Barbosa and Corseuil *IZA Journal of Labor & Development* (2014) 3:37
DOI 10.1186/s40175-014-0024-0

IZA Journal of Labor & Development
a SpringerOpen Journal

ORIGINAL ARTICLE Open Access

# Conditional cash transfer and informality in Brazil

Ana Luiza Neves de Holanda Barbosa* and Carlos Henrique Leite Corseuil*

* Correspondence:
ana.barbosa@ipea.gov.br; carlos.corseuil@ipea.gov.br
IPEA, Rio de Janeiro, Brazil

**Abstract:** This study analyses the impact of Bolsa Família (a Brazilian conditional cash transfer programme) in the allocation of labour inputs regarding formal and informal activities. The methodology relies on a discontinuity in the programme eligibility rule regarding children's age to attain the identification of a LATE parameter. Our results suggest that the cash transfer does not affect the occupational choice of Brazilian adults neither among formal and informal jobs nor the allocation of hours across sectors.

**JEL Codes:** I38, O17.

**Keywords:** Income transfer; Informal labour; Regression discontinuity design

## 1 Introduction

Bolsa Família Program (henceforth known as BFP) is a conditional cash transfer (CCT) which aims to improve the welfare of poor families. There are programmes with similar characteristics to BFP in several developing countries, both inside and outside Latin America. In some countries such as Brazil, the programme coverage is quite extensive. For example, in 2013, BFP had already benefited more than 13 million households. In this context, some concerns are raised regarding possible deleterious effects of this CCT on the labour supply of adults[1].

As CCT are usually targeted to low income households, there may be an incentive to marginally eligible households to receive earnings in ways non-verifiable to third parties, specially government or inspectors of the CCT program. In this sense informal activities could be preferable, even more when there is high prevalence of this kind of activities, as in Brazil, or Latin America in general.

The aim of this study is to investigate to what extent CCT programmes like BFP distorts the occupational choice of beneficiaries in the sense of making them more likely to shift supply of labour inputs pro informal activities. In particular, the goal is to analyse BFP impacts both at intensive (hours) and extensive (jobs) margins of labour supply.

This shift in labour supply pro informal activities would be deleterious to society, for at least three reasons. First, a change in the composition of employment, which reduces the share of the formal sector, compromises the state's ability to fund its policies, including those directly related to social welfare. A second reason for concern is the lower quality associated to informal jobs in dimensions such as lower wages, higher probability of workplace accidents and lack of social protection[2]. Lastly an informal attachment to labour market could have deleterious effects for the worker in the long run. Cruces et al. (2012) reports lower prospect of wage growth for those employed in the informal sector in early stages of their professional life.

Brazil can be considered an excellent country for analysis, given the availability in a household survey with national coverage (PNAD) of accurate information both about the formality status (formal or informal) of each job of each member of a sampled household and on the status of the household on both eligibility and participation in BFP (beneficiary or non-beneficiary).

There is, however, a major challenge for identification of BFP effects on any outcome of interest, as is the case for labour supply in the present study. There is room for self selection into the programme based on non-observable characteristics of the individuals. This makes it difficult to find an ideal comparison group that resembles the treatment group.

Our strategy for identifying the effect of BFP on labour supply composition among formal and informal activities seeks to isolate the real effect of the programme from outcome differences due to unobservable characteristics of individuals. We used a discontinuity in the BFP eligibility rule, which is a threshold value for the age of the youngest children at the household, as an instrument that shifts participation in the programme but does not change the outcome of interest (informality of labour inputs). This type of strategy is known as Regression Discontinuity Design – RDD[3]. In particular, since eligibility does not coincide with participation in the programme, we employ the version of this strategy known as a *fuzzy* RD. Our main identification hypothesis is to assume that the unobservable characteristics of adults in families whose youngest child reaches the age threshold for eligibility shortly *before* the beginning of the school year are very similar to those of families whose youngest child reaches the eligibility age threshold shortly *after* the beginning of the school year. This is considered a weak assumption, given the randomness component involved in determining the exact date of birth of an individual. In addition, we carried out tests whose results are consistent with the validity of this hypothesis.

Empirical evidence of the impact of CCT programmes on the occupational choices of adult programme beneficiaries is still incipient. What we do have are studies estimating the impact on the labour supply of individuals disaggregated by occupation type (Skoufias and Di Maro (2008); Ribas and Soares (2011) and Braw et al. (2012))[4,5].

These three papers use some version of the difference-in-differences estimator. The identification hypothesis in these works, therefore, is that the unobservable characteristics of either individuals or group of individuals that influence the selection of the programme are fixed in time. This could be a strong assumption as information about availability of the program and eligibility criteria may vary over time, and this may change the composition of treatment and control group relative to preferences over informal jobs[6].

The result found by Skoufias and Di Maro (2008) for Mexico indicates an effect on the participation of self-employed individuals immediately after the implementation of the programme. However, this effect disappears shortly thereafter. Ribas and Soares (2011) identifies a decrease in the participation in the formal sector due to BFP. Braw et al. (2012) reports an increase in working hours within the informal sector due to BFP.

Our results do not support the hypothesis that the BFP stimulate the informality among beneficiaries in any of the informality dimensions considered. We show that barely eligible households do have a higher probability of participation in BFP than barely ineligible ones. However this difference in participation does not translate into difference in the supply of labour inputs towards the informal sector. It is worth mentioning that this result is

robust for distinct margins of informality: namely intensive (hours) and extensive (jobs). Moreover jobs were considered in two alternative ways. First we consider only the main job of the household head. The other alternative is to consider the second most important job of the household (either the secondary job of the household head or the main job of his/her partner).

This paper is organized into four sections, besides this introduction and a final section devoted to concluding remarks. The next section presents the BFP rules, including eligibility criteria and benefit amounts. The third section describes the econometric procedures adopted for the evaluation of BFP's impact on the composition of the labour market. The database and the construction of our sample are presented in the fifth section, which also provides some descriptive statistics. The main econometric results are presented in the sixth section.

## 2 BFP Eligibility criteria

Programme eligibility depends on two main criteria. The first is associated with socio-economic status as measured by per capita household income (PCHI)—the sum of the gross income earned by all family members divided by the total number of individuals in the household. The second criterion is linked to family composition, in particular to the age of the youngest child. BFP also requires the recipient to comply with conditionalities associated with child health and education[7].

In September 2006, two household groups were considered eligible. The first group was composed of households with monthly PCHI between R$ 50.01 and R$ 100.00 (below the poverty line at the time) provided they had pregnant women, nursing mothers, children and adolescents aged 0 to 15 years old. This group received a variable benefit, depending on the number of members in the situations listed above, reaching the limit of R$ 45.00. The other group consists of families in extreme poverty, whose PCHI was less than or equal to R$ 50.00 monthly. Such families would receive a basic benefit of R$ 50.00, regardless of their composition, and a variable benefit if there were dependents among family members[8]. The maximum amount of BFP benefits that an extremely poor family could receive was R$ 95.00.

