Analysis of the demographic dividend in school enrollment in Brazil using hierarchic and hierarchic-spatial approaches

Juliana de Lucena Ruas Riani*

1) INTRODUCTION

This article aims to investigate the determinants of lower and middle school enrollment in Brazil, bringing together two traditions of educational studies.

The first tradition comes from economic demography literature about the demographic dividend. The main contribution in this area is the classical debate between COALE & HOOVER (1958) and SCHULTZ (1987) about the impacts of fertility variation and age structure on education. According to the former, lower fertility results in a lower youth dependency ratio, reducing the number of children enrolled in schools. This decrease provokes an increase in government savings due to reduced education costs. However, if not all school-age children are enrolled, the demographic dividend may result in an increase in school coverage via a rate of enrollment effect. Schultz questions this relation. He assumes that the government budget is not flexible in the short run, *i.e.* government spending may not shrink at the same rate as the reduced school-age cohort, possibly resulting in improved quality of education, by means of higher salaries for teachers or smaller class sizes. Most studies testing the hypothesis of the demographic dividend in education use aggregate data, with no control over the family environment in which the student is inserted.

The second tradition is related to studies about determinants of education which uses the Educational Production Function (EPF) theoretical framework. EPF analyzes how the various inputs of the educational process can affect educational indicators and be specified in the following manner[1]: $A = f(F, S, O, \varepsilon)$.

Where the educational response (A) is a function of vectors: of characteristics and background of the individual (F); of school and teacher inputs (S); of other relevant inputs, *e.g.* community factors (O) and of the random error term which reflects factors which were not measured and contribute to the educational result (ε).

This article seeks to evaluate the extent to which profound demographic changes which took place in Brazil with the significant drop in Total Fertility Rate and consequent reduced youth dependency ratio can result in a demographic bonus for the Brazilian education system, opening “windows of opportunities”. Nonetheless, this analysis also takes into consideration, in addition to demographic variables, other factors such as family background, restricted availability and other factors related to the profile of the educational system in municipalities.

Two different methodologies were used in this study: <u>hierarchic</u> and <u>hierarchic-spatial</u>. The former is important because it incorporates the hierarchic data structure,

* State Secretariat of Education of Minas Gerais – Brazil and Universidade de Itaúna

[1] For further details about the Educational Production Function, see HANUSHEK, E. A. (2002), HANUSHEK, GOMES-NETO & HARBISON (1996), LEE & BARRO (1992), LAZEAR, E. P. (1999), TODD, P. E. & WOLPIN, K. (2003) and KRUEGER, A. B. (1999).

improving parameter estimation. The latter, in turn, includes spatial effects in the analysis, resulting from the use of aggregate data by municipality and makes an important methodological contribution to educational studies.

2) Hierarchic Analysis of Enrollment Determinants

2.1) Model Building

In order to analyze the demographic dividend, as well as other family and context factors affecting school enrollment, the EPF theoretical framework was used as a starting point. Nevertheless, some modifications were made to take into account data structure and availability.

The first adjustment was use of hierarchic models in two levels for EPF estimation. The first level (level-1) contemplates individual data, while the second (level-2) contemplates municipality data. The advantage of EPF estimation using hierarchic rather than classic models is that in the latter case results can be biased because errors may not be independent and identically distributed (i.i.d.). Furthermore, these models solve the problem of units of analysis, which have independent variables at distinct levels of aggregation.

In an aggregated analysis of Brazil, it is very likely that these i.i.d. assumptions of the standard OLS estimation do not apply, since the relation between individual characteristics and their school attendance may vary in different municipalities or when one considers that there is dependence between school attendance of individuals covered within the same municipality. This could be due to varying availability of education.

Hierarchic model building considers that the intercept and/or slope are not the same for all level-2 units and that variation can occur due to the effect of some level-2 explanatory variable and/or due to the random component. During formalization of hierarchic models, each level is represented by a set of equations. Thus, the equation at the first level is described as follows:

$$\ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right)_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + e_{ij} \qquad (1)$$

Assuming that the level-2 variability occurs only in the intercept, second level equations are described as:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_j + u_{0j} \qquad (2)$$

$$\beta_{1j} = \gamma_{10} \qquad (3)$$

Where:

i = 1, 2, ..., nj level-1 units, which in this case are the individuals;

j = 1, 2, ..., J level-2 units, which are the municipalities;

β_{0j} is the intercept, assumed as random;

β_{1j} is the coefficient for the independent variable of level-1;

γ_{00}, γ_{01} and γ_{10} are the parameters for the fixed part of the model, common to all individuals;

W_j the vector of independent variables measured at level-2;
u_{oj} and e_{ij} are the random effects associated with level-2 and level-1, respectively.

u_{0j} and e_{ij} are assumed to be independent with normal distribution, mean equal to zero and constant variances σ^2_{uo} and σ^2_e.

The hierarchic logistic model was used to model probabilities of attending lower and middle school for each individual in the age group 7 to 14 years and 15 to 17 years. The dependent variable represented the attendance of each individual of each group; the value "1" of the variable was for attending school and "0" otherwise.

Among independent variables, those related to characteristics of individuals and their family environment are found in the level-1 of the hierarchic model. Individual characteristics, sex, skin color and rural/urban residence, were used as control variables. Variables representing the family environment are extremely important in determining school attendance and the chosen ones are the most commonly used in studies about educational determinants (SILVA & HASENBALG, 2001; RIANI, 2005; LAM & BARROS, 1995; RIOS-NETO, CESAR & RIANI, 2001; and BARROS *et alli*, 2001), which have proven impact on the individual's school performance. They are: i) level of education of the mother; ii) category of occupation of family head[2]; iii) female head of family; iiii) co-resident.

Regarding level-2 independent variables, they reflected demographic aspects and availability of education in the municipality. The first would capture possible demographic gains brought about by the demographic transition which has been under way in Brazil for the past decades and the second would evaluate the extent to which supply of educational services in each municipality interferes in access to school and is related to the profiles of the schools in the municipalities. Availability of teachers and concentration of students in large schools are important dimensions of the supply of education. As for the latter, a recent discussion of the Brazilian educational policy by the Ministry of Education (*Ministério da Educação* – MEC) is about nucleation of schools, seeking to agglomerate small schools in a single area, making government attention more effective, thus facilitating investments in infra-structure.

The demographic variables were: 1) the relative size of the cohort in a school age appropriate for attending a certain level of education o; 2) the size of the population in the municipality was included as a control variable for regressions. The supply of education variables were: 1) the ratio of number of teachers in each level of education to the population in appropriate age for attending that level of education; 2) number of establishments with 50 to 200 enrolled students; 3) number of establishments with 201 to 1000 enrolled students and 4) number of establishments with over 1000 enrolled students. The last three variables represent the nucleation of schools.

[2] Socio-occupational categories were constructed for the head of the family, taking in to consideration the educational level required for each occupation, the type of specialization and its functions and level of income, stratified into three occupational levels, which are: superior and managerial, middle and manual. This aggregation was proposed by SILVA (1973 and 1985), who seeks homogeneity in occupational categories. In order to avoid loss of information about families whose heads have no occupation, a fourth category was created for inactive individuals, *i.e.* unemployed or retired. The category of reference for regressions was the manual level.

Although EPF takes other educational inputs into consideration, such as quality of school and teacher infra-structure, in order for children and youth to have access to the education system, it is likely that only changes at the level of demand for and availability of education are the most influential factors. At the micro level, family background variables influence demand, *e.g.* a mother with a higher education level will have greater awareness of the importance of school for children.

At the macro level, the relative size of the cohort in school age and restricted availability are the most important factors causing changes in demand and availability of education, since they can make finding openings in schools difficult, particularly public schools. It was thus decided that other school factors would not be included. In fact, previous studies demonstrate that quality of school infra-structure and quality of teachers have no impacts on school attendance[3].

