Returns to Schooling and Quality of Education in Brazil: Evidence from Migrants Data

Luiz Mário Brotherhood   Pedro Cavalcanti Ferreira   Cezar Santos

Escola Brasileira de Economia e Finanças (EPGE/FGV-RJ)

July 20, 2016

Abstract

We provide a new index of educational services' quality of Brazilian states in 2010. This measure is constructed based on the notion that the financial returns obtained from an additional year of schooling can be seen as derived from the value that market forces assign to this education. We use migrants data to estimate returns to schooling of individuals that studied in different states but who work in the same labor market. We find very heterogeneous educational qualities across states: the poorest Brazilian region presents education quality index equal to approximately one third of the other regions. We compare our index with basic education test scores and conclude that there are important similarities and differences between them.

Keywords: quality of education, returns to schooling.

JEL Classification: I21, I26.

ANPEC Area: Labor Economics.

Resumo

Nesse artigo, construímos um novo índice de qualidade da educação dos estados brasileiros em 2010. A construção dessa medida é baseada na ideia de que o retorno financeiro obtido com um ano adicional de estudo pode ser considerado como originado do valor que as forças de mercado atribuem a essa educação. Usamos dados de migrantes para estimar retornos à escolaridade de indivíduos que estudaram em estados diferentes, mas que trabalham no mesmo mercado de trabalho. Obtemos um índice de qualidade educacional muito heterogêneo: a região brasileira mais pobre possui qualidade da educação igual a aproximadamente um terço das demais regiões. Comparamos esse índice com notas de exames do sistema de educação básico e concluímos que há importantes similaridades e diferenças entre as duas medidas.

Palavras-chave: qualidade da educação, retorno à educação.

Classificação JEL: I21, I26.

Área ANPEC: Economia do Trabalho.

1 Introduction

Education is very important for socioeconomic development.[1] A country's level of education has two dimensions: quantity and quality. The first one has been studied extensively.[2] The same is not true for quality. This dimension is very complex because it may involve subjective considerations, becoming very hard to be measured. Besides this, some authors argue that quality of education matters even more than quantity for economic growth.[3]

In this paper we provide a new index of educational services' quality of Brazilian states in 2010. This measure is constructed based on the notion that the financial returns obtained from an additional year of schooling can be seen as derived from the value that the market assigns to this education. Therefore, differences between returns to schooling of individuals who studied in different states, all else equal, is due to differences between the quality of educational services consumed by them.

At first, one can think of constructing such measures by using each state data separately. However, a possible drawback of this approach is that labor market characteristics vary significantly over Brazilian states. Thus, comparing educational returns in different states as education quality measures can lead to biased analysis. For example, suppose that labor markets are more developed in higher income states so that individuals with high-tech knowledge earn significantly more than less educated workers. On the other hand, suppose that technology companies are incipient in lower income states and that high-tech professionals earn not much more than less educated workers. In this case, estimating returns to schooling separately by state and interpreting it as education quality measures would lead one to conclude that educational services are better in richer states. However, in a counterfactual scenario it is possible that both workers have the same educational returns had they worked in the same firm.

To prevent this type of bias, we use data on individuals who obtained education in different states but who work in markets with similar characteristics. This is done by using 2010 census data on individuals who lived in São Paulo (SP), the Brazilian state with the largest population. The 2010 census contains information on migration that can be used to infer which migrants likely completed schooling in their state of birth. This strategy is similar to the one used in Schoellman (2012). Since migrants are probably positively self-selected,[4] we use Heckman's (1979) selection correction method.

Brazil is divided into five regions,[5] which are very unequal in terms of economic outcomes. The Northeast and North are the country's poorest regions, having per capita GDP

[1] See Sen (2000), and Banerjee and Duflo (2012).

[2] For reviews of this literature, see Sianesi and van Reenen (2003), and Krueger and Lindahl (2000).

[3] Hanushek and Wobmann (2007) argue that analysis limited to schooling quantity misses the core of what eduacation is all about. They point that extensive evidence on knowledge development and cognitive skills indicates that a variety of factors outside of school (family, peers, and others) have a direct and powerful influence on economic growth.

[4] See Ferreira and Santos (2007).

[5] Table 2 displays the distribution of states by each region.

in 2013 equal to R$ 12,954 and R$ 17,213 respectively, followed by the South (R$ 30,495), Midwest (R$ 32,322), and Southeast (R$ 34,789) regions. Compatible with this ranking, our method produces very heterogeneous educational quality indexes across states. Regional means are: Northeast 3.4%, Midwest 6.9%, South 8.9%, and Southeast 9.7%.[6]

We compare our index with 1995-2005 basic education test scores and conclude that there are important differences and similarities between them. On the one hand, our method pictures a more unequal scenario than test scores do: Northeastern states' mean score is equal to 92.3% of the others regions' mean for the case of basic education test scores, and 30.2% for the case of our index. This is evidence that there are aspects related to education quality that our method captures, but test scores do not. On the other hand, despite the differences there is a high correlation between both indexes.

We believe that this paper provides at least two important contributions. First, we contribute to the scarce literature on economics and educational quality in Brazil[7] by providing an alternative quality measure based on the value assigned to education by market forces. This can inform researchers and subsidize policy makers. Second, the high correlation between our index and test scores is evidence that supports the use of returns to education as an educational quality index. For some countries, educational quality measures are scarce, whereas data on earnings and schooling are easily available. Therefore, verifying the correlation between returns to schooling and test scores can support researchers interested in constructing education quality measures for developing countries.

