**Longitudinal Data Impact Analysis of the Brazilian
'Bolsa Família' Social Welfare Programme**

Aline Araújo Barbosa de Castro (alineabcastro@hotmail.com)
Carolina Ribeiro Antonio (carolinaribeiroantonio@hotmail.com)
Cinara de Jesus Santos (cinara.ufjf@gmail.com)
Tamires Mara Coelho (tamires_mcoelho@hotmail.com) *
Walmir dos Reis Miranda Filho (walmir.reis.miranda.filho@gmail.com)
Marcel de Toledo Vieira (marcel.vieira@ufjf.edu.br)
Ricardo da Silva Freguglia (ricardo.freguglia@ufjf.edu.br)

Universidade Federal de Juiz de Fora, Juiz de Fora, Brazil

Abstract

The 'Bolsa Família' Program (PBF), created in 2003 by the Brazilian government is a direct income transfer program that benefits poor and extremely poor families, through the commitment of health and education conditionalities. In order to ascertain what impacts were generated by the PBF in the living conditions of the beneficiaries, the 'Bolsa Família' Impact Assessment Survey was carried out (AIBF). The AIBF was a complex sampling longitudinal study that contrasted beneficiary families and non-beneficiaries of the program. Data collection was conducted in two rounds (2005 and 2009), and three observation groups were defined: BFP beneficiaries (Treatment group), registered in other social programs from the federal government, but non-BFP-beneficiaries (Control 1 group), and not registered in any social programs (Control 2 group). The main aim of this work is to evaluate the effectiveness of PBF with regard to health, using data collected in the two rounds of AIBF survey using the econometric method of difference in differences. Two types of dependent variables were analysed in this paper: (i) binary ones, which were analysed through logistic regression models; and (ii) continuous ones to which were fitted linear regression models. The explanatory variables considered were the following: the observation group to which the individual belonged; the time and the interaction between the two mentioned variables. Some of the analysed dependent variables, such as (a) existence of expenses with occasional medicines and (b) with health insurance for people over 14 years of age, and the amount spent on each of these costs have shown significant differences both between groups, over time and for the interaction term suggesting possible positive impacts on the health of the BFP beneficiaries. Thus it is possible to draw public policy conclusions about the impact of the program with the analysis set up we have considered.

Keywords: conditional cash transfer program, panel data, complex sampling, difference in differences.

1. The ‘Bolsa Família’ Program and the AIBF Survey

Social inequality and poverty levels are some challenges that developing countries face. In Brazil, the inequality between social classes began to be fought with the creation of the first direct cash transfer programs from the 1990s. Examples of these programs are the ‘Bolsa Escola’, the Food Grant and Gas Aid programs.

In 2003, the government implemented the ‘Bolsa Família’ Program (BFP) to join the transfer programs already created and increasing the focus of its actions. The PBF is a direct cash transfer program with conditionalities in the areas of education and health, which benefits poor and extremely poor families. In order to apply for the program, the family must register at the Single Registry of Social Programs of the Federal Government's Ministry of Social Development and Fight against Hunger (MDS), which is a tool for data collection and management aimed at identifying families with low income in Brazil. Registration is a prerequisite that does not imply receiving the benefit. In order to be covered, the family must be selected by the MDS.

According to the ‘Caixa Econômica Federal’ (CEF), which is the state bank in charge of the benefits, nearly 14 million Brazilian families are beneficiaries of PBF. The benefit amount at the beginning of 2015 ranged from R$ 77.00 to R$ 336.00 per month (equivalent to approximately US$ 30 to US$ 120), not including the benefit for Overcoming Extreme Poverty, according to income and family size. To remain in the program, beneficiaries must fulfill the following conditionalities: keep their children from 6 to 15 years old enrolled in schools, ensuring school attendance of at least 85%; make medical care and prenatal care (for women of childbearing age); keep the vaccination cards of children updated.

In order to contrast PBF beneficiary and non-beneficiary families, and understand the impact in the living conditions, the government conducted a longitudinal study ‘Bolsa Família’ Impact Evaluation Survey (AIBF) which was performed in two rounds (2005 and 2009). The survey is representative of three areas in the country: North / Midwest; Northeast; and South / Southeast, classifying families selected the following groups: (i) the treatment group (T), composed of PBF beneficiary households; (ii) the comparison group 1 (C1), composed of households that are registered and that receive other social benefits other than the PBF; and (iii) the comparison group 2 (C2), composed of households that reported never having received any benefit. Interviews were carried out in 2005 on 15,426 families in 269 municipalities from 23 states and the Federal District. In 2009, the survey tried to search the original sample members but only 74.12% (11,433) of households interviewed in 2005 were found again. For both rounds the following information were collected: information about the household; living conditions; education; health and spending on various items.

Therefore, the aim of this paper is, by analyzing the data of the AIBF survey, evaluate the effectiveness of PBF with regard to health, using the econometric method of difference in differences.