The 2000 Demographic Census provided by IBGE was the data source, with the exception of the number of teachers at each level of education, which had the INEP School Census of 2000 as the source.

2.2) Results

Five models were estimated for access to lower school education by an individual, as shown in Table 1 – lower school – and Table 2 – middle school. In model 1, only first level variables were included, allowing for measurement of non-conditional variability of level-2. In models 2, 3 and 4 the following variables were included, respectively: relative school-age cohort size, ratio of teachers to school-age population and number of establishments with a given number of enrolled students (nucleation variables).

The reason for separation of more relevant level-2 variables is the possibility of verification of the extent to which each contributes to reduced non-conditional variability of the intercept estimated in model 1, which can be obtained through calculation of the explained variance, according to the following formula:

$$\%\text{ of Explained Variance} = \frac{\hat{\tau}_{qq(non-conditioned)} - \hat{\tau}_{qq(conditioned)}}{\hat{\tau}_{qq(non-conditioned)}} \quad (4)$$

Lastly, model 5 – the final model – incorporates all variables included in previous models, in addition to categorical variables for population size, as a control.

The results for lower school are commented first (Table 1). Analysis of the random effect of model 1 (lower part of the table) shows that the hypothesis of the random-effect intercept is acceptable, since it was significant. In other words, it is acceptable that municipalities present distinct values for the probability of attending school for children between 7 and 14 years of age. Models 2-4 show that relative cohort size is the variable which best explains intercept variability (39.70%), while the ratio of teachers to school-age population is the one which explains least (7.54%).

Proceeding to analysis of fixed effects, level-1 variables were shown to be significant for practically all cases and in the expected direction. Results for individual

[3] See RIANI, 2005.

attributes, age, sex, skin color and rural/urban residence, showed that age has a negative relation with school attendance, and that women, white and yellow people and those living in urban areas are more likely to attend school.

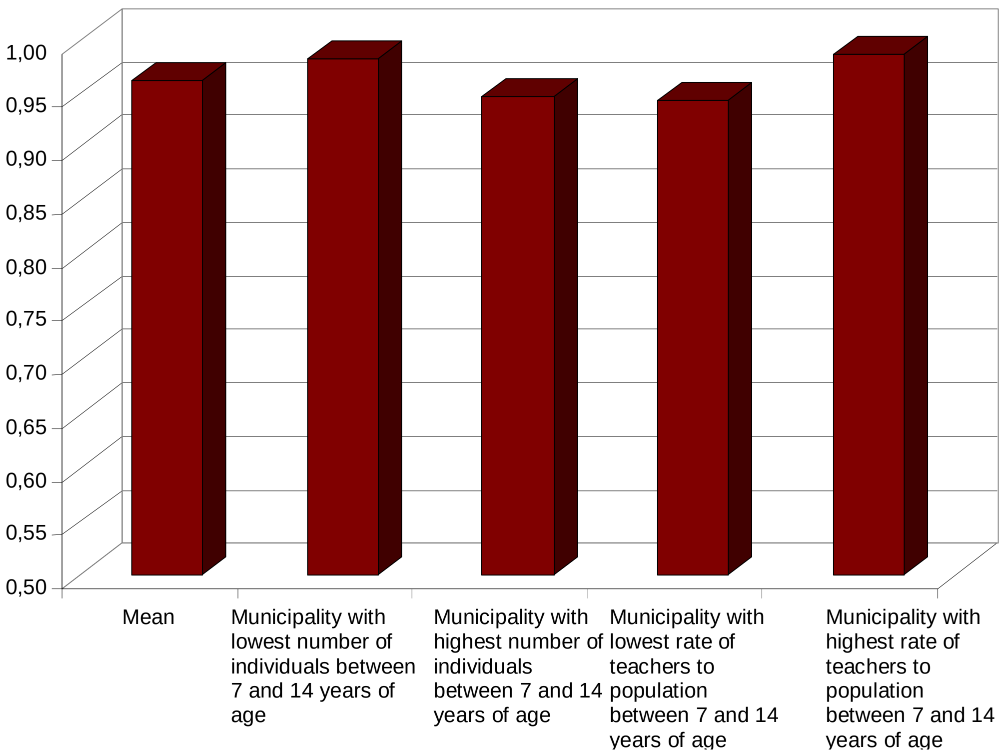
TABLE 1: Results of regression for school attendance probability – individuals between 7 and 14 years of age

Independent variables	Model 1		Model 2		Model 3		Model 4		Model 5	
Fixed Effect	Coefficient	Sig	Coefficient	Sig	Coefficient	Sig	Coefficient	Sig	Coefficient	Sig
Intercept	3,202	0,000	6,349	0,000	2,336	0,000	2,562	0,000	4,120	0,000
Relative cohort size – 7 to 14 years of age			-18,005	0,000					-11,765	0,000
Ratio of teachers to population between 7 and 14 years of age					13,844	0,000			10,845	0,000
Number of establishments with between 50 and 200 enrolled students							0,543	0,000	0,607	0,000
Number of establishments with between 201 and 1000 enrolled students							1,636	0,000	1,111	0,000
Number of establishments with up to 1000 enrolled students							0,723	0,000	0,924	0,000
Population up to 4999									-0,010	0,868
Population between 5000 and 9999									-0,028	0,648
Population between 10000 and 19999									-0,051	0,399
Population between 20000 and 49999									-0,103	0,087
Population between 50000 and 99999									-0,003	0,959
Population between 100000 and 199999									0,019	0,793
Age	-0,024	0,000	-0,024	0,000	-0,024	0,000	-0,024	0,000	-0,024	0,000
Urban/Rural residence; 1=urban and 0=rural	0,613	0,000	0,612	0,000	0,614	0,000	0,611	0,000	0,611	0,000
Skin color; 1=white and yellow and 0=black and brown	0,138	0,000	0,139	0,000	0,138	0,000	0,139	0,000	0,139	0,000
Sex; 1=male and 0=female	-0,199	0,000	-0,199	0,000	-0,199	0,000	-0,199	0,000	-0,199	0,000
Level of education – mother	0,177	0,000	0,178	0,000	0,177	0,000	0,178	0,000	0,178	0,000
Head of household is a woman	-0,228	0,000	-0,228	0,000	-0,228	0,000	-0,228	0,000	-0,229	0,000
Co-resident family	-0,216	0,000	-0,217	0,000	-0,216	0,000	-0,216	0,000	-0,217	0,000
Category of occupation of head of household – higher level	0,297	0,000	0,308	0,000	0,297	0,000	0,302	0,000	0,310	0,000
Category of occupation of head of household – middle level	0,013	0,136	0,014	0,091	0,012	0,141	0,014	0,102	0,015	0,081
Head of household has no occupation	-0,190	0,000	-0,188	0,000	-0,190	0,000	-0,189	0,000	-0,187	0,000
Random Effect										
Coefficient	0,581	0,000	0,350	0,000	0,537	0,000	0,427	0,000	0,287	0,000
% of variance explained			39,70		7,54		26,43		50,65	