Our method is very similar to Schoellman's (2012), who constructs measures of educational quality for several countries by estimating returns to schooling for immigrants in the United States. Kaarsen (2014) provides another measure of human capital quality by investigating the Trends in Math and Science Study test scores, exploiting the fact that those exams were applied in different countries for two distinct grades.

Some authors corroborate that returns to schooling of immigrants are positively correlated with mean educational quality in the source country. Chiswick and Miller (2010) and Bratsberg and Terrell (2002) verify that international test scores explain differences in the rate of return to schooling among immigrants in the United States. Li and Sweetman (2013) find the same for the case of Canada.

Methodologically, this paper is related to the literature that exploits migration for identification purposes. For example, Card and Krueger (1992) studies returns to schooling of cross-state migrants with the objective of estimating the education production function in the United States. Alesina and Giuliano (2010) study how strength of family ties affect the behavior of second-generation immigrants in the United States. Okoye (2012) presents evidence that investment in higher education produces positive externalities by verifying that returns to schooling of immigrants in the U.S. in 2000 are positively correlated to the proportion of higher educated workers in 1980 in the country of origin, even after controlling

[6]We do not consider the North region because in our dataset there is a small number of observations of migrants in São Paulo who were born in northern states.

[7]For a survey on field researches related to child education quality in Brazil, see Campos et al. (2006).

for other determinants of education quality. Moriconi and Peri (2015) use variations among behavior of first and second-generation cross-country European migrants to isolate the effect of culturally transmitted labor-leisure preferences on individual employment rates.

There is a large literature on returns to schooling in Brazil. For example, Jacinto and Rodeghiero (2012) use the 2007 Household Sample National Survey (Pesquisa Nacional por Amostragem de Domicílios – PNAD) and three different methods (Griliches' (1977), Heckman's (1976), and Garen's (1984)) to estimate returns to schooling. They find heterogeneous returns across schooling levels, and that higher education produces the highest return. Barbosa Filho and Pessôa (2008) use 1980-2004 PNADs and the 2000 census to estimate the Internal Rate of Return of education, obtaining a value equal to 18%. Resende and Wylie (2006) use the 1996-1997 Research on Living Standards (Pesquisa sobre Padrões de Vida – PPV) and Heckman's (1976) method to conclude that returns to schooling equal 12.6% for women and 15.9% for men. Sachsida et al. (2004) estimate returns using simple OLS, Heckman's (1976), and Garen's (1984) methods. They use PNAD data in three different ways: cross section for 1996, and pooled data and pseudopainel for 1992-1999. Estimated returns are between 10% and 22%. Our paper is naturally distinct from those since we are not interested in returns to schooling per se, but on educational quality measures.

This paper is organized in four more sections. Section 2 describes the datasets used, the sample selection strategy, and descriptive statistics. Section 3 explains the method used to construct educational quality measures, and analyzes the results. Section 4 compares our educational quality index with basic education test scores. Section 5 presents concluding comments.

2 Data

In order to estimate educational returns, we use data from the 2010 Brazilian census. This dataset is provided by the Brazilian Institute of Geography and Statistics (IBGE), and contains information related to individuals' residence characteristics, work, migration, schooling, mobility, and fecundity. We use data on individuals' earnings in the main job, hours worked per week, schooling attainment, age, state of birth, state of residence, race, gender, and urban/rural residence.

As Schoellman's (2012) data, the 2010 census does not provide direct information on where schooling was obtained. We follow the same strategy as him and use information on age, year of migration, and schooling attainment to infer which migrants likely completed schooling in their state of birth. Therefore, our sample includes individuals with age greater than or equal to 24, which is the expected date of graduation plus six years. This buffer is used in order to minimize measurement error from migrants who repeat grades, start school late, or experience interruptions in their education. We exclude migrants who are studying in São Paulo and, for individuals who were born and work in São Paulo, we exclude those who are studying in another state or those who previously lived in another state. We also exclude

Table 1: Descriptive statistics (means) and number of observations by state of birth and residence

	Years of schooling		Earnings per weekly hours		Number of observations			
	Living in	Living in	Living in	Living in	Living in SP		Living in other states	
State of birth	SP	other states	SP	other states	N	Percent	N	Percent
Rondônia	8.29	8.36	30.58	33.54	117	0.01	33,054	0.40
Acre	11.04	7.80	59.39	30.82	41	0.00	36,737	0.45
Amazonas	10.93	7.75	67.47	29.38	189	0.01	115,093	1.41
Roraima	9.81	8.48	36.29	35.56	13	0.00	12,585	0.15
Pará	9.13	7.33	54.00	26.97	879	0.06	270,569	3.30
Amapá	8.18	9.12	56.89	38.18	40	0.00	19,779	0.24
Tocantins	8.39	7.56	26.99	24.32	169	0.01	87,837	1.07
Maranhão	6.92	6.68	29.69	21.69	3,460	0.22	383,156	4.68
Piauí	6.11	6.01	26.82	18.20	5,151	0.32	242,363	2.96
Ceará	6.04	6.90	30.09	22.25	8,546	0.54	429,305	5.24
Rio Grande do Norte	6.47	6.93	34.26	23.04	2,034	0.13	202,705	2.48
Paraíba	5.71	6.25	28.64	20.64	6,261	0.39	284,200	3.47
Pernambuco	5.82	7.12	33.93	24.69	16,181	1.02	448,861	5.48
Alagoas	5.53	6.22	26.17	21.36	6,758	0.43	164,709	2.01
Sergipe	5.89	6.72	28.72	22.52	2,582	0.16	111,447	1.36
Bahia	5.93	6.83	28.36	22.45	28,156	1.77	760,400	9.29
Minas Gerais	6.73	7.30	40.46	28.65	29,229	1.84	1,307,237	15.97
Espírito Santo	7.61	8.03	55.83	34.07	832	0.05	188,301	2.30
Rio de Janeiro	10.60	9.33	101.51	46.20	4,309	0.27	556,174	6.79
São Paulo	9.32	8.85	47.34	47.22	1,445,752	91.07	134,533	1.64
Paraná	6.67	7.92	34.93	32.25	20,035	1.26	636,716	7.78
Santa Catarina	9.51	7.96	87.06	35.22	1,186	0.07	418,144	5.11
Rio Grande do Sul	10.46	8.00	94.26	33.59	1,818	0.11	806,930	9.86
Mato Grosso do Sul	8.23	7.91	40.08	32.42	1,686	0.11	102,544	1.25
Mato Grosso	7.63	8.12	33.76	31.63	731	0.05	97,630	1.19
Goiás	8.47	7.88	51.59	33.40	1,119	0.07	303,779	3.71
Distrito Federal	10.88	10.67	87.98	67.27	198	0.01	32,545	0.40
Total	9.09	7.51	46.46	29.79	1,587,472	100	8,187,333	100