2. Difference in Differences

According to Wooldridge (2013), the difference in differences estimation method is used in natural experiments, which are characterized primarily by exogenous events. In socio-economic studies, this event is usually the result of a policy change that transforms the environment in which survey units (people, companies, cities) act. In this work, the event in question is to receive benefits from income transfer programs, especially the PBF.

In addition, in experiments there is a control group, which is not affected by the event - in this analysis, individuals belonging to the C2 group - and a treatment group, which is allegedly affected by the event - for which were found only the T group, but also those belonging to C1 group, since many of these have received the benefit of PBF in the second round, assuming therefore that the impact of the program there for these two groups. For ethical reasons, the classification of individuals in the control group or treatment cannot be at random.

Finally, two points in time are required for a natural experiment - in this paper, the rounds of 2005 and 2009 - to compare the groups of control and treatment. Thus, the sample is generally divided into four groups: control group prior to the event; the control group after the event; treatment group prior to the event; and treatment group after the event. Although this study treatment group already received the PBF or other(s) benefit(s), the two-year interval between the program's inception (2003) and the first round of AIBF research (2005) was assumed to be insufficient for the production any impact on the health of individuals belonging to this group.

Let y be the dependent variable; dA be a dichotomous variable with values equal to zero if the individual belongs to the control group and zero if the individual belongs to the treatment group; dC be another dichotomous variable with values equal to zero if data was collected in 2005 and equal to one if the information was collected in 2009; and ε denotes the error term of the model. The estimation is obtained by fitting the following regression model:

$$y = \beta_0 + \beta_1 dA + \delta_0 dC + \delta_1 dAdC + \varepsilon \qquad (1)$$

Where β_0; β_1; δ_0 and δ_1 are the coefficients to be estimated which are associated, respectively, to: the constant term, variable dA; variable dC and the interaction of dA with dC. Note that δ_1 measures the changes effects – the impact of the PBF in the response variable – and its estimator $\hat{\delta}_1$ is obtained from:

$$\hat{\delta}_1 = \left(\bar{y}_{dC=1;dA=1} - \bar{y}_{dC=1;dA=0}\right) - \left(\bar{y}_{dC=0;dA=1} - \bar{y}_{dC=0;dA=0}\right) \qquad (2)$$

Where $\hat{\delta}_1$ is the estimator of the difference in differences. If y is a continuous variable, the estimation may be performed by ordinary least squares, when ignoring the sampling design. However, if y is a dichotomous variable, the relationship between y and the explanatory variables is non-linear. In such cases we adopt the logistic regression model.

The main dependent variables considered in the paper are health related variables: anthropometry of individuals; the existence of expenses related health plans, occasional use of medicines and hospitalization and the amount spent; and socioeconomic factors indirectly associated with the health of individuals (sanitary conditions of the household and free meal at school, for example).

3. Data Analysis and Concluding Remarks

The first analised variable refers to the occurrence of health plan expenditures for the residents older than 14 years of age, in the last 30 days. Through a regression model, we want to know which and how the explanatory variables influence the existence of spending with health insurance. The dependent variable was coded with the value one for each individual who had responded that made the refered spending the last 30 days, and zero otherwise. As y is a dichotomous variable in this case, we consider the logistic regression. Results are presented in Tables 1 and 2.

Table 1 – Logistic Model for the Occurrence of Spending with Health Plan Over the Last 30 Days for Residents Aged Over 14 Years Old (I)

Number of Observations	51,928
Likelihood Ratio Test Statistic	24,300.95
P-value	<0.0001
Pseudo R^2	0.5254

The model fits the data significantly better than a model with only the constant and has a good fit according to the calculated Pseudo R^2.

Table 2 – Logistic Model for the Occurrence of Spending with Health Plan
Over the Last 30 Days for Residents Aged Over 14 Years Old (II)

Coefficient	Estimate	St. Error	Z	P-value	Confidence Interval (95%)	
β_0	-1.781954	0.0523076	-34.07	<0.001	-1.884475	-1.679433
β_1	-1.677655	0.0597714	-28.07	<0.001	-1.794805	-1.560505
δ_0	3.333853	0.1255654	26.55	<0.001	3.08775	3.579957
δ_1	1.514243	0.13186	11.48	<0.001	1.255802	1.772684

All model coefficients are significantly different from zero. From calculating the odds ratios for the observed coefficients the following conclusions may be drawn for the health plan spending: individuals that belong to T and C1 groups in 2005, have 81% less chance of having spent on health plan; an individual belonging to the T group or the C1 group has 355% more chances of having had the considered spending in 30 days previous to data collection.

We also consider the dependent variable that refers to the amount spent on health plans for individuals over 15 years in the last 30 days. As this variable is continuous, we will apply the multiple linear regression model with estimation made by OLS, as we are currently ignoring the sampling design in our analysis. We have the following results for the model in Tables 3 and 4.