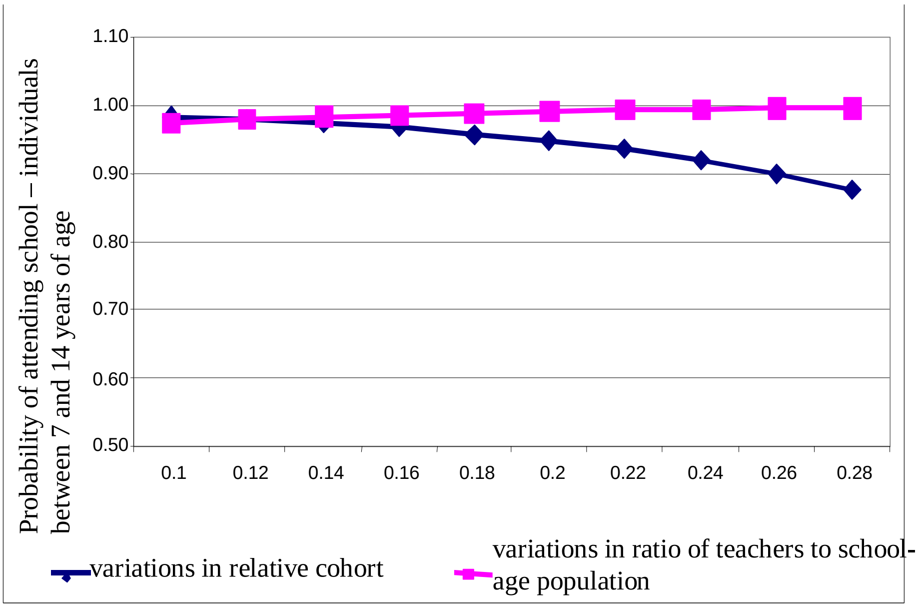
Regarding family background factors, it was observed that the mother's level of education is more important than the remaining variables, which is in accordance with results of previously discussed studies. For a better picture of the impact these factors have, Table 3 contains variations in the predicted value according to changes in family background factors, using coefficients from model 5. Increase of one standard deviation from the mean number of years of education for the mother results in an increase of 1.97% in the probability of children between 7 and 14 years of age to attend school, while the effect of other variables is lower. It should be noted that a slight increase in this probability, at this level of education, is due to the fact that it is practically universal, *i.e.* close to 100%.

Analysis of level-2 variables shows that lower demographic pressure resulting from a lower youth dependency ratio is one of the most important context factors determining access of children between 7 and 14 years of age to school. Smaller cohorts in this age group provoke an increase in probability of attending school at a rate close to what was verified in the simulation of variation in the mother's education. Restricted availability, on the other hand, although significant in its coefficient, has a smaller impact. Graphs 1 and 2 illustrate the impact of variations in school attendance probability for individuals between 7 and 14 years of age resulting from variations in the relative size of the cohort and restricted availability. From these, the higher impact of the youth dependency ratio can be more clearly seen, evidence that access by children between 7 and 14 years of age to school is improved by a higher demographic dividend, resulting from reduced fertility.

GRAPH 1: Impact of demographic pressure and restricted availability on probability of attending school for individuals between 7 and 14 years of age, considering municipalities with lowest and highest values.



GRAPH 2: Impact of demographic pressure and restricted availability on probability of attending school for individuals between 7 and 14 years of age



Results for individuals between 15 and 17 years of age can be found in Table 2. These results show that the hypothesis of considering the intercept with random effect is also acceptable for this level of education. When level-2 variables are added, contrarily to the previous case, restricted availability is the factor which best explains intercept variability. Since nucleation variables are not significant for this probability, another regression model which does not take them into account was estimated (Model 6).

Regarding level-1 variables, results closely resemble those found for individuals between 7 and 14 years of age, with the following differences: i) stronger negative impact of age, sex and families headed by women; ii) position variables for occupation of head of family have higher impact, be it more positive for those with higher socioeconomic status, or more negative for heads of families not participating in the work force; iii) families living together had a positive impact, opposite of the result for the previous case.

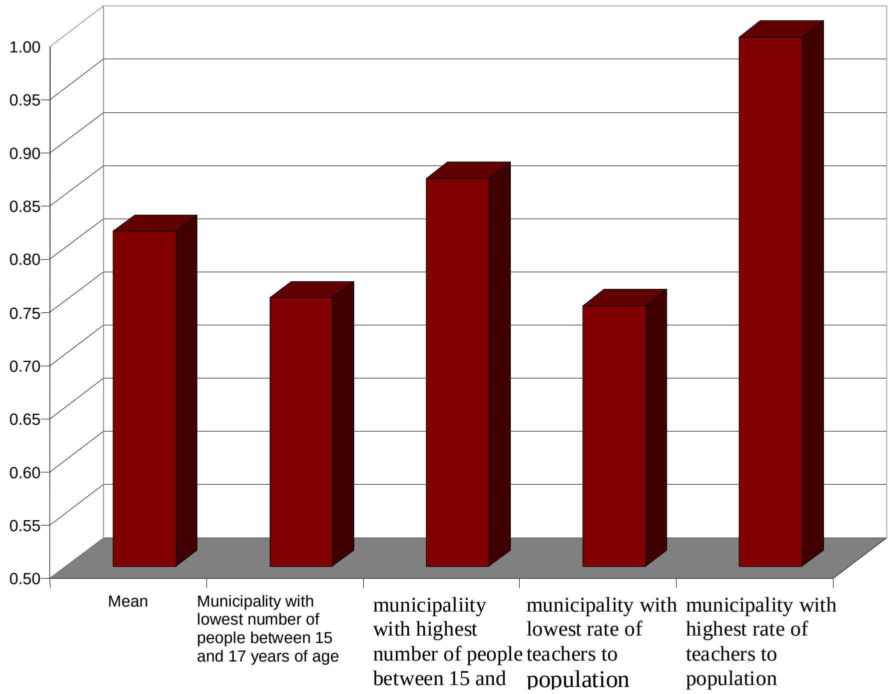
Table 3 shows that in middle school, variation resulting from simulations[4] of changes in level-1 variables is much higher than in lower school, especially with respect to the role of maternal education. This is due to a lower rate of people attending school in the sample for ages 15 to 17, 81% compared to 95% in the sample for ages 7 to 14.

Moving on to the analysis of municipal variables, it is noticeable that for this level of education, restricted availability has an important positive impact. When considering the municipality with highest number of middle school teachers according to population between 15 and 17 years of age, the probability of attending school approaches 100%, as shown in Graph 3. Also worthy of note is the positive sign for the relative school-age cohort size, an unexpected result. However, when this level-2 variable is taken into consideration in isolation in the analysis (Model 2), its sign becomes negative. Other studies have shown that this sign becomes positive when restricted availability is included in the regression. A possible explanation is the low level of coverage for this age group, far

[4] Simulations for middle school were carried out considering Model 6 coefficients.

from saturation, making an increase in demand, when controlled by restricted availability, have a positive effect.

GRAPH 3: Impact of demographic pressure and restricted availability on probability of attending school for individuals between 15 and 17 years of age, considering municipalities with lowest and highest values.



GRAPH 4: Impact of demographic pressure and restricted availability on the probability of attending school for individuals between 15 and 17 years of age

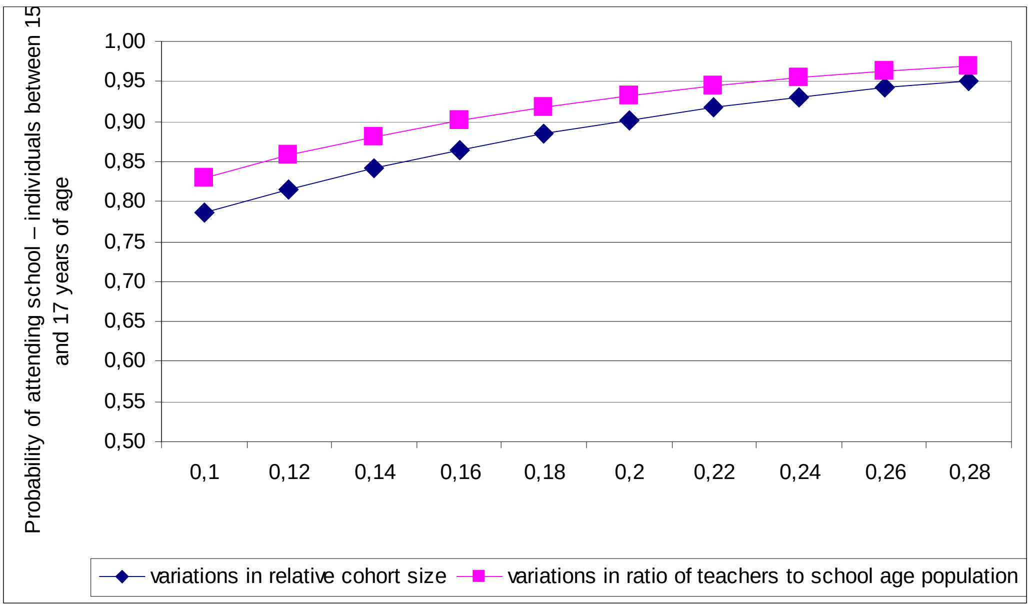


TABLE 2: Results of regression for probability of attending school – individuals between 15 and 17 years of age