Despite several studies that seek to reap the benefits of BFP exploring the condition of eligibility based on the PCHI, in this paper we chose to explore the eligibility condition based on the age of the youngest child. As we will see later our methodology depends, therefore, on individuals not being able to manipulate the value of the variable of the eligibility criteria chosen. Given the existence of a birth certificate in Brazil, it seems more reasonable to assume that there is no way to manipulate the age of the children than it is to assume that there is no way to manipulate the PCHI. This argument is an immediate consequence of the fact that the income of an informal occupation can be hidden from those responsible for running BFP, unlike what happens with the income derived from formal occupations[9].

It is important to note how household selection takes place within the programme[10]. The selection of BFP beneficiary households takes place under the aegis of the Ministério do Desenvolvimento Social e Combate à Fome (MDS), Caixa Economica Federal (CEF) and the municipalities[11]. The National Secretariat of Citizenship Income (SENARC), under the MDS, is the agency responsible for the programme. Among other important

actions, SENARC establishes the criteria for selecting beneficiaries—that is, those who receive the benefit and how much they shall receive. It also defines the questionnaire used to fill an administrative file with all potential beneficiaries (*Cadastro Único*) and the criteria for suspending and cutting benefits, among several other operational parameter definitions. The role of CEF is quite significant for BFP. Besides being the operator and payment agency, it is responsible for processing the information collected by municipalities that makes up the *Cadastro Único*, as well as calculating the PCHI and how much each family should receive, and paying the monthly benefit through magnetic debit cards. The most important role of municipalities is the responsibility for identifying families to be benefited, by gathering information that will be fed into the *Cadastro Único*. Sátyro and Soares (2009) argue that, ultimately, municipal officers are the ones who decide who will or will not be a potential beneficiary of *Bolsa Família*, in the sense that all the information handled by CEF or analysed by SENARC is collected by them. According to the authors, this "is a strategic role – without committed and well managed municipalities, the very functioning of BFP would be compromised" (SOARES; SÁTYRO, p.17). It is clear, then, that given the form of BFP management, beneficiaries tend to have an incentive to work in the informal sector (with their incomes hidden from government agencies) and still remain eligible for the BFP benefit.

In the next section, it will become evident that, the process of excluding households that become ineligible is more important than the process of adding new households to its rank of beneficiaries. In particular, the process of excluding households that become ineligible according to the criterion of the age of the youngest child is central to our analysis. According to BFP bylaws, exclusion does not occur immediately following the 16th birthday of the youngest child, but rather after the end of his/her school year.

As such, we shall base our empirical strategy—to be explained in the next section—on comparisons of households that appear in the 2006 PNAD and in which the youngest child was about to turn 16 years old on December 31, 2005 with households in which the youngest child had just turned 16 years old by December 31, 2005.

## 3 Methodology

### 3.1 The identification problem and its consequences for estimation

The identification of the effect of BFP on occupational choice is not a trivial task, given that there is room for self-selection of individuals in the programme. This problem arises when the probability of an individual entering the programme depends on unobservable characteristics that may also influence their occupational choice.

The previous section provides evidence that there is room for this kind of problem, in the case of BFP. Basically, the probability of selection will be greater for an eligible household that is better capable of becoming visible to the eyes of the municipal agents.

For a better view of the problems arising from self-selection based on unobservable characteristics, consider the following regression model relating the supply of labour to the informal sector from household "i" ($Y_i$) to the age of the youngest child on the last day of the previous year ($I_i$), as well as participation in BFP ($T_i$)[12]:

$$Y_i = \alpha_1 + \beta_1(I_i - c) + \lambda_1 T_i + \gamma_1 T_i(I_i - c) + \delta_1 X_i + \varepsilon_i, \quad (7)$$

where c is a normalization constant representing the age limit of the youngest child to the eligibility of the household (in our case, 16 years, exactly, on December 31,

2005), X denotes controls regarding characteristics of the household head (age, gender and education), household size and location. Our parameter of interest is $\lambda_1$, which represents the effect of program participation on the supply of labour input to the informal sector.

The fact that there is self-selection in BFP, based on unobservable characteristics, leads to the following property of this model:

$$E[\varepsilon_i|T_i; X_i, I_i] \neq 0 \tag{8}$$

Consequently, $\lambda_1$ cannot be identified by comparing the occupational choices of those who participated (T = 1) and those who did not participate (T = 0), even when keeping the other explanatory variables of the model (X) constant. This is because, according to (8), when T is varying ε is also varying. Therefore, we will not be able to separate the effect on Y arising from the variation of T from that arising from the variation in ε. In terms of estimation, this means that $\lambda_1$ cannot be estimated consistently by ordinary least squares.

In this case, the ideal scenario is to have some source of variation in T among individuals that does not affect ε. Put another way, it would be interesting to have a variable at hand that affects the decision of individuals to participate in the program and which is not related to unobservable characteristics of individuals. This variable serves as an instrument for the correct identification of our parameter of interest, because we do not want to know the impact of this variable, but its variation is used to shift participation, while keeping everything else constant[13].

Generally, we resort to an exogenous eligibility criterion as the instrument for the correct identification of the treatment. This exogenous criterion may be due to a random draw or a discontinuity in the rule of eligibility. In the latter case, we have what is conventionally called a fuzzy regression discontinuity[14] in literature.

As discussed in the previous section, there is a discontinuity in BFP's eligibility rule, which prompts us to choose the fuzzy regression discontinuity approach. The identification strategy associated with this approach can be viewed in Figure 1, which shows participation in the program (T) in September 2006 (date of the 2006 PNAD) as a function of the gap that separates the age of the youngest child from the 16-year mark as of 31/12/2005. Thus, positive (negative) values on the horizontal axis denote households where the youngest child had not yet (already had) turned 16 years of age by 31/12/2005 and were, therefore, (in)eligible to the program in 2006.

Each point on the figure represents average values for household groups that were grouped according to the age of the youngest child and, therefore, also in accordance with the aforementioned gap. The aggregation used two-month intervals, so that the point at 1 on the horizontal axis shows the share of beneficiaries among households whose youngest child was less than two months away from turning 16 years of age, by 31/12/2005; as opposed to the point −1, which shows the share of beneficiaries among households whose youngest child had just turned 16 years of age in the last two, by 31/12/2005.

