São Paulo School of
ECONOMICS

Working Paper

6/2015

Dec 2015

CENTER FOR APPLIED MICROECONOMICS

Estimating the returns to education using a parametric control function approach: evidences for a developing country

André Portela Souza
Eduardo Zylberstajn

Estimating the returns to education using a parametric control function approach: evidences for a developing country

André Portela Souza
São Paulo School of Economics and Center for Applied Microeconomic Studies - FGV[1]

Eduardo Zylberstajn
São Paulo School of Economics - FGV and Fipe[2]

December 5, 2015

Abstract

This paper investigates the causal effect of education on earnings in Brazil by employing a new method proposed by Klein and Vella (2010) that obtains identification on the presence of conditional heteroskedasticity. In contrast to traditionally used IV methods, this approach yields unbiased estimates in the absence of instruments and allows for interpretation of the coefficients that are not confined to local average treatment effects. Results indicate that the average return to education in Brazil was relatively stable around 14% from 1995 to 2003, declined afterwards until 2011 and has remained at around 11.4% since then. The results suggest that the OLS estimations are biased downward and we interpret this bias as a sign of under-education premiums that are likely to occur in environments where the more talented ones are dropped from school and moved into the labor market earlier in life.

Keywords: return to education, wage equation, control function approach, under-education premium

JEL Classification: C3, I21, J31

---

[1] Corresponding author: email: andre.portela.souza@fgv.br; tel.: +55 11 3799-3358; fax: +55 11 3799-3350.

[2] ezylberstajn@fipe.org.br

The authors are thankful to André Portela Souza, Bruno Oliva, Cristine Pinto, Guilherme Stein, Hélio Zylberstajn, Naércio Menezes-Filho, Priscilla Albuquerque Tavares, Reynaldo Fernandes and Vladimir Ponczek for helpful comments and suggestions.

1. Introduction

Earnings inequalities have declined during the past decade in most of Latin American countries. In Brazil, the Gini coefficient declined from 0.567 in 2001 to 0.495 in 2013[3]. Still, Brazil remains a very unequal country, and recent developments suggest that this trend might be at risk: the Gini coefficient has been stable for the last three years.

Earnings inequality level in a given country is the result of the distribution of the individual productivity characteristics and the returns of these attributes in the population. In particular, the returns to education play a key role to determine earnings and are likely one of the most explored issues in the empirical economics literature. However, to these days several empirical challenges make it difficult to consistently estimate the returns to schooling. This is, in turn, a serious problem researchers and policy makers have to address in order to better understand income inequality dynamics and to foster investments in human capital.

This paper employs a novel technique developed by Klein and Vella (2010)[4] to estimate the causal effect of schooling on wages in Brazil for the period 1995 to 2013. The difficulty in identifying the causal effects of education on earnings arise from the endogeneity of educational choices to wages. Since Mincer (1974) established a methodology to estimate wage equations, several authors have documented and tried to deal with the OLS bias, usually by employing instrumental variables (IV) strategies (see, for instance, Griliches, 1977; Angrist and Krueger, 1991; Duflo, 2001; Heckman et al. 2006; Card, 1999, in particular, presents a detailed survey on this subject).

Although large in quantity, the empirical literature lacks in robustness of results. Angrist and Krueger (1991) estimate returns to schooling in the US labor market between 6-10%, above the OLS estimates of 5-7%. Oreopoulos (2007) estimates that one extra year of education yields an average increase of 13% in wages, compared to a downward biased OLS estimate of 8%. Carneiro and Lee (2008) and Chen (2008) also make use of IV to estimate an average return to education between 13-15% for American men, but the validity of the instruments they use (proxies of the costs of school attendance) is questioned by authors such as Cameron and Taber (2004).

In Brazil, several studies have estimated the returns to education (see, for instance, Langoni, 1973; Senna, 1976; Tannen, 1991; Barros and Reis, 1991), but the difficulty in dealing with endogeneity has been a constant issue. More recently, Teixeira and Menezes-Filho (2012) use a national survey with data from 1997 to 2007 and employ an IV strategy to find that returns were between 5.5-9.4%, lower than their OLS estimate of 11.6%.