Table 2: Saeb mean test scores

	1995	1997	1999	2001	2003	2005
Brazil	244.2	242.3	230.1	227.5	229.1	226.7
North	227.2	227.5	215.2	214.1	214.9	213.0
Rondônia	231.8	230.2	220.1	224.2	220.4	223.7
Acre	219.2	219.3	209.1	208.9	218.8	218.5
Amazonas	229.3	227.1	217.5	210.4	214.8	207.7
Roraima	228.8	219.1	215.6	213.9	222.7	219.7
Pará	228.9	230.3	215.4	217.0	214.1	214.0
Amapá	225.0	221.7	219.8	214.7	219.3	214.4
Tocantins	220.5	225.8	207.6	210.2	210.8	210.7
Northeast	226.0	235.4	218.1	212.9	216.0	212.2
Maranhão	214.0	224.3	211.7	208.2	214.7	203.8
Piauí	225.4	240.9	221.4	219.8	217.1	212.1
Ceará	230.5	241.1	222.0	213.1	217.9	215.9
Rio Grande do Norte	227.8	231.0	214.0	211.8	212.1	208.6
Paraíba	228.4	230.9	222.3	215.5	214.5	210.6
Pernambuco	224.8	233.3	213.8	210.5	215.9	212.9
Alagoas	223.7	225.4	216.3	211.3	212.5	208.1
Sergipe	236.7	238.4	220.6	216.1	214.9	222.5
Bahia	228.4	243.7	220.4	214.8	218.5	214.2
Southeast	252.0	243.9	235.1	234.6	235.9	233.9
Minas Gerais	255.7	265.5	236.9	236.2	238.2	241.2
Espírito Santo	235.3	239.0	234.6	231.5	231.7	234.5
Rio de Janeiro	245.0	238.0	241.4	236.6	237.8	230.5
São Paulo	254.1	239.6	232.7	233.6	234.7	231.5
South	250.0	253.0	239.4	238.5	241.0	238.9
Paraná	249.1	251.4	236.4	231.5	236.1	235.5
Santa Catarina	246.3	254.4	241.4	239.9	218.4	237.5
Rio Grande do Sul	253.3	254.3	241.9	245.7	244.3	244.1
Midwest	247.6	246.9	232.7	229.7	230.8	228.9
Mato Grosso do Sul	241.5	248.7	231.1	231.8	231.6	233.0
Mato Grosso	233.3	231.5	225.1	222.6	222.4	219.9
Goiás	246.4	249.3	233.1	226.5	229.9	224.6
Distrito Federal	265.1	253.3	240.2	244.8	243.7	247.6

Minimum score is 0 and maximum score is 500.

individuals with age greater than 65.

The 2010 census also lacks information on the exact numbers of years of schooling attainment. Instead, it is possible to construct a categorical educational variable that identifies the following years of schooling intervals: from 0 to 3 years, 4-7, 8-10, 11-14, and 15 years or more. We deal with this limitation in two alternative ways. First, we impute individuals' years of schooling in the first four intervals as the interval mean, and 15 years for individuals in the last interval. Second, in Subsection 3.1 we estimate returns to schooling using dummy variables for each educational category and calculate the weighted mean return using the fraction of individuals in each interval as weights. Both methods produce qualitatively similar results.

Table 1 contains descriptive statistics and number of observations by state of birth. Note that our sample includes individuals who do not live in São Paulo because those observations are used in Heckman's selection correction method. For most states of birth, mean years of schooling are higher for individuals who live in São Paulo. This is evidence that migrants are positively selected. We exclude the North region and Distrito Federal in our main analysis because they present small number of observations of migrants. Including those observations produces estimates with large standard errors, making inference questionable.

To compare estimated returns to schooling with another educational quality measure, we use Sistema Nacional de Avaliação da Educação Básica (Saeb) test scores for the years 1995-2005. The Saeb exam is administered by the Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep), an institution associated to the Ministry of Education and, since 1995, is composed by biennial Mathematics and Portuguese tests applied to samples of students in the 4th and 8th grade of the Elementary School (Ensino Fundamental), and the 3rd year of the High School (Ensino Médio), in public and private schools.[8]

Table 2 contains mean Saeb test scores by year, state and region. During 1995-2005, Minas Gerais, Rio Grande do Sul, and Distrito Federal students had the highest scores, and Roraima, Tocantins, and Maranhão students had the lowest scores. For all years, the Southeast region ranked first, followed by the South region, and the North and Northeast regions shared the last position. At national level, there is a high downward trend in mean scores.