The comparison between households with young children who recently turned 16 or who will turn 16 shortly after 31/12/2005 brings noticeable differences to the likelihood of participating in the programme in 2006. Reading Figure 1 from the right to the left clearly shows that participation in the programme drops sharply around the mark corresponding to the threshold for one of the eligibility criteria. Reasonably assuming that

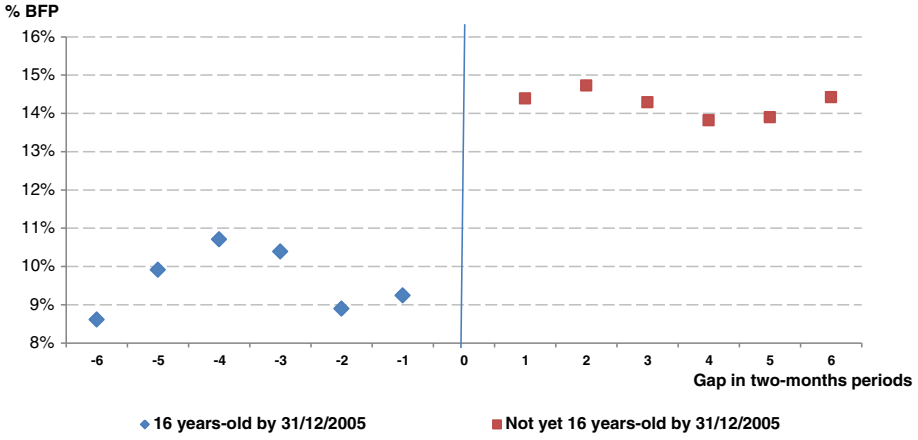


**Figure 1 Share of Households that Receive the BFP benefit.** Source: 2006 PNAD.

everything else should be very similar between these two household groups (including the values of ε), the age of the youngest child around 16 years is the exogenous variation that shifts T, keeping everything else constant, and, therefore, allows us to identify the effect of Bolsa Família.

In the remainder of this section we will discuss identification in a more precise fashion. A number of contributions have been made to guide researchers on the assumptions required to identify the effect of a programme in this context, as well as some operational procedures for its estimation. Our intent is to follow the recommendations of this recent literature, which can be accessed in the papers by Imbens and Lemieux (2008) and Lee and Lemieux (2010), among other sources.

### 3.2 The solution via discontinuity

The identification and, consequently, the consistent estimation of parameter $\lambda_1$ in equation (7) require some additional assumptions. For the sake of exposition, we assume, at first, that Bolsa Família uniformly affects the occupational choice of individuals (a context that is often denoted in literature as 'homogeneous effects'). In this context, the following properties for households where young children are in an age around c are sufficient to identify our parameter of interest[15]:

$$E[\varepsilon_i|I_i = c^- = E\varepsilon_i|I_i = c^+] = 0 \quad (9)$$

$$E[T_i|I_i = c^-] \neq E[T_i|I_i = c^+] \quad (10)$$

In the case of the above equations, $c^-$ ($c^+$) denotes ages that are very close to c, but smaller (larger) than the exact value. In our context $c^-$ ($c^+$) represents households with young children aged slightly under (over) 16 on 31/12/2005.

The hypothesis represented by (9) has the following interpretation within the context of this article: the ability to become visible to program inspectors does not differ between households where the youngest child was about to turn 16 on 31/12/2005 and households where the youngest child has recently turned 16 on that same date. The hypothesis represented by (10) requires that these same two groups of households differ

with respect to the probability of selection into the programme. It is expected that the group where the youngest child had not turned 16 would have a higher probability of selection.

We have postponed a more detailed discussion on the validity of these hypotheses to the next section. For now, we assume that both are valid and proceed with the presentation of our identification strategy. From equation (7), it is easy to see that identification comes from the following ratio:

$$\lambda_1 = \{E[Y_i|c^-] - E[Y_i|c^+]\}/\{E[T_i|c^-] - E[T_i|c^+]\} \quad (11)$$

This strategy is equivalent to the use of an instrumental variable. In this case, the eligibility works as an instrument for participation. In intuitive terms, it explores the contrast between individuals with ages that are very close to the eligibility threshold. The variation in eligibility for these individuals induces an exogenous variation in BFP participation which, in turn, induces a causal effect on the probability of being employed in the formal sector.

For operational purposes, we will use a sample restricted to households with a youngest child aged around "c", given the unique role of these households in the identification of the parameter of interest. In this sample, we apply the instrumental variable estimator, where the first stage may be represented by the following equation:

$$T_i = \alpha_2 + \beta_2(I_i - c) + \lambda_2 D_i + \gamma_2 D_i(I_i - c) + \delta_2 X_i + \zeta_i \quad (12)$$

where $D_i$ indicates whether the household is eligible through the age criterion, i.e.:

$$D_i = 1.(I_i < c) \quad (13)$$

To put it in another way, in our strategy, $D_i$ serves as an instrument for $T_i$. Another important issue from the operational perspective is what should be considered "near the age limit for eligibility". For this definition, we follow the procedure suggested by Imbens and Kalyanaraman (2012), noting that such a procedure is proposed for the case of sharp discontinuity. In the case of fuzzy discontinuity, as in our estimation, there is a procedure designed to identify the optimal size of the window, but it is known that this size must be a little larger than that indicated for the sharp case.

#### *3.2.1 Heterogeneous effects and interpretation of the identified parameter*

In the literature on impact assessment, it is common to relax the hypothesis of homogeneous effect, allowing the benefits of the programme in question to vary between individuals. Thus, equation (7) shall present $\lambda_{1i}$ as opposed to $\lambda_1$. Nothing changes for the matter of identification and estimation procedures described previously.

However, there is a significant change in the interpretation of the parameter identified by the right-hand side of the equation (11) and estimated by instrumental variables. The identified parameter, denoted by LATE, becomes the expected value of the effect of the programme for a subgroup of the population known as "*compliers*". This subgroup consists of households whose participation is defined by eligibility. In other words, they participate when they are eligible and do not participate when they are not. In our article, this translates into households participating in 2005 with children who are 16 years old in December and who do not participate in 2006, when they become ineligible[16].

Besides the changed interpretation, the environment with a heterogeneous effect also requires two additional identification hypotheses[17]. The first, known as "monotonicity", requires that there be no household that decides to participate in the program when it is not eligible and decides not to participate when it is eligible (what literature calls "*defier*" behaviour). That is, besides the compliers, we admit the existence of only two other types of households for whom participation is not guided by eligibility: the one that always participate regardless of being eligible or not (*"always taker"*) and the one that never participate regardless of being eligible or not ("*never taker*").