The KV (2010) strategy uses a control function approach. Identification relies on heterosckedasticity of the error terms, which yields nonlinearity to the control term, allowing for non-biased estimates of the coefficient of interest. As Saniter (2012) points out, this approach makes use of second moment restrictions (variance) instead of first moment exclusion restrictions (inherent to any IV approach) and therefore has two key advantages. First, interpretation of the coefficient of interest is possible for the entire population of interest (i.e. inference is made on the average treatment effect – ATE), while

[3] According to Instituto Brasileiro de Geografia e Estatística (IBGE).
[4] Henceforth KV (2010).

any IV estimate only provides the causal effect for the subsample of individuals who are affected by variations in the instrument (compliers). Second, because there is no need for instruments, inference can be made at any moment in time. On the other hand, IV estimates are often also bounded by the time of the variation induced by the instrument. An example of such case is found in Ichino and Winter-Ebmer (2004), who use own and father's World War II involvement as instruments for schooling; thus, their results are estimates for a particular cohort of Germans and cannot be replicated for previous or subsequent cohorts.

However, the implementation of the KV (2010) estimator is not simple. On the contrary, it is difficult because of their semiparametric approach, which brings a heavy computational burden to the estimation procedure. We then turn to Farré, Klein and Vella (2013), who implement a parametric estimator for the KV identification strategy. Using a sample drawn from the National Longitudinal Survey of Youth (NLSY 1979), they find that accounting for endogeneity increase the estimate returns to schooling from 6.8 to 11.2%. They interpret their findings as compatible with "the over education penalty" hypothesis. There are factors associated with some individuals becoming over educated (actual education level above the predicted education) that are penalized in the labor market.

The contributions of this paper to the literature are at least twofold. First, this is the first time, to our knowledge, that evidences of returns to education in a developing country are estimated by the control function approach. The comparison of the results is interesting in itself. Although the results may be qualitatively similar between developed and developing countries, the nature of the underlying process may be different. Indeed, to provide a preview of the results found in this paper, the estimated average return to education in Brazil fell from 14.2% in 1995 to 11.4% in 2013. The OLS estimates are downward biased in the entire period by 2.3 percentage points on average. We interpret our finding as compatible with the "under education premium" hypothesis, where underlying factors and/or individual unobserved characteristics associated with under education are positively associated with higher wages in the labor market. Developing countries are characterized by factors associated with lower education attainment such as the presence of credit constraints, low school quality, high returns to labor market experience for young adults, etc. Horowitz and Wang (2004), for instance, show that under credit constraint families specialize the time allocation of their children. It is possible that higher ability children are selected out of the school and into the labor market earlier in life. We provide some indirect evidence of this hypothesis.

Second, our findings contribute to the debate of the causes of the inequality decrease in Latin America in general, and in Brazil in particular. It has been well documented that Latin American countries have experienced decreased inequalities in labor earnings recently and such trends are mainly explained by the decrease in the returns to education (e.g., Azevedo et al. (2013); Lopez-Calva and Lustig (2010); Lustig et al. (2013); and Manacorda et al. (2010)). In general, their estimates assume exogeneity of schooling. Our findings provide robust estimates of declining returns to education from 1995 to 2013 and thus corroborate their conclusions.

These results are presented in greater detail in section 4. First, section 2 presents the econometric model and the implementation procedure of the estimator. Section 3 describes the data used in the estimation. Section 5 discusses in greater depth the OLV-IV bias and section 6 presents some evidence to support the robustness of our findings. Section 7 concludes.

2. Empirical exercise

2.1 Econometric model and identification

In this section we describe the KV approach. The ultimate goal of any linear model that estimates the causal effect of education on earnings is to obtain consistent estimates for the parameter δ in a wage equation of the following kind:

$$w_i = x_i\beta + \delta s_i + u_i \tag{1}$$

Where w_i represents hourly log wages, s_i represents years of schooling, x_i is a vector of exogenous regressors and u_i the error term for individual *i*.

OLS estimation of (1) is inconsistent because of the endogeneity bias caused by omitted variables correlated to both wage and schooling (e.g., individual abilities). To better illustrate KV proposed strategy, let's first write the education equation and then rewrite the model in a control function setting:

$$s_i = x_i\varphi + v_i \tag{2}$$

where x_i may be (but not necessarily is) identical for both equations. v_i the error term for individual *i*. From this point on, we refer to (1) as the wage equation and to (2) as the education equation.

If there is endogeneity the covariance between u_i and v_i is different from zero. Then we can write $u_i = \lambda v_i + e_i$, where $cov(v_i, e_i) = 0$. Replacing this expression in equations (1) and (2) provides us the controlled function:

$$w_i = x_i\beta + \delta s_i + \lambda v_i + e_i \tag{3}$$

where v_i is the control term and λ is a measure of the degree of endogeneity. Because v_i is a perfect combination of s_i and x_i, the regressors are collinear and OLS is thus infeasible.

However, we can also write:

$$\lambda = \frac{cov(u_i, v_i)}{var(v_i)} = \frac{cov(u_i, v_i)}{\sigma_v \sigma_v}\frac{\sigma_u}{\sigma_u} = \rho\frac{\sigma_u}{\sigma_v} \tag{4}$$

where σ_j (j=u, v) denotes the standard deviations of the error terms u and v and $\rho = cov(u_i, v_i)/\sigma_u\sigma_v$ is the correlation coefficient between them.