3 Returns to schooling as educational quality measures

Our objective is to construct new measures of the quality of educational services by state in Brazil through the estimation of returns to schooling. Our strategy builds on the idea that financial returns obtained from an additional year of schooling can be seen as derived from the value that the market assigns to this education. Therefore, differences between returns to schooling of individuals who studied in different states, all else equal, is due to differences

[8] Inep's Índice de Desenvolvimento da Educação Básica (Ideb) is a broader educational quality index than Saeb because it embodies both Saeb test scores and approval rates. However, we do not use it because it is available only from 2005, a relatively recent year. Our dataset contains information on individuals who were working in 2010, and most of them studied many years before 2005.

between the quality of educational services that they consumed.

A first approach to implement this idea empirically is to estimate the following augmented Mincerian regression:

$$\log(W_{ij}) = \alpha_j + \beta_j S_{ij} + \gamma X_{ij} + u_{ij}, \tag{1}$$

where i and j index individual and state, respectively, W denotes earnings per weekly hours worked, S denotes years of schooling, X is a vector of control variables that includes gender, age, age squared, race, and urban residence dummy, and u is an error term. α_j is an intercept that can vary across states, and β_j is the return to schooling for individuals working in state j.

Column (1) of Table 3 displays returns to schooling estimates of equation (1). In this specification the Northeastern states present the highest returns. For example, one additional year of schooling in Piauí is associated to a 10% increase in earnings. Santa Catarina has the lowest return, equal to 7%.

Interpreting these estimates as educational quality measures is problematic because labor market characteristics vary significantly across Brazilian states, making it possible that two different markets reward the same schooling quality differently. To overcome this problem, we use data only on individuals who work in São Paulo but who obtained education in different states. Therefore, define now j as state of birth instead of state where individual i works.[9] Column (2) of Table 3 provides returns to schooling estimates using only individuals who work in São Paulo in our baseline sample. For comparison with the previous result, Figure 1 plots estimates for the first two models. Note that the two methods produce very different estimates. For example, Northeastern states' estimates are the largest in Model 1, but are the smallest in Model 2. Rio de Janeiro (RJ), Espírito Santos (ES), Rio Grande do Sul (RS), and Santa Catarina (SC) also have very distinct estimates. This result is consistent with the idea that skilled labor is scarce in the poorest states, so that market forces offer a high reward for education in those regions. After we use data only on individuals that work in the same labor market, we are able to obtain a superior measure of education quality as valued by market forces.

Model 2 estimates are still questionable if we want to interpret returns as educational quality measures. Earnings in São Paulo are obviously not observed for individuals who do not work there. If the decision to work in São Paulo is determined by variables that are correlated to individuals' years of schooling, estimation of (1) by OLS produces biased and inconsistent coefficients. Therefore, we use Heckman's (1979) selection correction method (Heckit Model) and postulate that earnings in São Paulo are observed if

$$\delta_j + \eta S_{ij} + \phi X_{ij} + \psi E_{ij} + v_{ij} > 0, \tag{2}$$

where δ_j are intercepts that vary across state of birth and E_{ij} is the (expected) earnings per hour of individual i if she decides to work in São Paulo in relation to working in another

[9]See Section 2 for a discussion on how the sample was selected so that we can attribute a high probability that individuals completed schooling in their state of birth.

Table 3: Returns to schooling estimates

	(1) Model 1	(2) Model 2	(3) Heckit Model
Maranhão	0.100	0.0397	0.0328
	(0.0006)	(0.0044)	(0.0042)
Piauí	0.102	0.0311	0.0231
	(0.0008)	(0.0033)	(0.0031)
Ceará	0.0986	0.0455	0.0368
	(0.0005)	(0.0032)	(0.0030)
Rio Grande do Norte	0.0938	0.0435	0.0376
	(0.0007)	(0.0058)	(0.0057)
Paraíba	0.0970	0.0386	0.0310
	(0.0007)	(0.0035)	(0.0034)
Pernambuco	0.0932	0.0439	0.0351
	(0.0005)	(0.0024)	(0.0023)
Alagoas	0.0954	0.0457	0.0373
	(0.0008)	(0.0036)	(0.0034)
Sergipe	0.0999	0.0491	0.0393
	(0.0009)	(0.0059)	(0.0057)
Bahia	0.0954	0.0429	0.0335
	(0.0004)	(0.0016)	(0.0016)
Minas Gerais	0.0820	0.0834	0.0746
	(0.0002)	(0.0017)	(0.0017)
Espírito Santo	0.0888	0.112	0.102
	(0.0005)	(0.0090)	(0.0090)
Rio de Janeiro	0.0885	0.137	0.131
	(0.0003)	(0.0050)	(0.0048)
São Paulo	0.0807	0.0850	0.0820
	(0.0002)	(0.0003)	(0.0003)
Paraná	0.0795	0.0705	0.0614
	(0.0003)	(0.0020)	(0.0019)
Santa Catarina	0.0700	0.112	0.103
	(0.0003)	(0.0085)	(0.0084)
Rio Grande do Sul	0.0865	0.112	0.105
	(0.0003)	(0.0080)	(0.0079)
Mato Grosso do Sul	0.0824	0.0709	0.0635
	(0.0007)	(0.0063)	(0.0062)
Mato Grosso	0.0701	0.0720	0.0644
	(0.0006)	(0.0103)	(0.0106)
Goiás	0.0737	0.0916	0.0813
	(0.0005)	(0.0085)	(0.0083)
N	5,954,565	1,058,650	5,478,685

Standard errors in parentheses. All estimates are significant at one percent.