The second additional hypothesis is more abstract and it is known as independence. In our context, this hypothesis states that the potential outcomes of participation and occupational choice conditioned to each of the possible eligibility statuses (eligible or not) is independent of the eligibility status actually observed for the household. For example, an eligible household which actually participates in the programme would have an unobserved participation result if it were not eligible. The hypothesis in question requires that these two potential outcomes of participation, when it is eligible and when it is not, be independent of whether the household is eligible or not. The same goes for potential occupational choices associated to the condition of being eligible or not.

It is important to note that these two additional hypotheses are not testable, unlike the first two cases described in section 3.2, which will be considered again in the next section.

## 4 Preliminary empirical information

### 4.1 Data and descriptive statistics

As seen in the previous section, the implementation of our identification strategy requires the availability of household-level information about the children's dates of birth, supply of labour to informal sector and about receiving the BFP benefit. The National Household Sample Survey (PNAD) 2006, conducted by the Brazilian Institute of Geography and Statistics (IBGE) and used in this study, contains this information.

PNAD is an annual household survey that covers the entire country and collects information on the features of the homes and their residents. In particular, the information on the date of birth and job (in)formality were always present in the survey's traditional module. In the 2006 PNAD, there is an additional module on the characteristics of the access to some income transfers from social programmes at the households. In this additional module, there is a question specifically about Bolsa Família[18]. The 2006 PNAD surveyed 410,241 people, in 145,547 households nationwide.

Regarding the definition of informality, we work with the one used in most studies on informality with Brazilian data, which distinguishes employees by having an employment registration signed by the employer (*carteira assinada*) and allocates all employers to the formal sector and the self-employed to the informal.

At this point of the analysis, the definition of our sample is subject to two constraints. The first restriction was imposed by our methodology, which makes exclusive use of observations around the threshold that defines eligibility. In our context, this means restricting the sample to households with the youngest child at around 16 years of age.

The second filter refers to workers with problematic occupational insertions[19]. Further on we will comment on additional restrictions.

Figure 2a and b show two different measures of labour input devoted to the informal sector, as a function of the gap in months (measured on 31/12/2005) that separates the age of the youngest child from his/her 16th birthday. The measures are share of household working hours in the informal sector (graph 2a) and share of household where the main job of the head is an informal one. These figures are constructed analogously to Figure 1. Thus, positive (negative) values on the horizontal axis represent households (in)eligible for the programme throughout 2006 by the criterion of the age of the youngest child. And, as in Figure 1, the points on the graphs represent average values of household groups that were grouped together according to the age of the youngest child, in two-month periods (see Figure 3).

It is noted that either measure of employment in the informal sector is quite similar among the two groups being compared. That is, labour input devoted to the informal sector does not seem to vary with the household's eligibility to BFP. It is worth noting

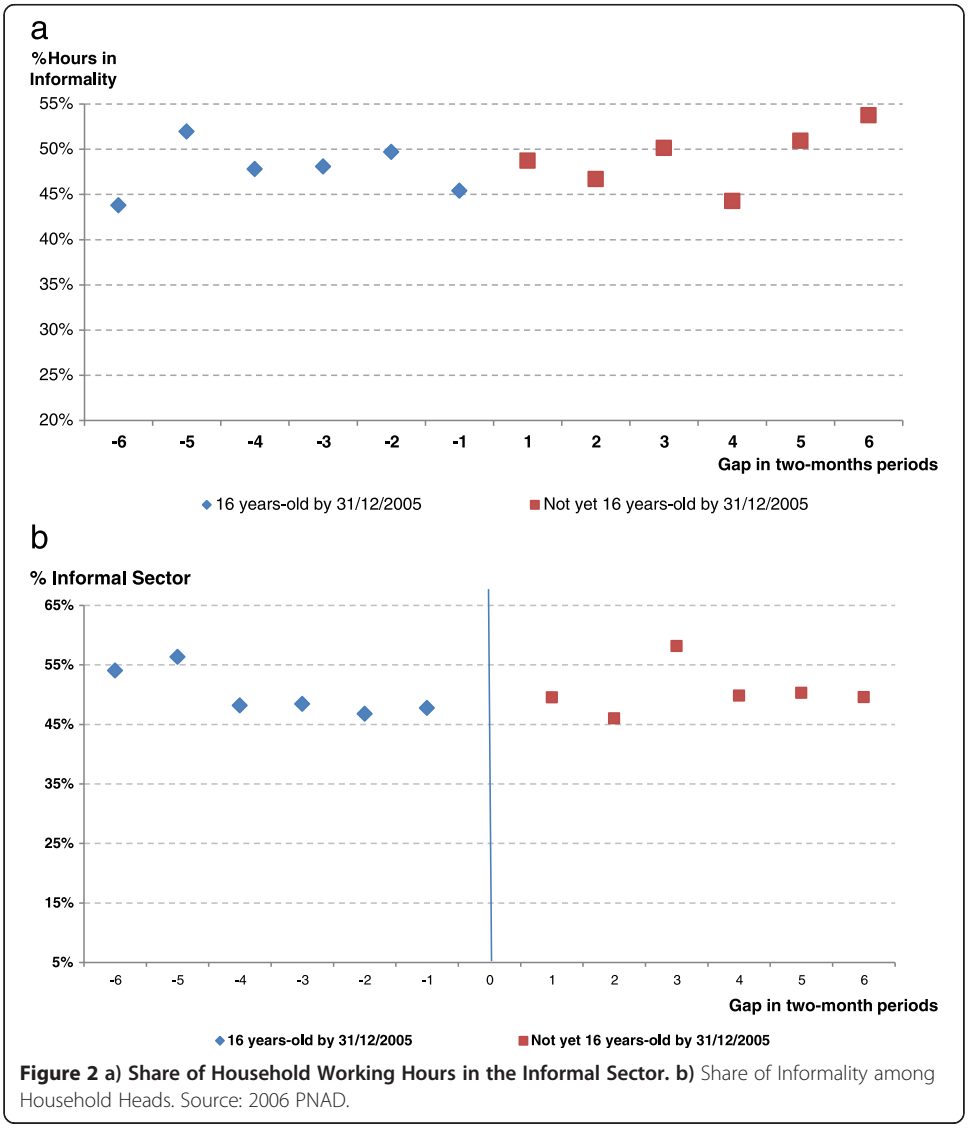


**Figure 2 a) Share of Household Working Hours in the Informal Sector. b)** Share of Informality among Household Heads. Source: 2006 PNAD.