Independent variables	Model 1		Model 2		Model 3		Model 4		Model 5		Mode
Fixed Effect	Coefficient	Sig	Coefficient	Sig	Coefficient	Sig	Coefficient	Sig	Coefficient	Sig	Coefficient
Intercept	1,440	0,000	1,912	0,000	1,141	0,000	1,518	0,000	1,000	0,000	1,113
Relative cohort size – 15 to 17 years of age			-7,081	0,000					9,421	0,000	9,241
Ratio of teachers to population between 15 and 17 years of age					8,115	0,000			10,310	0,000	10,285
Number of establishments with between 50 and 200 enrolled students							-0,136	0,016	-0,015	0,783	
Number of establishments with between 201 and 1000 enrolled students							-0,098	0,081	0,081	0,119	
Number of establishments with up to 1000 enrolled students							0,637	0,000	0,416	0,000	
Population up to 4999									-0,693	0,000	-0,788
Population between 5000 and 9999									-0,725	0,000	-0,781
Population between 10000 and 19999									-0,665	0,000	-0,711
Population between 20000 and 49999									-0,609	0,000	-0,635
Population between 50000 and 99999									-0,406	0,000	-0,413
Population between 100000 and 199999									-0,191	0,003	-0,197
Age	-0,441	0,000	-0,441	0,000	-0,441	0,000	-0,441	0,000	-0,441	0,000	-0,441
Urban/Rural residence; 1=urban and 0=rural	0,470	0,000	0,470	0,000	0,469	0,000	0,470	0,000	0,468	0,000	0,468
Skin color; 1=white and yellow and 0=black and brown	0,162	0,000	0,162	0,000	0,162	0,000	0,162	0,000	0,162	0,000	0,162
Sex; 1=male and 0=female	-0,445	0,000	-0,445	0,000	-0,445	0,000	-0,445	0,000	-0,444	0,000	-0,444
Level of education – mother	0,158	0,000	0,158	0,000	0,158	0,000	0,158	0,000	0,158	0,000	0,158
Head of household is a woman	-0,362	0,000	-0,362	0,000	-0,362	0,000	-0,362	0,000	-0,362	0,000	-0,362
Co-resident family	0,069	0,003	0,069	0,003	0,069	0,003	0,069	0,003	0,069	0,003	0,069
Category of occupation of head of household – higher level	0,479	0,000	0,481	0,000	0,483	0,000	0,479	0,000	0,485	0,000	0,485
Category of occupation of head of household – middle level	0,141	0,000	0,141	0,000	0,141	0,000	0,141	0,000	0,141	0,000	0,141
Head of household has no occupation	-0,048	0,000	-0,048	0,000	-0,047	0,000	-0,048	0,000	-0,047	0,000	-0,047
Random Effect											
Coefficient	0,294	0,000	0,290	0,000	0,262	0,000	0,282	0,000	0,220	0,000	0,223
% of variance explained			1,43		10,88		3,93		25,03		24,08

TABLE 3: Variations in probability of attending school according to family background variables for different samples*

Independent variables	Individuals between 7 and 14 years of age		Individuals between 15 and 17 years of age	
	Mean plus standard deviation	Mean minus standard deviation	Mean plus standard deviation	Mean minus standard deviation
Level of education – mother	1,97	-3,83	9,75	-14,59
Head of household is a woman	-0,34	0,31	-2,88	2,62
Co-resident family	-0,15	0,14	0,15	-0,15
Category of occupation of head of household –	0,28	-0,30	2,24	-2,43
Category of occupation of head of household –	0,02	-0,02	0,95	-0,98
Head of household has no occupation	-0,33	0,30	-0,40	0,40

Note: * Model 5 coefficients were used for the 7 to 14 years of age sample and Model 6 coefficients for the 15 to 17 years of age sample.

In summary, results show that lower demographic pressure, resulting from a lower youth dependency ratio, is one of the most important contextual factors for school attendance for children between 7 and 14 years of age. School attendance of children between 15 to 17 years of age, in turn, availability of education, represented by the ratio of teachers to school age population, is the most relevant municipal factor, demonstrating the need for public policy increasing availability of this level of education in Brazilian municipalities.

Furthermore, these results show the great influence of maternal education on attendance at both age groups, bearing witness to the inequality of opportunities in Brazil.

3) Hierarchic-spatial Analysis of Enrollment Determinants

3.1) Building the Model

Hierarchic models take spatial heterogeneity into account, while considering that there is a random effect connected with each level-2 unit, in this case municipalities. Nonetheless, spatial autocorrelation is not taken into account, since one of its assumptions is independence among effects of level-2 areas, *i.e.* errors at this level are independent and with constant variance. This assumption may not be realistic, considering that Brazil has high social and economic heterogeneity, which translates into large spatial heterogeneity; closer areas can be expected to be more similar than more distant areas, thus begetting macro level spatial correlation.

Autocorrelation and spatial heterogeneity are the spatial effects resulting from aggregate data. Spatial dependence takes place when variables observed in a certain location depend on values collected in neighboring localities and heterogeneity when error means and variances are not constant in space.

In other words, it can be said that aggregated variables of a municipality may be influenced by variables of neighboring municipalities, resulting in spatial autocorrelation. On the other hand, spatial units with differentiated behavior regarding their neighbors may exist, suggesting spatial heterogeneity. When these spatial effects are not corrected, coefficients of the hierarchic model may be inconsistent and inefficient.

In order to detect spatial autocorrelation, a diagnostic was carried out of independent level-2 variables in hierarchic models seen in the previous section, as well as rate of coverage, which represents the dependent variable in the hierarchic model. Nevertheless, since spatial analysis requires that measures be aggregated into geographical units, these are measured in terms of municipal rates instead of dichotomical ones, as in the previous model.

The diagnostic was performed by means of Moran[5] Statistic I, which is a tool for exploratory spatial data analysis. Results in Table 4 show that all variables have positive spatial autocorrelation, since Moran I was positive and significant in all cases. Thus, another EPF specification was made which attempts to include spatial effects in the analysis.

TABLE 4: Moran Statistic I

Variables	Moran I	Prob
Ratio enrollment for ages 7 to 14	0,579	0,000
Ratio enrollment for ages 15 to 17	0,392	0,000
Restricted availability		
Ratio of lower school teachers to school-age population	0,428	0,000
Ratio of middle school teachers to school-age population	0,439	0,000
Nucleation variables – Lower School		
Number of establishments with between 50 and 200 enrolled students	0,206	0,000
Number of establishments with between 201 and 1000 enrolled students	0,530	0,000
Number of establishments with more than 1000 enrolled students	0,215	0,000
Nucleation variables – Middle School		
Number of establishments with between 50 and 200 enrolled students	0,077	0,000
Number of establishments with between 201 and 1000 enrolled students	0,085	0,000
Number of establishments with more than 1000 enrolled students	0,100	0,000
Demographic dividend		
Relative cohort size for ages 7 to 14	0,811	0,000
Relative cohort size for ages 15 to 17	0,652	0,000
Population size		
Population between 0 and 4999	0,161	0,000
Population between 5000 and 9999	0,030	0,000
Population between 10000 and 19999	0,037	0,000
Population between 20000 and 49999	0,053	0,000
Population between 50000 and 99999	0,014	0,000
Population between 1000000 and 199999	0,052	0,000
Above 200000	0,193	0,000

5 Moran Statistic I can be found according to the following formula: $I = \frac{N}{S_0} \frac{\sum_i \sum_j W_{ij}(x_i - \mu)(x_j - \mu)}{\sum (x_i - \mu)^2}$ Where: N is the number of observations; W_{ij} is the W Matrix element corresponding to the i and j couple; x_i and x_j are the values of the variables in locations i and j; S_0 is the sum of all weights ($S_0 = \sum_i \sum_i w_{ij}$) and for standardized matrices $S_0 = N$; μ is the average of x.