Table 3 – Multiple Linear Regression Model for the Value Spent with Health Plan
Over the Last 30 Days for Residents Aged Over 14 Years Old (I)

Number of Observations	2,931
Likelihood Ratio Test Statistic	279.73
P-value	<0.0001
R^2	0.2228

The model fits the data significantly better than a model with only the constant and has a good fit according to the calculated R^2.

Table 4 – Multiple Linear Regression Model for the Value Spent with Health Plan
Over the Last 30 Days for Residents Aged Over 14 Years Old (II)

Coefficient	Estimate	St. Error	z	P-value	Confidence Interval (95%)	
β_0	202.4885	6.093942	33.23	<0.001	190.5396	214.4373
β_1	-128.7025	7.070837	-18.20	<0.001	-142.5668	-114.8382
δ_0	-201.4459	14.34674	-14.04	<0.001	-229.5767	-173.3152
δ_1	128.7857	15.23687	8.45	<0.001	98.90961	158.6617

Similarly to the findings obtained for the previous case, all covariates are significant. Individuals belonging to T or C1 groups in 2005, spend R$ 128.70 less with health plans whereas individuals belonging to T or C1 group in 2009 spent R$ 128.79 more.

The variable analysed below is the occurrence of expenses with occasional use of medication for people over 14 years in the 30 days before data collection. This variable is also dichotomous and was coded with the value one for each individual who responded that have made spending the last 30 days, and zero otherwise. Results are presented in Tables 5 and 6.

Table 5 – Logistic Model for the Occurrence of Spending with Occasional Use Medicines
Over the Last 30 Days for Residents Aged Over 14 Years Old (I)

Number of Observations	86 570
Likelihood Ratio Test Statistic	235.86
P-value	<0.0001
Pseudo R^2	0.0024

The model fits the data significantly better than a model with only the constant but has a very low Pseudo R^2.

Table 6 – Logistic Model for the Occurrence of Spending with Occasional Use Medicines Over the Last 30 Days for Residents Aged Over 14 Years Old (II)

Coefficient	Estimate	St. Error	z	P-value	Confidence Interval (95%)	
β_0	-0.8734621	0.0402954	-21.68	<0.001	-0.9524396	-0.7944847
β_1	-0.1500729	0.0418432	-3.59	<0.001	-0.2320839	-0.0680618
δ_0	-0.4712732	0.0608037	-7.75	<0.001	-0.5904462	-0.3521003
δ_1	0.2549308	0.0629846	4.05	<0.001	0.1314832	0.3783783

Again, all covariates are significant at the 5% significance level. Calculating the odds ratios for the observed coefficients, the following conclusions are drawn for the occurrence of the expenses with occasional use of medicines: the chance of a person belonging to T or C1 group in 2005 to have made any spending is 13.94% lower; and the chance of an individual, in 2009, belonging to the T group or the C1 group have had any expenditure with occasional medicines 29.03% higher.

Finally, we analyse the amount spent in Brazilian Reals, with occasional use of medicines for residents aged 14 or older, in the 30 days previous to the data collection. Multiple linear regression model has been adopted as the response variable is assumed continuous. Results are presented in Tables 7 and 8.

Table 7 – Multiple Linear Regression Model for the Value Spent with Occasional Use Medicines Over the Last 30 Days for Residents Aged Over 14 Years Old (I)

Number of Observations	54,482
Likelihood Ratio Test Statistic	7,428.03
P-value	<0.0001
R^2	0.2903

The model fits the data significantly better than a model with only the constant and has a good fit according to the calculated R^2.

Table 8 – Multiple Linear Regression Model for the Value Spent with Occasional Use Medicines Over the Last 30 Days for Residents Aged Over 14 Years Old (II)

Coefficient	Estimate	St. Error	z	P-value	Confidence Interval (95%)	
β_0	48.12939	0.7381513	65.20	<0.001	46.68261	49.57617
β_1	-13.28288	0.767695	-17.30	<0.001	-14.78757	-11.77819
δ_0	-46.16618	0.8407616	-54.91	<0.001	-47.81408	-44.51828
δ_1	13.28857	0.8736508	15.21	<0.001	11.57621	15.00093

All coefficients are significant at the 5% level. Individuals that belong to T or C1 groups in 2005, spend about R$ 13.28 less with medicines of occasional use and individuals that belong to the T group or the C1 group spend on average R$ 13.00 more in 2009 compared to the groups resulting from other combinations.

Our results suggest some positive impacts of the PBF although our analysis certainly needs to be refined by allowing for complex sampling design of the AIBF survey in the estimation procedures and controlling for other covariates in the regression models.

Reference

Wooldridge, J. M. (2013) *Introductory Econometrics: A Modern Approach – Fifth Edition*. South-Western, Cengage Learning.