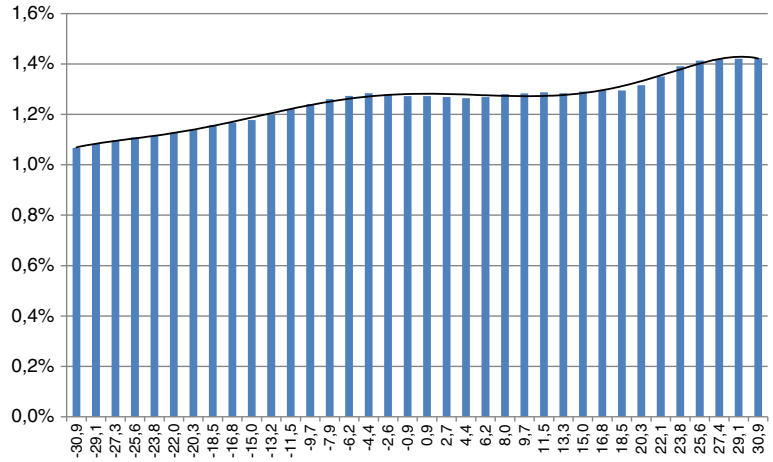


**Figure 3 Estimate of the density function of the gap in months between the age of the youngest child on 31/12/2005 and exactly 16 years of age.** Source: 2006 PNAD.

that the eligible group showed a significantly greater participation in BFP than the ineligible group. Therefore, it is noted that eligibility influences participation which, in turn, does not seem to influence informality. This result will be confirmed in our estimates with the instrumental variable method, adapted to the context of fuzzy regression discontinuity. Before going to the results, however, we shall report some evidence that shows that the method is, indeed, adequate for our purposes.

### 4.2 Testing identification hypotheses

In summary, the identification of our parameter of interest relies on four hypotheses: *i)* condition of local exclusion (around c), represented by (9); *ii)* validity of the instrument (also around c), represented by (10); *iii)* monotonicity; and *iv)* independence. Ideally, we would like to be able to test the validity of all of them. However, the last two are not testable hypotheses. Therefore, the provision of evidence about the validity relies on the first two.

#### *4.2.1 Validity of the Instrument*

The hypothesis of the validity of the instrument is easily tested, since it consists in comparing the conditional means of observed variables, as expressed in equation (10). The coefficient $\lambda_2$ in the equation (12) captures exactly the difference between the two sides of equation (10). Therefore, testing the significance of this coefficient is one way of testing the hypothesis of the instrument's validity.

Note that we are including in this sample a group with a high probability of behaving like an "always taker", composed of households with PCHIs less than or equal to R$ 50.00. As stated earlier, this group is eligible regardless of the household's age composition. Thus, the propensity to participate in the programme for this group of households may not vary with the age of the youngest child.

On the other hand, households with high PCHIs tend to behave like "never takers", since the probability of being registered tends to be very low, regardless of the age of the youngest child. In fact, the percentage of beneficiaries is 0.7 per cent when we analyse households with PCHIs over R$ 700.00[20].

These groups may hinder the identification of $\lambda_1$ since, as discussed in section 3, the source of identification in our empirical strategy is the group of "compliers". Therefore, the smaller the share of "compliers" or greater the participation of "always takers" and "never-takers", the lower the power of our identification strategy. In particular, a decrease in the share of "compliers" tends to invalidate our second identification hypothesis, represented by (10).

As such, we tested this hypothesis with three different samples. In the first sample, we did not add any restriction relative to what was mentioned in the previous section (sample I). In the second sample, we excluded households with PCHIs over R$ 700.00 (sample II). In the third sample, we excluded households with PCHIs lower than R$ 50.00 and over R$ 700.00 (sample III).

Table 1 above reports the estimates for $\lambda_2$ in three pairs of columns corresponding to the three samples. In each pair of columns we report the coefficients estimates and their standard errors in the first column; the second statistic is in brackets below the first one.

The results in Table 1 indicate that among the 12 specifications considered (4 windows x 3 filters), only one coefficient is not significantly different from zero at the 95 per cent confidence level. In all other specifications, the estimates are positive and significantly different from zero, a result which supports the hypothesis that the eligibility criteria based on the age of the youngest child positively affects the propensity to participate in the program, for households where the age of the youngest child was close to 16 on 31/12/2005. In other words, we have evidence that the propensity to benefit from BFP drops significantly when the household does not meet the eligibility criteria for BFP's variable benefit.

A more formal test of the strength of the instrument is provided by a F-statistics in the first stage that compare OLS and IV models. We perform these tests for the regressions using the optimal bandwidth (first line). The results confirm the power of our instrument for the sample with no further restrictions (first column). For the other samples with additional restrictions (second and third column) one has to be more

**Table 1 Instrument validity test (household head)**

| | **Total sample** | | **Sample I** | | **Sample II** | |
|---|---|---|---|---|---|---|
| | **All households receiving the PBF benefit** | | **Households with PCHIs < R$ 700.00** | | **Households with PCHIs > R$ 50.00 and PCHIs < $ 700.00** | |
| | **Coef.** | **Optimal window** | **Coef.** | **Optimal window** | **Coef.** | **Optimal window** |
| λ**2** | 0.057 | 1.61 | 0.081 | 1.78 | 0.081 | 1.76 |
| (standard error) | (0.049) | | (0.030) | | (0.030) | |
| λ**2** (bandwidth = **110**) | 0.062 | 1.77 | 0,069 | 1.96 | 0.076 | 1.93 |
| (standard error | (0.025) | | (0.029) | | (0.029) | |
| λ**2** (bandwidth = **125**) | 0.066 | 2.01 | 0.059 | 2.23 | 0.065 | 2.20 |
| erro-padão) | (0.024) | | (0.027) | | (0.027) | |
| λ**2** (bandwidth = **150**) | 0.089 | 2.41 | 0.062 | 2.67 | 0.066 | 2.64 |
| (standard error) | (0.042) | | (0.025) | | (0.026) | |
| F-test[1] | 67,781 | | 5,009 | | 7,781 | |
| p-value | 0.000 | | 0.000 | | 0.000 | |

Source: 2006 PNAD.
Note: F-statistics are related to regressions with results reported in the first line.

complacent with critical values of the F-test in order to reject the hypothesis of weak instruments. Since these two samples were built to increase the power of the instrument, and the opposite happened, we will have the unrestricted sample as our baseline.

In the second column of each pair of columns we report the window size used. As mentioned earlier, there is no indicated procedure to determine the optimal size of the window in the case of fuzzy discontinuity. Thus, we consider alternative sizes, given that the optimal size for the window, in the case of fuzzy discontinuity, should be slightly higher than that recommended in the case of acute discontinuity. The four different window sizes considered originate the four lines of results. In the first line we consider the size recommended by the procedure for acute discontinuity. In the other lines we consider larger windows: 10 per cent larger in the second line, 25 per cent larger in the third line and 50 per cent larger in the fourth line. The indicated window sizes were around two months.