Spatial econometrics incorporates both spatial effects described above, with regression models. However, these models only deal with aggregated variables not allowing for inclusion of individual variables related to family background in the analysis.

Development of a methodology combining hierarchic and spatial methodologies was thus sought. This model was named hierarchic-spatial and is an important methodological contribution of this study. Characteristics of the spatial model – autocorrelation and spatial heterogeneity – are thus incorporated into the hierarchic model, which is non-spatial in nature, albeit considering different levels of aggregation in the same analysis.

The advantage of combining both methodologies is that the hierarchic model allows for joint analysis of variables related to individuals and the contextual (municipal) ones at different levels of aggregation, reducing the problem of ecological fallacy. In turn, the spatial model allows for possible spatial correlation of dependent and independent variables and/or errors at the second level of the hierarchic model, which will enable better estimation of parameters for regressions.

This field of study is starting to become more developed with use of Bayesian methods, which use the WinBUGS computer program in addition to MLwiN. However, what is proposed here is development of a technique which uses HLM and SpaceStat programs. A similar attempt was made by MORENOFF (2003) for a study about the weight of children upon birth. This study was the starting point for the methodology in this analysis.

MORENOFF (2003) assumes that the spatial process takes place through the autoregressive process of space lag[6], whose formula is as follows:

$$y = \rho W y + X\beta + \varepsilon \qquad (5)$$

Where: ρ is the autoregressive spatial parameter; W is the spatial weight matrix; X is the matrix of independent variables; β is the matrix of regression coefficients and ε is the vector of the random error term.

The equation above is the structural form of the spatial lag model. The reduced form makes it easier to notice how the spatial process works, also known in this case as spatial multiplier:

$$y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1}\varepsilon \qquad (6)$$

The term (I - ρW)-1 is Leontief's inverse, which connects the variable y_i to all x_i in the system through the spatial multiplier, and y_i to ε for all locations in the system, not only to the error in i (ANSELIN, 2003).

[6] There is another spatial model which is the spatial error model. It assumes that there is spatial correlation in variables which were not included in the model. Its formula is:

$$y = X\beta + \lambda W\varepsilon + u$$

Where: λ is the autoregressive coefficient; ε is the error term with spatial dependency and u is the random error term.

Equation (6) is important because it shows the spatial effect operates through X covariables and the non-observed error term, *i.e.* there is spatial dependency among observed and non-observed variables. The alternative found by MORENOFF to introduce the spatial effect into the hierarchic analysis was the inclusion of spatial lags for independent variables[7] into the hierarchic model, *i.e.* inclusion of WX. This procedure allows for correction of spatial dependency only for observed variables.

The strategy adopted in this study is made up of three steps. First, second level residuals of the non-conditional hierarchic model is obtained, *i.e.* when level-2 variables are not included in the analysis, only those of level-1.

Next, spatial econometric tests are performed on this residual[8]. With these tests it is possible to determine whether the spatial effect is in the lagged form or spatial error. Tests showed that the spatial effect takes place more intensely in the form of spatial lag – a result not included in this study.

The hierarchic-spatial model is estimated during the third stage, with two different specifications. The first consists in adding residual lags to the second level of the hierarchic model, *i.e.* in equation (2). Spatial dependency is thus corrected both in observed and non-observed variables.

The second specification consists in including spatial lag of independent contextual variables in level-2 of the hierarchic regression. This is the same strategy adopted by MORENOFF. It should be noted that this strategy does not control spatial effects of non-observed variables, only those of observed variables. However, it is useful because it enables observation of which covariable is more spatially correlated with independent variables.

3.2) Results

The results of the hierarchic-spatial model for school attendance by individuals between 7 and 14 years of age and 15 to 17 years of age can be found in Tables 5 and 6. To make comparisons easier, results for the hierarchic model in the previous section are also included in the tables. Considering that in this part variables of interest are contextual, results for level-1 variables are not shown; however, they do not differ from previous ones.

Analyzing individuals between 7 and 14 years of age first, Table 5 shows that inclusion of spatial lag of the residuals significantly reduces the impact of demographic pressure on both types of residuals – Bayesian and MQO. As an example, in the hierarchic-spatial model which includes spatial lags of the Bayesian residuals, a decrease in the relative cohort size of one standard deviation from the mean provokes a 0.31% variation in the probability of children between 7 and 14 years of age attending school, from 0.961 to 0.964. When the spatial term is not included in the analysis, variation is 1.13%. Regarding the variable of restricted availability, decrease in its effect is far lower, making it more important than relative cohort size.

---

[7] Spatial lag for a given variable in location j is its average in neighboring locations. Neighbors are determined by the spatial weight matrix.

[8] HLM generates two types of residue: Bayesian and MQO. The analysis was made for both residues.

When spatial lags of independent variables are added, the relative school-age cohort size in neighboring municipalities is the lagged factor which most influences access by children to school in a given municipality. A negative variation of one standard deviation from the mean relative cohort size in neighboring municipalities results in an increase of 0.46% in the probability of attending school, reaching 0.966. It should be noted that this small variation in probability is due to the fact that attendance by children between 7 and 14 years of age is practically at its maximum, *i.e.* 100%.

Regarding individuals between 15 and 17 years of age, table 6 shows that inclusion of spatial effects also reduces the impact of covariates. It is important to note that including lags of the residuals changes the direction of the demographic pressure variable, making it negative, which is the expected sign. On the other hand, in the model which includes only lags of independent variables, the sign remains positive, despite a smaller coefficient than in the non-spatial model (hierarchic only). This result indicates that relative cohort size is under other contextual effects, which are spatially correlated, since spatial residual lag is under spatial effects of observed and non-observed variables.

Among variables with lag, those of demographic pressure and restricted availability are the most important, with significant effect of the former. An increase in the relative 15 to 17 year-old cohort size in neighboring municipalities of one standard deviation from the mean increases the probability of attending school for this age group by 2.32%.

Based on these results it is clear that when attempts are made at controlling spatial autocorrelation in hierarchic models in order to reduce bias and solve efficiency problems in the estimates, the effect of level-2 variables decreases. On the other hand, however, spatial lags of contextual variables have a significant effect (hierarchic-spatial model with lag of independent variables).

These results indicate that level-2 variables have <u>direct</u> and <u>indirect</u> effects on the probability of attending school. The <u>direct effect</u> is the impact on the probability of attending school resulting from changes in contextual variables in the municipality itself. The <u>indirect effect</u>, in turn, is the impact resulting from changes in contextual variables in neighboring municipalities. The former effect is captured by the level-2 coefficients in both specifications of the hierarchic-spatial model (with residual lag and covariable lag), while the latter is captured by spatial lag coefficients of variables in the hierarchic-spatial model in which they are included. It may be said that the final result of a variation in the relative cohort size and ratio of teachers to school age population – main contextual factors in this case – is the sum of both these effects, since these variables are highly correlated in space, allowing them to change together.