If the assumption that the errors are heteroskedastic holds, then $\sigma_j = H_j(x_i^j)$, where H_j is the heteroskedasticity function and $x_j \subseteq x$. Equation (3) can now be rewritten as:

$$w_i = x_i\beta + \delta s_i + \rho\frac{H_u(x_i^u)}{H_v(x_i^v)}v_i + e_i \tag{5}$$

which is the final estimation equation. Notice that x_i^u and x_i^v may be different or identical in both equations.

There are two important identification assumptions in order to properly estimate the equation above. First, $H_u(x_u)/H_v(x_v)$ needs to vary across x, so that regressors are not collinear. KV (2010) has named this assumption as the *variable impact property* (VIP), which simply requires that the heteroskedasticity is present in either one of both equations in such a way that the quotient of the two functions is not constant across x. A good candidate (but not the only one) for providing this is age. What is needed is that the heteroskedasticity due to age in the education equation is different than the one in the wage equation, which is economically plausible: on the one hand, schooling in Brazil has been expanding throughout cohorts and being universalized; on the other hand, we expect heterosckedasticity in the wage errors to increase as age increases due to heterogeneous experience and human capital accumulation (other than education). Also, we include in the regressors state dummies. For the wage equation we include indicator variables for the current state of residence. Local labor market conditions vary across states and this may affect not only the wage levels but the wage dispersion as well. For the schooling equation, we add the indicator variables for the state of birth. They are proxies for different school attainment costs both direct (school accessibility) and indirect (opportunity costs to attend school such as child labor wages). Again, they may affect both education levels and dispersion. We show evidence in support of this in section 3.

The second assumption is that the errors correlation must be constant and independent of the regressors, i.e. $\rho = corr(u_i, v_i|x_i) = corr(u_i, v_i) = const$. KV (2010) refer to this as the *constant correlation condition* (CCC). Put differently, the CCC requires that the degree of endogeneity in the model does not depend on x. Formally, the error terms u_i and v_i are assumed to have a multiplicative structure composed by a heteroskedastic part and an homoskedastic part. Let the $u_i^*$and v_i^* be respective homoskedastic error part. The error terms are defined as follows:

$$u_i = H_u(x_i^u)u_i^* \text{ and } v_i = H_v(x_i^v)v_i^*. \tag{6}$$

The correlation coefficient ρ can be written as

$$\rho = \frac{cov(u_i,v_i)}{H_u(x_i^u)H_v(x_i^v)} = \frac{H_u(x_i^u)H_v(x_i^v)cov(u_i^*,v_i^*)}{H_u(x_i^u)H_v(x_i^v)} = cov(u_i^*, v_i^*). \tag{7}$$

In this setting, KV interpret u_i^* and v_i^* as measures of unobserved abilities. The CCC assumption imposes that after conditioning out on x_i´s the return to unobserved ability is constant. The heterogeneity of the returns to unobserved abilities comes entirely from the heterogeneity of the individuals socio-economic characteristics described in the x_i´s.

2.2 Implementation

The implementation of the KV (2010) estimator can be done either parametrically and non-parametrically. Notably, a non-parametric approach has the advantage of not imposing normality of errors for consistency nor a functional form for the heteroskedasticity functions. KV (2010) describe a semi-parametric method, for which they also provide proof of consistency; their method is implemented by Saniter (2012), Klein and Vella (2009), Schroeder (2010) and Farré, Klein and Vella (2009). Nonetheless, all these papers describe the computational burden of adopting such strategy. Saniter (2012), for instance, who estimated the returns to schooling in Germany, reports having used 500 computer cores at the same time.

Perhaps because of these implementation difficulties, the KV (2010) estimator has been used in only a few studies. Aside from the studies mentioned above, Schroeder (2010) estimates the impact of microcredit borrowing on per-capita household consumption in Bangladesh using conditional second moments. Wang (2010) adapts the KV (2010) estimator to a Chinese database and finds that the returns to education in urban China are in the range 2.3-4.6%, below the OLS estimates of 3.9%-7.3%. Figueirêdo et al. (2014) use the KV (2010) approach to assess the influence of family background in student achievement in ENEM, a Brazilian national exam that is undertaken when students graduate from high school.

Farré, Klein and Vella (2013) follow a different path. They choose a functional form for the heteroskedasticity functions (that take an exponential form), thus largely simplifying the implementation of the estimator. They also simulate a Monte Carlo exercise in which the true heteroskedastic functions are exponential but the estimation is done parametrically assuming either hetereoskedasticity in only one of the equations or a quadratic form for the heteroskedastic functions. In both cases, results show that the estimator performs quite well[5].