#### *4.2.2 Condition of exclusion*

The condition of exclusion around c is not so easy to test, since it involves the behaviour of a variable that has not been observed. This hypothesis requires the continuity of the unobservable variable around c. There are two common procedures in the literature to provide indirect evidence about the validity of this hypothesis. In the first procedure, there is a search for evidence of continuity in the control variables (x). It is expected that a sharp discontinuity in the observed variables may reflect discontinuities in unobserved variables.

Table 2 presents a comparison of certain observable characteristics of the household head and of the family as a whole, recorded in the 2006 PNAD among households where the age of the youngest child is slightly less and slightly more than 16 on 31/12/2005. To define the groups marked as eligible and ineligible, we used a radius of two months, as in the previous item. We also followed the tables in the previous section regarding the delineation of three distinct samples, according to the use of per capita household income for this purpose.

**Table 2 Balancing—characteristics of the eligible and the non-eligible**

| | **Total sample** | | **Sample I** | | **Sample II** | |
|---|---|---|---|---|---|---|
| | **All households receiving the PBF benefit** | | **Households with PCHIs < R$ 700.00** | | **Households with PCHIs > R$ 50.00 and PCHIs < $ 700.00** | |
| | **Eligible** | **Non-eligible** | **Eligible** | **Non-eligible** | **Eligible** | **Non-eligible** |
| Individual characteristics | | | | | | |
| **Gender** (% de men) | 78.9 | 75.5 | 78.2 | 77.6 | 78.5 | 77.1 |
| **Years of schooling** | 6.9 | 7.2 | 6.0 | 6.0 | 6.1 | 6.1 |
| **Age** | 45.3 | 44.4 | 44.8 | 43.7 | 44.8 | 43.6 |
| Household characteristics | | | | | | |
| No. of people in the family | 3.6 | 3.5 | 3.6 | 3.5 | 3.6 | 3.5 |
| Resident in the NO/NE regions | 38.7 | 40.6 | 39.9 | 42.4 | 38.4 | 42.0 |
| Total per capita income | 508.1 | 544.1 | 306.3 | 304.4 | 313.3 | 310.1 |
| Number of observations | 284 | 310 | 238 | 250 | 232 | 245 |

Source: 2006 PNAD.

In general, we can say that the numbers in a group never stray too far from the respective comparison group. For some variables such as age of the head of household and the number of people in the household, the proximity between groups is striking in any of the three samples considered. Even in the case of other variables, the values tend not to differ significantly between the eligible and ineligible groups. The biggest changes are recorded for variable geographical location and household income. Even the difference in this last variable is limited to the first sample only.

In the second procedure, to provide evidence regarding the validity of the condition of exclusion, the evidence is derived from the expected consequences on the behaviour of variables observed in a scenario where this assumption is no longer valid.

In our context, we can think of two types of scenarios where the condition of exclusion ceases to be valid: *i)* if the household manipulates information about the age of the youngest child, and *ii)* if the household conceives another child when the age of the youngest child is approaching 16.

If the first scenario is correct, one would expect a discontinuity in the distribution of households according to the age of the youngest child, with a pronounced concentration of households with the age of the youngest child at slightly less than 16. Graph 3 shows a nonparametric estimation of the density function of the variable 'gap' restricted to the domain of −30 months to +30 months[21]. The hypothesis of manipulation would gain force if there is an increase in the mass of the distribution immediately to the right of zero. The results do not seem to indicate that there is a relatively higher discontinuity around the age defined as the program eligibility threshold[22]. If there is some variation around zero, it goes in the opposite direction, that is, a mass increase to the left of zero. We applied the test of manipulation proposed by McCrary (2008) as a robustness check for potential non-random sorting about the age of the youngest child. Our results indicate no statistical evidence of a discontinuity in the density function of the households according to the age of the youngest child at the cutoff.

That means that the evidence above does not corroborate the hypothesis of manipulation of the information about the age of the youngest child. Added to this evidence is the fact that the programme agent asks for the birth certificates of beneficiary residents.

In the second scenario considered above, we observed that only 0.09 per cent of households with children aged between 15 and 16 years have other children aged between 0 to 1 year[23].

As such, the discontinuity in the definition of eligibility based on the age of the youngest child will be explored in this article as being exogenous.

## 5 Main results

In this section we present three sets of results for the effect of BFP on labour input allocated to informality, shown in equation (7) by coefficient $\lambda_1$. The difference among these three sets relates to the different dimensions of informality we consider.

Table 3 shows the main results of this work. These are the estimated values referring to effect of Bolsa Família on the proportion of working hours of the household dedicated to informal activities. The tables follow the same pattern as the previous one, with sample restrictions related to household income varying across the columns and restrictions related to bandwidth size for the age of the youngest child varying across the rows.

**Table 3 Impact of BFP on the hours worked in the informal sector**

| | Total sample | | Sample I | | Sample II | |
|---|---|---|---|---|---|---|
| | All households receiving the PBF benefit | | Households with PCHIs < R$ 700.00 | | Households with PCHIs > R$ 50.00 and PCHIs < $ 700.00 | |
| | Coef. | Optimal Window | Coef. | Optimal Window | Coef. | Optimal Window |
| **Bolsa-Família** | −2.006 | 1.67 | −1.778 | 2.08 | −2.152 | 2.00 |
| (standard error) | (1.224) | | (1.028) | | (1.215) | |
| Bolsa-Família | | | | | | |
| **Bolsa-Família** (bandwidth - **110**) | −2.391 | 1.84 | −1.866 | 2.29 | −2.463 | 2.20 |
| (standard error) | (1.497) | | (1.78) | | (1.525) | |
| **Bolsa-Família** (bandwidth - **125**) | −1.965 | 2.09 | −2.633 | 2.60 | −2.072 | 2.50 |
| (standard error) | (1.134) | | (1.568) | | (1.327) | |
| **Bolsa-Família** (bandwidth - **150**) | −1.431 | 2.50 | −1.752 | 3.12 | −1.638 | 3.00 |
| (standard error) | (0.974) | | (1.425) | | (1.313) | |

Source: 2006 PNAD.

The point estimates of our parameters of interest are negative in all twelve situations reported in the table (3 samples × 4 window sizes). That is, if anything Bolsa-Familia shifts the supply of working hours to the formal sector. However, we cannot reject the hypothesis that all estimates may be zero at the 95 per cent confidence level (Table 3). That is, we were unable to identify any effect of Bolsa Família in the proportion of working hours of the household dedicated to informal activities[25].

One should note that magnitudes of point estimations are not small, however standard errors are also large. So it may be the case that BFP induces formal attachment to labour market but we just could not estimate it precisely. Assuming that this is the case one possible reasoning for the counter intuitive impact of BFP is that poor households may live far from formal job opportunities, and hence the money from BFP may be used to finance transportation cost required in applications for formal jobs.