Graphs 5 through 8 improve understanding of direct and indirect effects. Graph 5 shows the change in percentage in the probability of attending school for individuals between 7 and 14 years of age resulting from a decrease of one standard deviation from the mean relative school age cohort size in the municipality itself and in its neighbors for different models. The first three columns measure the increase in percentage in the event of a decrease in demographic pressure in the municipality itself, considering that in the first, simulations were made based on coefficients of the hierarchic model in the previous section[9], which makes no distinction between direct and indirect effects. The second and

[9] Model 5 was used for school attendance of children between 7 and 14 years of age and Model 6 for those between 15 and 17 years of age.

third columns have simulations calculated through use of hierarchic-spatial models with Bayesian lag of the residuals and level-2 lags of the covariables respectively. In this case, it is clear that this would be the direct effect of the increase caused by the decrease in relative cohort size. However, since the first model captures spatial autocorrelation of observed and non-observed variables, while the second model only considers observed variables, results are different. The higher the correlation of variables included in the model with those omitted, and the higher their spatial autocorrelation, the higher is the lag among direct effects found in both hierarchic-spatial models. The fourth column measures the increase in percentage when a decrease in the value of this variable takes place in neighboring municipalities, which was considered an indirect effect of the demographic dividend. Lastly, the final column is the sum of the third and fourth columns, *i.e.* the sum of direct and indirect effects of the relative school-age cohort size.

The same procedure was adopted in Graph 6 for the ratio of teachers to school age population. However, in this case an increase of one standard deviation from the mean was considered as variation in restricted availability. Graphs 7 and 8 resemble the first two, although they measure the impacts on probability of individuals between 15 and 17 years of age to attend school. Negative variation in probabilities, *i.e.* a decrease in percentage of probabilities shown in Graph 6 takes place due to the unexpected sign for relative cohort size.

The graphs show that when control over spatial autocorrelation of data is attempted, a reduction in the impact of the relative cohort size and the ratio of teachers to school age population indeed occurs. However, when adding the indirect effect of these variables, *i.e.* the effect via neighbors, it can be observed that an increase or decrease is close to what was found in the non-spatial hierarchic model.

This result may raise an important question. In spite of the bias in the hierarchic model estimate due to its lack of control over spatial autocorrelation of data, interpretation of its results may not be entirely wrong, if the fact that both direct and indirect effects of contextual variables are taken into consideration, while the hierarchic-spatial model separates these effects.

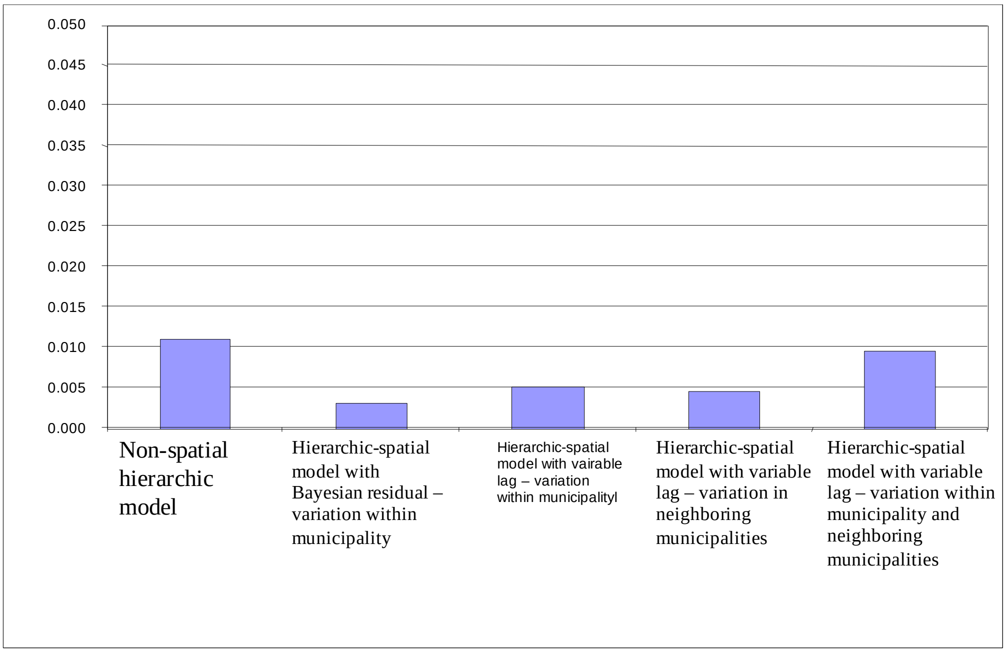
TABLE 5: Results of different models for probability of attending school – individuals between 7 and 14 years of age.

Independent variables	Hierarchic Model (model 5)		Model with Bayesian residual lag		Model with MQO residual lag		Model with independent variable lag	
Fixed effect	Coefficient	Sig	Coefficient	Sig	Coefficient	Sig	Coefficient	Sig
Intercept	4,120	0,000	3,134	0,000	3,047	0,000	3,613	0,000
Relative cohort size ages 7 to 14	-11,765	0,000	-2,802	0,000	-2,438	0,000	-4,744	0,000
Ratio of teachers to population between 7 and 14 years of age	10,845	0,000	8,144	0,000	8,289	0,000	9,295	0,000
Number of establishments with between 50 and 200 enrolled students	0,607	0,000	0,350	0,000	0,352	0,000	0,460	0,000
Number of establishments with between 201 and 1000 enrolled students	1,111	0,000	0,629	0,000	0,619	0,000	0,846	0,000
Number of establishments with up to 1000 enrolled students	0,924	0,000	0,210	0,175	0,317	0,040	0,718	0,000
Population up to 4999	-0,010	0,868	-0,286	0,000	-0,329	0,000	-0,283	0,000
Population between 5000 and 9999	-0,028	0,648	-0,236	0,000	-0,268	0,000	-0,258	0,000
Population between 10000 and 19999	-0,051	0,399	-0,217	0,000	-0,249	0,000	-0,236	0,000
Population between 20000 and 49999	-0,103	0,087	-0,218	0,000	-0,250	0,000	-0,252	0,000
Population between 50000 and 99999	-0,003	0,959	-0,055	0,302	-0,083	0,116	-0,093	0,148
Population between 100000 and 199999	0,019	0,793	-0,058	0,345	-0,063	0,301	-0,038	0,601
Spatial residual lag			0,781	0,000	0,690	0,000		
Spatial lag relative cohort size ages 7 to 14							-6,418	0,000
Spatial lag for ratio of teachers to population ages 7 to 14							5,040	0,000
Spatial lag for number of establishments between 50 and 200							0,696	0,000
Spatial lag for number of establishments between 201 and 1000							0,477	0,000
Spatial lag for number of establishments with up to 1000 enrolled students							-0,129	0,733
Spatial lag for population up to 4999							0,519	0,000
Spatial lag for population between 5000 and 9999							0,124	0,269
Spatial lag for population between 10000 and 19999							0,106	0,337
Spatial lag for population between 20000 and 49999							0,068	0,537
Spatial lag for population between 50000 and 99999							0,031	0,796
Spatial lag for population between 100000 and 199999							0,241	0,089
Random Effect								
Coefficient	0,287	0,000	0,192	0,000	0,190	0,000	0,268	0,000