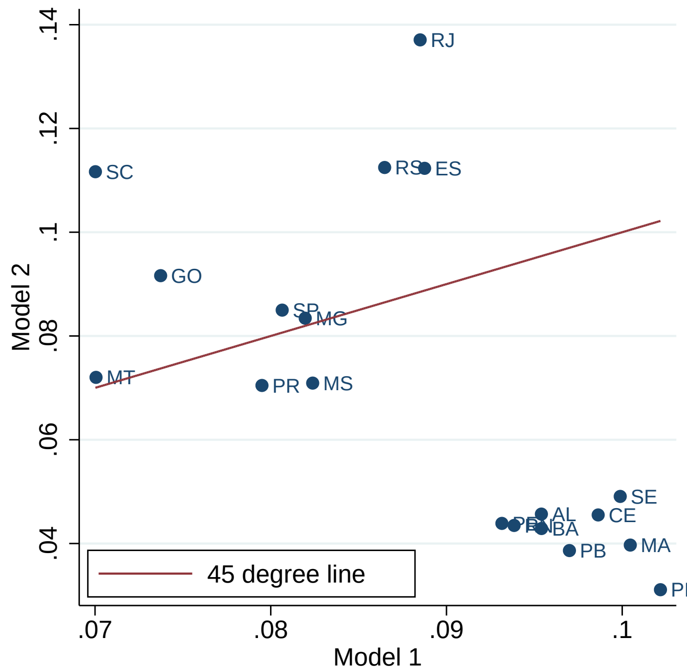

Figure 1: Model 1 and Model 2 returns to schooling estimates

state. In particular, for an individual who works in São Paulo, E_{ij} is equal to her earnings divided by the expected earnings if she works in another state. For an individual who works in another state, E_{ij} equals the expected earnings if she works in São Paulo divided by her current earnings. To calculate expected earnings we use fitted values of linear regressions. That is, we first run a series of regressions of earnings per weekly hours on years of schooling, gender, age, age squared, race, and urban residence dummy for each possible combination of state of birth and a dummy variable that indicates residence in São Paulo. Then, for example, the expected earnings of working in São Paulo for an individual who currently has a job in Rio de Janeiro is computed as the fitted value of the regression that uses data on individuals who work in São Paulo and were born in Rio de Janeiro. Additionally, we posit that

$$u_{ij} \sim N(0,\sigma^2), \quad v_{ij} \sim N(0,1), \quad \mathrm{corr}(u_{ij}, v_{ij}) = \rho. \tag{3}$$

We estimate this model using the Maximum Likelihood method. For comparison, Figure 2 plots estimates for Model 2 and Heckit. Observe that all states have lower estimates in Heckman's model, except for São Paulo. This is evidence that migrants are positively selected and our method corrects the selection bias by increasing São Paulo's returns in relation to the other states. Also, except for São Paulo, the vertical difference between estimate and the 45 degree line is very similar for all states, indicating that selection biases have similar magnitudes across states.

Column (3) of Table 3 and Figure 3 display returns to schooling estimates that can be

