	32.24
	3	32.17	30.93	33.41
	4	32.81	31.05	34.58
Big Cities	0	22.26	19.39	25.14
	1	21.18	19.11	23.25
	2	22.89	20.80	24.98
	3	22.26	20.25	24.26
	4	21.73	19.22	24.24

Note: The table presents the predicted probability of participating in the Formal Labor Market over the BFP Exposure level conditional on the control variables assuming the values reported on Table 19. For the covariates considered in logarithm scale in the models (parents' wages and municipality characteristics), we applied the log transformation on the values reported on table. For each indicated sample the model is estimated again, and then the prediction is performed with the respective set of coefficients. The 95% Lower and Upper Confidence Intervals are estimated through the Delta Method. Source: Author's calculations with the data from MDS, MTE, IBGE, and INEP.