We repeated the same exercise to estimate the effect of BFP on the probability of the main occupation held by the head of the household being informal[24]. In this case, we aim to identify some adjustment in the extensive margin of labor input as our previous results indicate no impact in the intensive margin. The results reported in Table 4 go in the same direction as those estimated for the household hours in Table 3. That is, although the point estimate is negative, the program has no impact on the propensity of household heads occupying informal jobs.

Note that the results for the extensive margin reported in Table 4 demand some extra caution in their interpretation. The sample in this case is restricted to occupied household heads, which means a selective group. Even though, we see as informative the robustness of the results shown in Table 3 for another exercise where both the measure of informality and the sample are different from the previous one.

Finally, as the last set of results we report in Table 5 the estimated values for the BFP effect on the probability of the secondary occupation of the household (which can be the secondary occupation of the household head or the main occupation of another household member) being informal. Once more, at the 95 per cent confidence level, we

**Table 4 Impact of BFP on the occupational choice of the household head**

| | Total sample | | Sample I | | Sample II | |
|---|---|---|---|---|---|---|
| | All households receiving the PBF benefit | | Households with PCHIs < R$ 700.00 | | Households with PCHIs > R$ 50.00 and PCHIs < $ 700.00 | |
| | Coef. | Optimal Window | Coef. | Optimal Window | Coef. | Optimal Window |
| **Bolsa-Família** | −3.237 | 1.61 | −2.417 | 1.78 | −2.429 | 1.76 |
| (standard error) | (3.329) | | (1.460) | (1.485) | | |
| Bolsa-Família | | | | | | |
| **Bolsa-Família** (bandwidth - **110**) | −2.339 | 1.77 | −2.846 | 1.96 | −2.629 | 1.94 |
| (standard error) | (1.612) | | (1.797) | | (1.580) | |
| **Bolsa-Família** (bandwidth - **125**) | −1.996 | 2.01 | −2.607 | 2.23 | −2.697 | 2.20 |
| (standard error) | (1.387) | | (1.875) | | (1.747) | |
| **Bolsa-Família** (bandwidth - **150**) | −1.122 | 2.41 | −2.539 | 2.67 | −2.648 | 2.64 |
| (standard error) | (0.973) | | (1.614) | | (1.563) | |

Source: 2006 PNAD.

cannot reject the hypothesis that all estimates may be zero (Table 5). That is, although the point estimate is negative, the program has no impact on secondary occupational choice of the household, between formal and informal jobs.

## 6 Conclusions

In this study, we sought to identify the extent to which Bolsa Família Program has changed either hours of work or occupational choices towards informal jobs. Unlike what was found in the existing empirical literature, our results suggest that the programme has no impact neither in the allocation of hours nor on the occupational choice of beneficiaries between formal and informal jobs.

**Table 5 Impact of BFP on the secondary occupational choice**

| | Total sample | | Sample I | | Sample II | |
|---|---|---|---|---|---|---|
| | All households receiving the PBF benefit | | Households with PCHIs < R$ 700.00 | | Households with PCHIs > R$ 50.00 and PCHIs < $ 700.00 | |
| | Coef. | Optimal window | Coef. | Optimal window | Coef. | Optimal window |
| **Bolsa-Família** | −1.196 | 1.74 | −1.997 | 1.81 | −1.904 | 1.91 |
| (standard error) | (1.339) | | (1.355) | | (1.347) | |
| Bolsa-Família | | | | | | |
| **Bolsa-Família** (bandwidth - **110**) | −1.614 | 1.92 | −1970 | 1.99 | −1.626 | 2.10 |
| (standard error) | (1.563) | | (1.366) | | (0.983) | |
| **Bolsa-Família** (bandwidth - **125**) | −1.304 | 2.18 | −1.347 | 2.27 | −1.301 | 2.38 |
| (standard error) | (1.627) | | (1.236) | | (1.208) | |
| **Bolsa-Família** (bandwidth - **150**) | −1.107 | 2.62 | | 2.72 | −0.854 | 2.86 |
| (standard error) | (1.023) | | (1.235) | | (0.956) | |

Source: 2006 PNAD.

It is worth mentioning that this result is solid in a wide range of situations that we have considered. Twelve separate household samples were used to estimate our parameter of interest. In addition, for each sample we estimate the effect of the programme both in the allocation of working hours and occupational choice, regarding both the main occupation of the household head and in the secondary occupation in the household. In all these situations, the estimate was statistically nil.

One of the major contributions of our article is the application of a method that allows us to deal with the problem of self-selection based on unobservable characteristics. We exploited a discontinuity present in the eligibility criteria of the programme, about the age of the youngest child. We assumed that the discontinuity around a critical age value (16 years old) represents an exogenous variation in eligibility that could be accompanied by a variation in programme participation also around this critical value. This last variation would, in turn, identify the desired effect on the occupational choice of the household heads.

Finally, it is worth mentioning two additional observations. On the one hand, two of the four hypotheses behind our method are testable. We presented evidence in favour of both testable hypotheses, which puts us in a relatively comfortable position about the suitability of the method to our context. On the other hand, the method restricts the identification of Bolsa Família's effect on the group of households with at least one child and whose youngest (or only) child is around 16 on 31/12/2005. Nothing prevents the effect from being different from what we reported for families with younger children.

## Endnotes

[1]In this regard, see Chapter 4 of Fiszbein and Schady (2009) for a review of the effects of CCT programs on the labour supply.

[2]In principle the result on wages could be driven by selection of lower ability workers in this sector. However studies using longitudinal data and (individuals) fixed effect estimators confirm lower wages to informal jobs.

[3]See Imbens and Lemieux 2008.

[4]While the first study analyses Progresa, in Mexico, others analyse BFP.

[5]Teixeira (2010) as well as Marino and Mendes (2013) also conduct this type of analysis. However, they rule out, by assumption, the influence of unobservable factors in the selection of BFP beneficiaries.

[6]For instance, threshold values for earnings may vary in real terms if they are fixed in nominal terms.

[7]In the case of child health, prenatal testing for pregnant women and medical care for nursing mothers between 14 and 44 years are required, in addition to paediatric monitoring of the growth and development of children aged up to six years. In the case of education, the conditionalities are associated with the access to - and maintenance of - children and youth in school.

[8]Two other benefits were incorporated into BFP after 2006: the Youth Variable Benefit and the Benefit to Overcome Extreme Poverty (BSP).

[9]In practice, the municipal agents identify potential beneficiaries and can track the status of BFP beneficiary households in databases containing formal income sources,

such as the Continuous Cash Benefit (BPC) and Annual Report of Social Information (RAIS) databases.