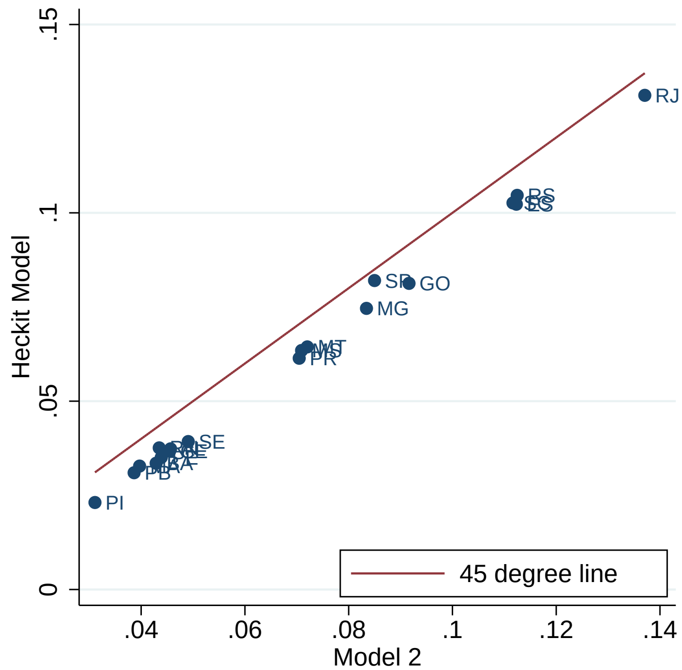

Figure 2: Model 2 and Heckit returns to schooling estimates

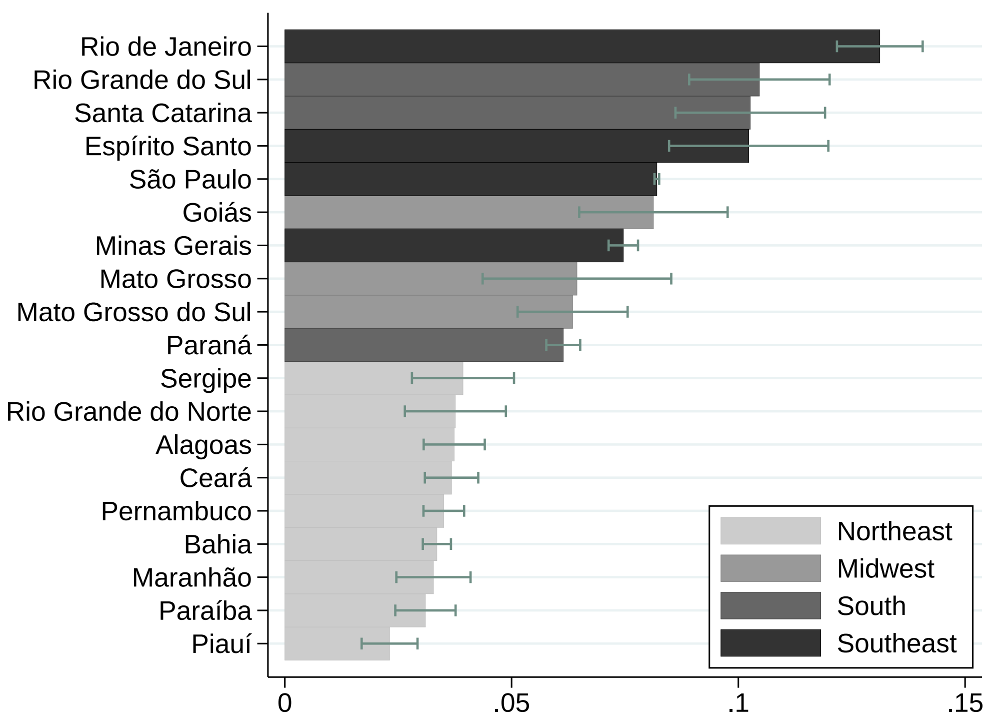

Figure 3: Heckit estimates and 95% confidence intervals

Table 4: Heckit selection equation elasticities

	Elasticity
Schooling	0.0453
	(0.0012)
Earnings Ratio	0.0320
	(0.0074)
Age	-0.0098
	(0.0001)
Race	
Black	0.1428
	(0.0041)
Pardo	0.1459
	(0.0025)
Other	0.1792
	(0.0095)
Woman	-0.0086
	(0.0020)

Standard errors in parentheses. All estimates are significant at one percent. Elasticities in terms of the following variations. Years of schooling and Earnings ratio: one standard deviation increase centered in the mean value. Age: from 35 to 36 years.

interpreted as educational quality measures, our main result. Rio de Janeiro and Piauí present the highest and lowest estimates, respectively. That is, after controlling for selection issues, if we take two individuals born in Rio de Janeiro that work in São Paulo and have the same observable characteristics, except for the fact that one has one more year of schooling than the other, it is expected that the earnings of the more educated individual are 13.1% higher than the less educated one. In contrast, one additional year of education in Piauí, one of the poorest states in Brazil, increases earnings by only 2.3%. The Northeast region unambiguously presents the lowest educational quality, while the other regions display heterogeneity. Mean schooling returns by region are as follows: Northeast 3.4%, Midwest 6.9%, South 8.9%, Southeast 9.7%.

Table 4 shows elasticities related to the coefficients in the selection equation (2). A one standard deviation increase in schooling (expected earnings derived from working in São Paulo in relation to other states) produces a 4.5% (3.2%) higher probability of working in São Paulo. An individual who is 36 years old presents 0.9% lower probability of working in São Paulo than an individual who is 35.

Correlation between error terms estimate is $\hat{\rho} = 0.7$, and the p-value associated to the test $\rho = 0$ is approximately equal to zero. Therefore, we reject the null and conclude that there is selection bias in Models 1 and 2 estimates.

3.1 Categorical schooling variable

In the previous analysis we used imputed schooling data because the 2010 census does not provide the exact number of years of schooling. An alternative to imputation is using a categorical schooling variable that identifies the following years of schooling intervals: from 0 to 3 years, 4-7, 8-10, 11-14, and 15 years or more. We proceed in two steps. First we estimate Heckman's model modifying equation (1) to be

$$\log(W_{ij}) = \alpha_j + \sum_{k=1}^{5} \beta_{jk} D_{ijk} + \gamma X_{ij} + u_{ij}, \tag{4}$$

where k assumes the 5 possible values of the categorical schooling variable and D_{ijk} is a dummy that indicates if individual i's schooling belongs to interval k. We also use the categorical schooling variable in the selection equation (2). This step produces 4 schooling coefficients for each state (one of them is omitted to avoid collinearity). Second, we compute for each state the weighted mean of schooling coefficients using the fraction of individuals in each schooling interval as weights. The result is an average of marginal effects for each state.

Figure 4 plots Heckman's model estimates using imputed and categorical schooling variables. Both methods produce qualitatively similar results. Note that averages of marginal effects have different magnitudes than returns estimated previously. This happens because those estimates no longer have the interpretation of expected increase in earnings due to one additional year of schooling.

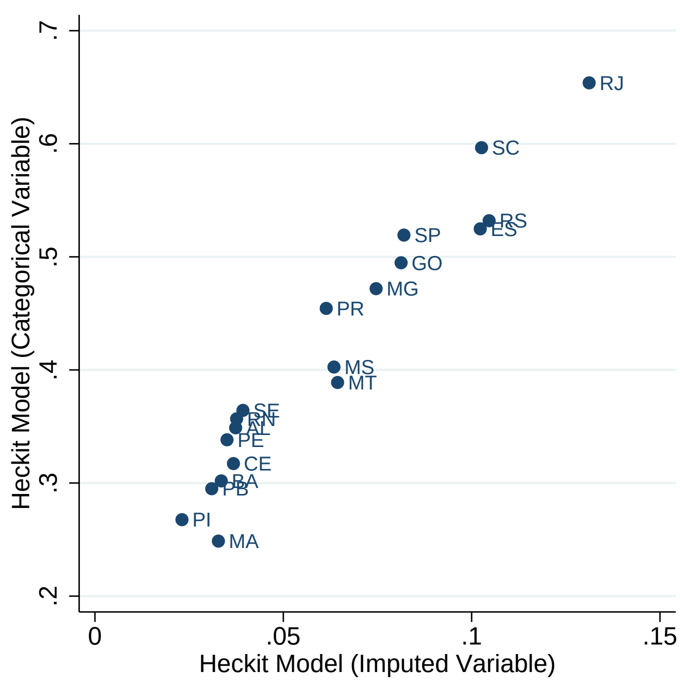

Figure 4: Returns to schooling estimates (imputed and categorical schooling variable)