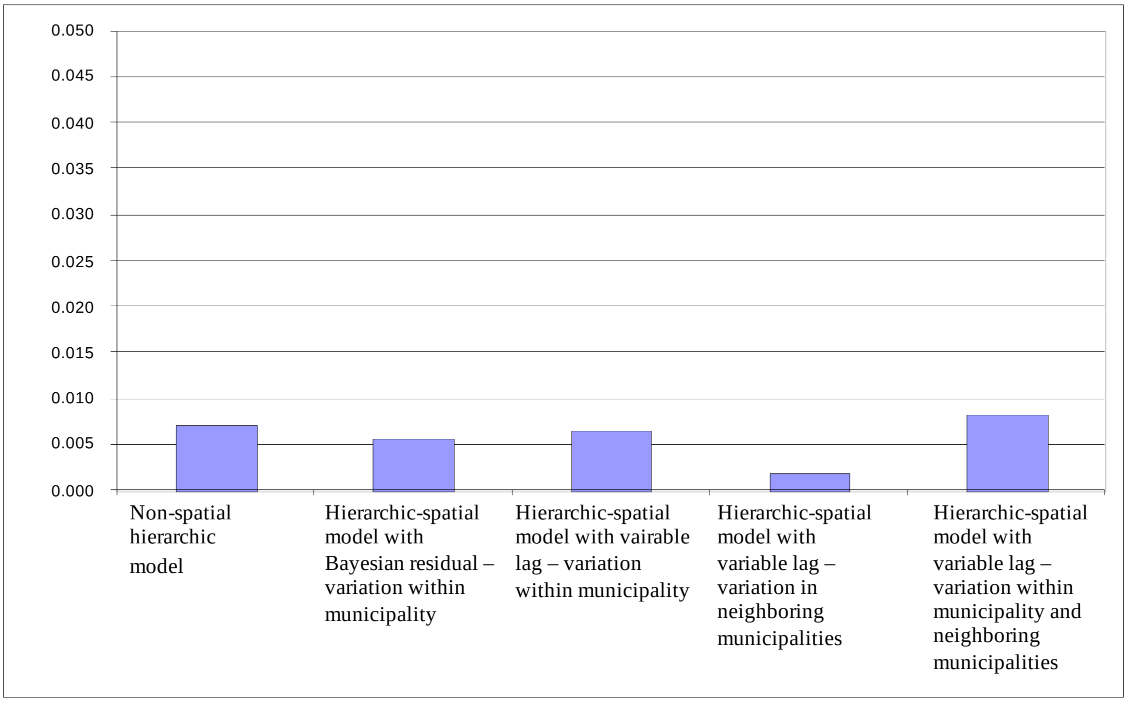
TABLE 6: Result of different models for probability of attending school – individuals between 15 and 17 years of age.

Independent variables	Hierarchic Model (model 6)		Model with Bayesian residual lag		Model with MQO residual lag		Model with independent variable lag	
Efeito Fixo	Coefficient	Sig	Coefficient	Sig	Coefficient	Sig	Coefficient	Sig
Intercept	1,113	0,000	1,909	0,000	1,910	0,000	0,637	0,000
Relative cohort size ages 7 to 14	9,241	0,000	-4,501	0,000	-4,157	0,000	5,762	0,000
Ratio of teachers to population between 15 and 17 years of age	10,285	0,000	9,602	0,000	9,277	0,000	9,284	0,000
Population up to 4999	-0,788	0,000	-0,624	0,000	-0,658	0,000	-0,714	0,000
Population between 5000 and 9999	-0,781	0,000	-0,589	0,000	-0,622	0,000	-0,696	0,000
Population between 10000 and 19999	-0,711	0,000	-0,557	0,000	-0,584	0,000	-0,625	0,000
Population between 20000 and 49999	-0,635	0,000	-0,503	0,000	-0,525	0,000	-0,548	0,000
Population between 50000 and 99999	-0,413	0,000	-0,284	0,000	-0,301	0,000	-0,328	0,000
Population between 100000 and 199999	-0,197	0,003	-0,129	0,009	-0,138	0,006	-0,163	0,010
Spatial residual lag			1,004	0,000	0,789	0,000		
Spatial lag relative cohort size ages 7 to 14							13,695	0,000
Spatial lag for ratio of teachers to population ages 7 to 14							4,308	0,000
Spatial lag for population up to 4999							-0,166	0,062
Spatial lag for population between 5000 and 9999							-0,578	0,000
Spatial lag for population between 10000 and 19999							-0,513	0,000
Spatial lag for population between 20000 and 49999							-0,404	0,000
Spatial lag for population between 50000 and 99999							-0,370	0,001
Spatial lag for population between 100000 and 199999							-0,049	0,695
Random Effect								
Coefficient	0,223	0,000	0,125	0,000	0,126	0,000	0,213	0,000

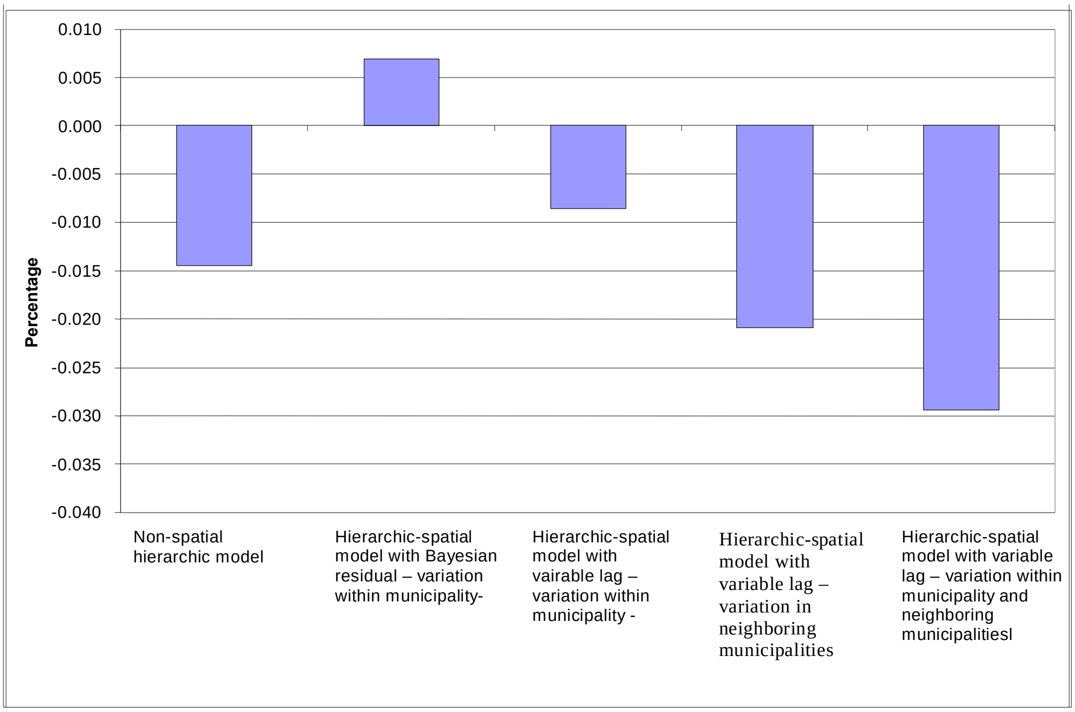
GRAPH 5: Variation in the probability of attending school resulting from decrease of one standard deviation from the mean relative school-age cohort size – simulations according to different models for individuals between 7 and 14 years of age