[10]All the information regarding the management of BFP are from Soares and Sátyro (2009).

[11]MDS is the Brazilian Ministry of Social Development and Fight against Hunger. It was created in 2004 with the goal of promoting the social inclusion, food and nutrition security, full social assistance and a minimum citizen income to the families living in the poverty (http:///www.mds.gov.br). CEF is the Federal Savings and Loans Bank.

[12]We used the age on 31/12/2005 because on this date we are able to identify households whose youngest child's age has recently crossed the border of eligibility (c), making those households ineligible in 2006, as well as households that remained eligible in 2006 because the age of the youngest child only crosses the border immediately after the turn of the year. These households will be crucial in our identification strategy, as will be detailed later.

[13]Therefore, we can identify a causal relationship according to the definition usually applied by economists, in which the causality between variables X and Y can be identified if we observe the response of Y to a change in X, *keeping everything else constant.*

[14]The term "fuzzy" refers to the relationship between participation and eligibility, which is not deterministic in the case of acute regression discontinuity.

[15]For the sake of simplifying notation we suppress X from the conditioning part of the expectation operators from now on.

[16]There is a discussion in the literature about how interesting is the identification of a valid parameter for such a restricted group (see Heckman [1997]). However, in our case, it can be argued that it is of interest to identify this group in view of the change in the eligibility criteria that occurred in July 2008, which now considers eligible households with young children up to 17 years.

[17]The work of Imbens and Angrist (1994), which coined the term LATE, presents a more technical discussion about the identification of this parameter. The book by Angrist and Pischke (2009) provides a more intuitive approach to the same topic.

[18]The question is: "In September 2006, did a resident of this household receive money from the social program Bolsa Familia?" Followed these possible answers: 1) Yes; 2) No.

[19]We classified as employed during the reference week individuals who executed any paid work during it, or the ones who executed an unpaid work in that week for at least fifteen hours, or who have paid jobs of which they are temporarily away. It is not considered here as occupied individuals who executed any work for their own consumption or construction during the reference week.

[20]Usually, the "always taker" and "never taker" groups are not easily identifiable and, therefore, this procedure of restricting the sample is unusual. Note that we are taking advantage of the two-dimensional nature of BFP's eligibility rule (age composition and PCHI) to identify the "never-takers".

[21]The estimate was achieved through the Kernel method with Epanechnikov weighting.

[22]It can be argued that there is more interest, on the part of the households, to manipulate such information in front of a programme agent than in front of an IBGE interviewer. However, since the 2006 PNAD collected information on BFP and IBGE is a government agency, it is possible that a household prone to manipulating information in front of programme agents showed the same likelihood to do so in front of an IBGE interviewer.

[23]Hypothetically, there would be yet another kind of attitude that could compromise the exogenous character that we attribute to the criterion based on the age of the youngest child, which is changing the age at which children leave home. In particular, it would be a problem for our analysis if the children left home below the age limit for eligibility in an environment without Bolsa Familia and had postponed this move as BFP beneficiary to maximize the period during which they receive the benefit. There is no information available to check this hypothesis.

[24]This was estimated as a linear probability model even with a binary dependent variable. The parameter identification and interpretation would complicate considerably for a limited dependent model.

[25]Results are obtained from local linear regressions. We include higher order polynomials for robustness checks and it does not substantially affect our results.

**Competing interests**
The IZA Journal of Labor & Development is committed to the IZA Guiding Principles of Research Integrity. Both authors declare that they have observed these principles.

**Acknowledgements**
The authors thank Sergei Soares for several explanations on the institutional aspects of Bolsa Familia. We benefited from valuable comments from the editor and an anonymous referee, as well as from our colleagues Pedro Herculano de Souza, Miguel Foguel and Fabio Soares. We also benefited from comments of the participants in the 2014 edition of IZA-World Bank labor and development seminar in Perú, in the 2013 edition of Itaú International Seminar of Impact Evaluation, and a seminar at the Economic Department of PUC-Rio. Finally, we are grateful to Italo Cabral de Souza for the assistance in processing the 2006 PNAD.
Responsible editor: David Lam

Received: 15 October 2014 Accepted: 21 November 2014
Published online: 31 December 2014

**References**

Angrist JD, Pischke JS (2009) Mostly Harmless Econometrics. Princeton, Princeton University Press
Avaliação de Impacto do Programa Bolsa Família – 1[A] Rodada (AIBF I). Secretaria de Avaliação e Gestão da Informação. Brasília: 2012. Available at: http://www.mds.gov.br
Braw A, Giilgan D, Hoddinotti J, Roy S (2012) Bolsa Família and Household Labour Supply. International Food Policy Research Institute
Cruces G, Ham A, Viollaz M (2012) Scarring effects of youth unemployment and informality: Evidence from Argentina and Brazil, CEDLAS
Fiszbein A, Schady NR (2009) Conditional Cash Transfers: Reducing present and future poverty. The World Bank, Washington, D.C.
Heckman J (1997) Instrumental variables: a study of implicit behavioral assumptions used in making programme evaluations. J Hum Resour 32:441–462
Imbens G, Angrist J (1994) Identification and estimation of local average treatment effects. Econometrica 62:467–476
Imbens G, Kalyanaraman K (2012) Optimal bandwidth choice for the regression discontinuity estimator. Rev Econ Stud 79(3):933–959
Imbens G, Lemieux T (2008) Regression discontinuity designs: a guide to practice. J Econometrics 142(2):615–635, Elsevier
Lee DS, Lemieux T (2010) Regression discontinuity designs in economics. J Econ Lit 48(2):281–355
Marino E, Mendes S (2013) The Impact of Government Income Transfers on the Brazilian Job Market, vol n.1, 43rd edn. Estudos Econômicos, São Paulo, pp 29–50
McCrary J (2008) Manipulation of the running variable in the regression discontinuity design: a density test. J Econometrics 142(2):698–714, Elsevier
Ribas RP, Soares FV (2011) Is the effect of conditional transfers on labour supply negligible everywhere? Available at: http://www.iza.org/conference_files/worldb2011/ribas_r6802.pdf
Skoufias E, Di Maro V (2008) Conditional cash transfers, adult work and poverty. J Dev Stud 44(7):935–960
Soares S, Sátyro N (2009) O programa Bolsa Família: desenho institucional, impactos e possibilidades futuras. Texto para discussão n° 1424. Brasília, IPEA
Teixeira CG (2010) A Heterogeneity Analysis of the Bolsa Família Programme Effect on Men and Women's Work Supply, IPC *Working Paper* no 61. IPC, Brasília, Available at: http://www.ipc-undp.org/pub/IPCWorkingPaper61.pdf