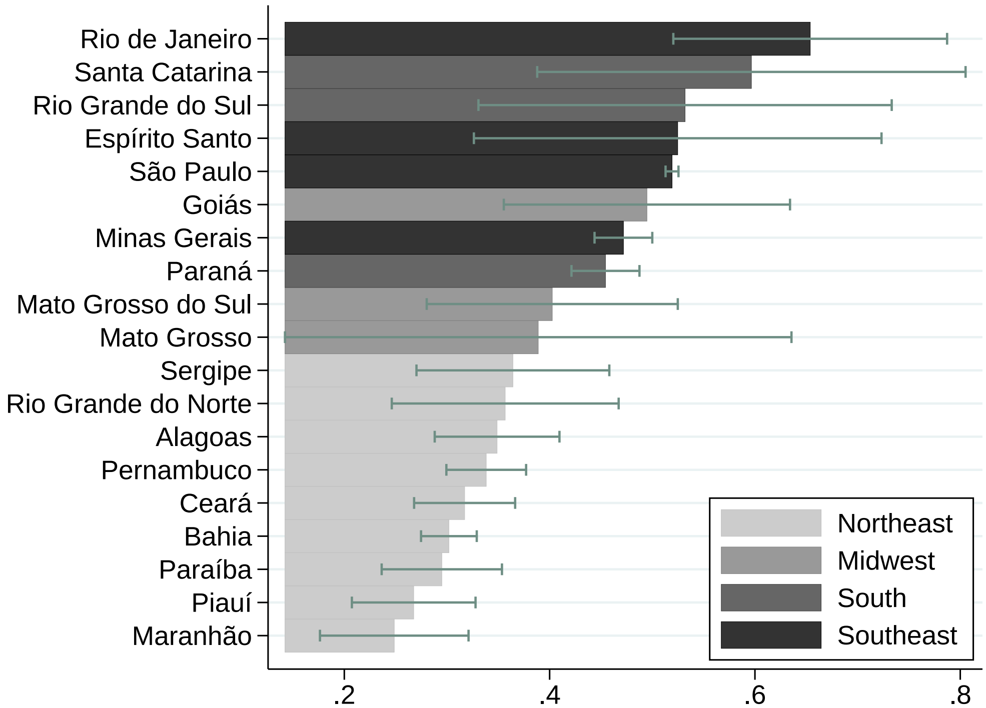

Figure 5: Heckit estimates and 95% confidence intervals (categorical schooling variable)

Figure 5 display returns estimates. Confidence intervals are significantly larger in this case because we estimate 160 additional parameters. Besides this, standard errors increase when we compute the weighted average of marginal effects.

4 Saeb test scores

In this section we compare returns to schooling, interpreted as educational quality measures, with Saeb test scores. We use data on Saeb exams applied in the years 1995-2005. To make an adequate comparison, we re-estimate educational returns using the subsample of individuals who were probably studying in grades for which the Saeb exams were applied. This amounts to selecting individuals with age between 24 and 32. For this subsample, there are states for which there is a very small number of migrants in São Paulo, making educational returns' standard errors very large for those cases. Because of this we exclude states for which there are less than 100 migrants, implying that Espírito Santo and Mato Grosso are not included in this analysis.

Figure 6 displays standardized returns to schooling and mean Saeb test scores, along with some correlation statistics. The two measures are highly correlated: the correlation coefficient equals 0.88 and a one standard deviation increase in return to schooling is associated to a 82.5% standard deviation increase in Saeb test scores. However, there are significant differ-

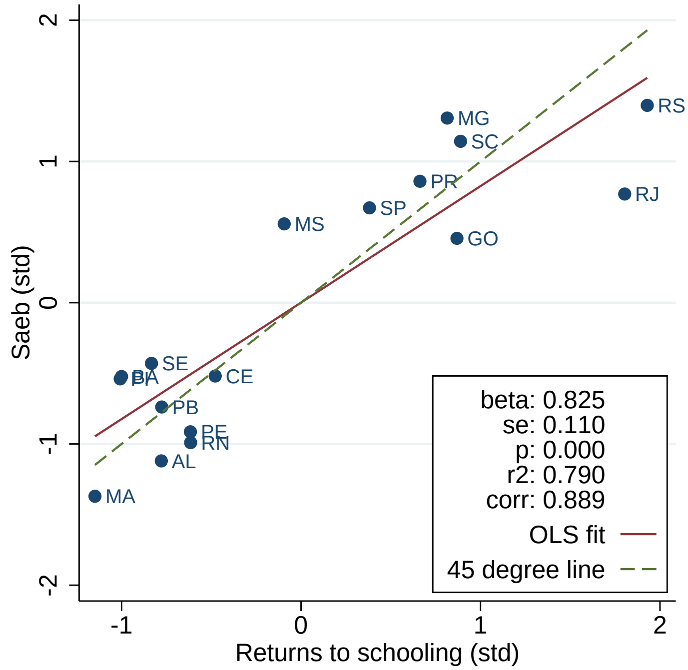

Figure 6: Returns to schooling and Saeb test scores (standardized)

ences between both indexes: the Northeast region's mean Saeb scores are equal to 91.6% of the others regions' mean. For the case of our educational quality index, this number equals 30.2%. That is, our measure suggests a larger discrepancy between regions' educational qualities than Saeb scores do. If we think of our index as the value that market forces assign to education, this is evidence that there are educational components that the market captures, but test scores do not.