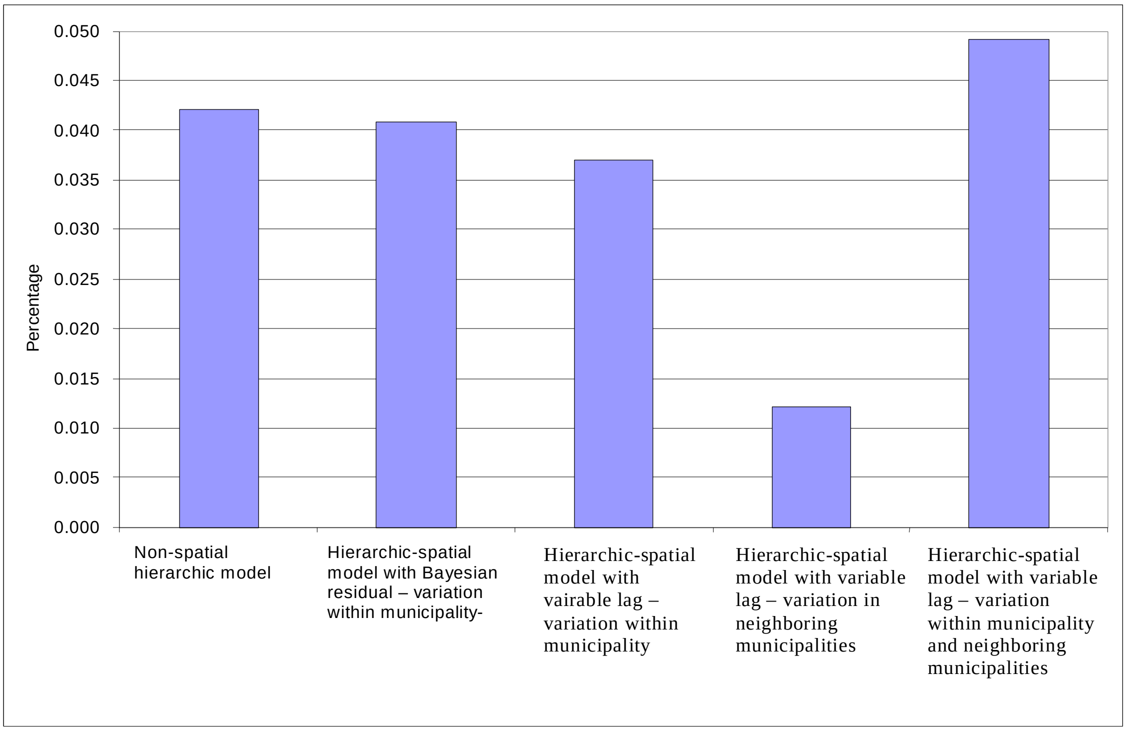
GRAPH 6: Variation in the probability of attending school resulting from increase of one standard deviation from the mean ratio of teachers to school-age population – simulations according to different models for individuals between 7 and 14 years of age



GRAPH 7: Variation in the probability of attending school resulting from decrease of one standard deviation related to the mean relative school-age cohort size – simulations according to different models for individuals between 15 and 17 years of age



GRAPH 8: Variation in the probability of attending school resulting from increase of one standard deviation from the mean ratio of teachers to school-age population – simulations according to different models for individuals between 15 and 17 years of age



4) CONCLUSION

This study sought to analyze the demographic dividend by means of two distinct methodologies, which are: hierarchic and hierarchic-spatial. These models made it possible to incorporate non-demographic variables into the study, thus bringing family factors and restricted availability of education in municipalities into the analysis, in addition to spatial correlation resulting from georeferenced data.

Results of the first methodology showed that lower demographic pressure has an important effect on access to education for children between 7 and 14 years of age, making up the most important contextual factor. Smaller cohorts in this age group increase the probability of attending school. This result is in accordance with COALE & HOOVER (1956).

For access by individuals between 15 and 17 years of age, the impact of the demographic dividend was the opposite of that found in the lower age group. On the other hand, restricted availability was the main factor for greater access to education for these individuals. Thus, there is clear need for public policy to expand availability of middle school education in Brazilian municipalities.

When spatial autocorrelation is incorporated into the analysis – the hierarchic-spatial model – two types of impacts of independent variables can be observed, which were called direct and indirect effects. The former is the impact on school attendance resulting from changes in contextual variables in the municipality itself, while the latter is the impact related to changes in contextual variables of neighboring municipalities. It has been observed that when spatial dependency is controlled, direct effects drop significantly, making restricted availability the most important contextual factor for access to education for both age groups. Also worthy of note is the fact that for attendance by individuals between 15 and 17 years of age the coefficient of relative school-age cohort size in the model which includes spatial residual lag becomes negative, as expected. This suggests that relative cohort size in the 15 to 17 year-old age group is correlated with other factors which were not included in the analysis, which are spatially correlated.

Regarding the indirect impact of the demographic dividend, results showed its important effect, *i.e.* variations in relative school-age cohort size in neighboring municipalities cause an important change in the probability of children attending school.

5) BIBLIOGRAPHY

ANSELIN, L. **Spatial econometrics**. Dallas: University of Texas, School of Social Science, 1999. 284p.

ANSELIN, L. Spatial externalities, spatial multipliers, and spatial econometrics. **International Regional Science Review**, v.26, n.2, p.153-166, 2003.

BARROS. R. P. *et al.* Determinantes do desempenho educacional do Brasil. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v.31, n.1. p.1-42, abr. 2001.

BRYK, A. S., RAUDENBUSH, S. W. **Hierarchical linear models**: applications and data analysis methods. 2.ed. Newbury Park, California: Sage, 2002. 485p.

COALE, A., HOOVER, E. **Population growth and economic development in low-income countries**: a case study of India's prospects. Princeton, NJ: Princeton University, 1958. 389p.
HANUSHEK, E. A, GOMES-NETO, J. B., HARBISON, R. W. Efficiency-enhancing investments in school quality. In: BIRDSALL, N., SABOT, R. H (Eds.) **Opportunity forgone**: education in Brazil. Washington, DC.: Inter-American development Bank.1996. p.385-424.
HANUSHEK, E. A. Publicly provided education. In: AUERBACH, A. J., FELDSTEIN, M. (Eds.) **Handbook of public economics**. Amsterdam: Elsevier, 2002. v.4, p. 2045-2141.
JONES, G. J. **Population dynamics and education and health planning**. Geneva: International Labour Office, 1990. 88p. (Wold Employment Programme. Paper, 8)
KRUEGER, A. B. Experimental estimates of education production functions. **The Quarterly Journal of Economics,** v.114, n.2, p.497-532, May. 1999
KRUEGER, A. B. Reassessing the view that American schools are broken. **Economic Research bank of New York**, v.4, n.1, p.29-46, Mar. 1998.
LAZEAR, E. P. **Educational production**. Cambridge: National Bureau of Economic Research, 1999. 48p. (Working paper, 7349)
LEE, J., BARRO. R. J. **Schooling quality in a cross section of countries**. Cambridge: National Bureau of Economic Research, 1997. 33, [9] p. (Working paper, 6198)
MORENOFF, J. D. Neighborhood mechanisms and the spatial dynamics of birth weight. **American Journal of Sociology**, v.108, n.5, p.976-1017, Mar. 2003.
RIANI, J. L. R. Impactos da estrutura etária em indicadores de educação no Brasil, 1991. **Revista Brasileira de Estudos Populacionais**, v.18, n1/2, p.15-34, jan./dez, 2002
RIANI, J. L. R. **Determinantes do resultado educacional no Brasil: família, perfil escolar dos municípios e dividendo demográfico numa abordagem hierárquica e espacial**. Tese de Doutorado. Universidade Federal de Minas Gerais, 2005.
RIOS-NETO, E. L. G., CÉSAR, C. C., RIANI, J. L. R. Estratificação educacional e progressão escolar por série no Brasil. **Pesquisa e Planejamento Econômico**, Rio de Janeiro, v.32, n.3, p.395-415, Dez. 2002.
SCHULTZ, T. P. School expenditures and enrollments, 1960-1980: the effects of income, prices and population growth. In: JOHNSON, D. G., LEE, R. D. (Eds.) **Population growth and economics development**: issues and evidence. Madison, WI: University of Wisconsin, 1987. p.413-478.
SILVA, N. V. **Atualização da escala socioeconômica de ocupação para 1980**. Laboratório Nacional de Computação Científica, CNPq, 1985. 17p. (Mimeogr.)
SILVA, N. V. **Posição social das ocupações**. Rio de Janeiro: IBGE, Centro de Informática, 1973. (Mimeogr.)
SILVA, N. V., HASENBALG, C. **Recursos familiares e transições educacionais**. Versão Preliminar apresentada no Workshop de Demografia da Educação. Associação Brasileira de Estudos Populacionais, ABEP, Salvador, Bahia, Junho 2001
TODD, P. E., WOLPIN, K. I. On the specification and estimation of the production function for cognitive achievement. **The Economic Journal**, v.113, n.485, p.F3-F33, Feb. 2003.