Despite differences, the high correlation between both indexes presents evidence that returns to schooling can be used as a proxy variable for educational quality in the case where the latter is not available. For some countries, educational quality measures are scarce, whereas data on earnings and schooling are easily available. Therefore, verifying the correlation between returns to schooling and test scores is relevant for researchers interested in investigating education themes in developing countries through the construction of education quality measures.

5 Conclusion

In this paper we provide a new measure of educational quality of Brazilian states in 2010, based on the idea that the financial returns obtained from an additional year of schooling can be seen as derived from the value that the market assigns to this education. Using a strategy similar to Schoellman's (2012), we use census data on migrants in São Paulo in order to investigate returns to schooling of individuals who obtained education in different states, but who work in the same labor market.

We find that educational quality is very heterogeneous across states. Our index implies that educational quality is more unequal across states than 1995-2005 Saeb test scores imply. This is a relevant result for the debate on educational quality in Brazil. Additionally, our index displays high correlation with Saeb test scores, presenting evidence that returns to schooling can be used as proxy variable for educational quality in the case where the latter is not available. This is relevant, for example, for researchers interested in studying education themes in developing countries.

References

A. Alesina and P. Giuliano. The power of the family. *Journal of Economic Growth*, 15(2): 93–125, 2010. ISSN 1573-7020. doi: 10.1007/s10887-010-9052-z. URL http://dx.doi.org/10.1007/s10887-010-9052-z.

A. Banerjee and E. Duflo. *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. PublicAffairs, 2012.

F. H. Barbosa Filho and S. Pessôa. Retorno da educação no Brasil. *Pesquisa e Planejamento Econômico*, 38(1), 2008.

B. Bratsberg and D. Terrell. School quality and returs to education of U.S. immigrants. *Economic Inquiry*, 40(2):177–198, 2002. ISSN 1465-7295. doi: 10.1093/ei/40.2.177. URL http://dx.doi.org/10.1093/ei/40.2.177.

M. Campos, J. Fullgraf, and V. Wiggers. A qualidade da educação infantil brasileira: alguns resultados de pesquisa. *Cadernos de Pesquisa*, 36:87–128, 4 2006. ISSN 0100-1574. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0100-15742006000100005&nrm=iso.

D. Card and A. Krueger. Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy*, 100(1):1–40, 1992. URL http://EconPapers.repec.org/RePEc:ucp:jpolec:v:100:y:1992:i:1:p:1-40.

B. Chiswick and P. Miller. The effects of school quality in the origin on the payoff to schooling for immigrants. IZA Discussion Papers 5075, Institute for the Study of Labor (IZA), 2010. URL http://EconPapers.repec.org/RePEc:iza:izadps:dp5075.

P. Ferreira and C. Santos. Migração e distribuição regional de renda no Brasil. *Pesquisa e Planejamento Econômico*, 37(3), 12 2007.

J. Garen. The returns to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica*, 52(5):1199–1218, 1984.

Z. Griliches. Estimating the returns to schooling: Some econometric problems. *Econometrica*, 45(1):1–22, 1977. URL http://EconPapers.repec.org/RePEc:ecm:emetrp:v:45:y:1977:i:1:p:1-22.

E. Hanushek and L. Wobmann. *Education Quality and Economic Growth*. The World Bank, 2007.

J. Heckman. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement*, 5(4):475–492, 1976.

J. Heckman. Sample bias as a specification error. *Econometrica*, 47(1):153–162, 1979.

P. Jacinto and L. Rodeghiero. Retorno em escolaridade: um estudo para a região metropolitana de Porto Alegre. *Revista Brasileira de Estudos Regionais e Urbanos*, 2012. ISSN 0034-7140.

N. Kaarsen. Cross-country differences in the quality of schooling. *Journal of Development Economics*, 107:215–224, 2014. ISSN 0304-3878. doi: http://dx.doi.org/10.1016/j.jdeveco.2013.11.005. URL http://www.sciencedirect.com/science/article/pii/S0304387813001673.

A. Krueger and M. Lindahl. Education for growth: Why and for whom? Working Paper 7591, National Bureau of Economic Research, March 2000. URL http://www.nber.org/papers/w7591.

Q. Li and A. Sweetman. The quality of immigrant source country educational outcomes: Do they matter in the receiving country? CReAM Discussion Paper Series 1332, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London, Dec 2013. URL https://ideas.repec.org/p/crm/wpaper/1332.html.

S. Moriconi and G. Peri. Country-specific preferences and employment rates in Europe. CREA Discussion Paper Series 15-11, Center for Research in Economic Analysis, University of Luxembourg, 2015. URL https://ideas.repec.org/p/luc/wpaper/15-11.html.

D. Okoye. Education policy and rate of return estimates. *Working Paper*, 2012.

M Resende and R. Wylie. Retornos para educação no Brasil: Evidências empíricas adicionais. *Economia Aplicada*, 10(3):349–365, 2006.

A. Sachsida, P. Loureiro, and M. Mendonça. Um estudo sobre retorno em escolaridade no Brasil. *Revista Brasileira de Economia*, 58:249–265, 2004. ISSN 0034-7140. URL http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-71402004000200006&nrm=iso.

T. Schoellman. Education quality and development accounting. *The Review of Economic Studies*, 79(1):388–417, 2012.

A. Sen. *Development as Freedom*. Anchor, 2000.

B. Sianesi and J. van Reenen. The returns to education: Macroeconomics. *Journal of Economic Surveys*, 17(2):157–200, 2003. ISSN 1467-6419. doi: 10.1111/1467-6419.00192. URL http://dx.doi.org/10.1111/1467-6419.00192.