The heavy computational demand in the nonparametric estimation arises for reasons that become clear below and we will highlight them when the method is described. In order to avoid the nonparametric burden, we adopt the same parametric approach proposed by FKV (2013), treating $H_u(x_i^u)$ and $H_v(x_i^v)$ as exponential functions of the individuals' observable characteristics. More specifically, we define the following heteroskedastic function:

$$H_j{}^2(x_j) = \exp(z\theta_j) \tag{8}$$

[5] Also using simulation techniques, Klein and Vella (2010) show that even with little heteroskedasticity the OLS bias is eliminated. These results are encouraging to the use of the estimation technique described in this section.

This approach makes the estimation feasible, with the risk of losing efficiency due to mis-specification of the functional form. It should be noted that z may or may not be equal to x. In our benchmark specification, we use all individual characteristics in both equations. The only subtle change that we make is the use of state of residence dummies in the wage equation and state of birth dummies in the education equation. We provide some robustness checks in section 6.

Following FKV (2013), estimation is done following a three-step procedure:

1) Estimate $\hat{v}$ through OLS in the education equation (Eq. (2)): $\hat{v} = s - x\hat{\varphi}_{OLS}$.

2) Using a Poisson regression[6], estimate θ_v from Eq. (8) by regressing $\hat{v}^2$ on z. Then, compute the standard deviation of the reduced form error as $\widehat{H}_v = \sqrt{\exp(z\hat{\theta}_v)}$.

3) Using $\hat{v}$ and $\widehat{H}_v$ and the assumed form for $H_u = \exp(z\theta_u)$, solve the following nonlinear least squares problem[7]:

$$\min_{\beta,\rho,\delta,\theta_u} \sum_{i}^{n} \left(e_i = w_i - x_i\beta - \delta s_i - \rho \frac{\sqrt{\exp(z_i\theta_u)}}{\widehat{H}_{vi}} \hat{v}_i \right)^2 \tag{9}$$

Solving equation (9) yields the coefficient estimates of interest, particularly $\hat{\delta}$.[8] Estimation was performed for several years of data. Standard errors are defined as square roots of the diagonal elements of the covariance matrix, which in turn is obtained numerically (see Nocedal, 1996).

[6] FKV (2013) estimate $\ln(\hat{v}^2)$ using OLS. In our case, the error is not normally distributed, making the traditional log-linear approach inappropriate to recover $\widehat{H_v^2}$. To overcome this issue, we use a Poisson regression that fits a model $H_j{}^2(x_j) = \exp(z\theta_j + \varepsilon)$, in line with our specification in equation (6).

[7] Once again we remind that in our benchmark specification we have $z = x$ except for state of birth dummies.

[8] The optimization was programmed in SAS, using the NLP procedure (Non Linear Programming) and the default Conjugate Gradient Methods (CONGRA) optimization technique. Convergence in step 3 was typically achieved with 300 iterations or less. Notice that if the estimation was to be done non- (or semi-) parametrically, step 3 implies that instead of estimating $\hat{\theta}_u$, one needs to estimate the unknown functions H_j at each iteration, which is precisely the main computational burden in the non-parametric strategy. On the other hand, by imposing the functional form $H_{ij} = \sqrt{\exp(z_i\theta_j + \varepsilon_i)}$ the optimization of equation (7) is done in approximately two minutes using a common server with standard CPU and RAM specifications.

3. Data

The analysis presented in the next sections are based on the *Pesquisa Nacional por Amostra de Domicílios* (National Household Sample Survey, henceforth PNAD) for the years of 1995 to 2013. PNAD is an annual household survey conducted by the *Instituto Brasileiro de Geografia e Estatística* (IBGE) and it is a nationally representative sample. In the most recent wave (2013), 362,555 individuals were surveyed.

We adopt a parsimonious specification, so we use few variables and compatibilization of different waves is straightforward[9]. Because our interest lies on returns to schooling, we restrained our sample to individuals between 25 and 55 years of age (to minimize censored schooling and selection into the labor market biases). Also, we have excluded individuals that did not report any labor income in the week of reference. Individuals with missing values in any of the variables used in the estimation were also dropped. We considered the log hourly wage from the main job as the labor income variable and dropped the top and bottom 1% observations to exclude outliers. Table 1 provides an overview of all the variables used in this study and Table 2 displays some summary statistics for the 2013 dataset.

Table 1 – Variables description

Variable	Description
LnWage	Log of hourly wage
YrsEduc	Years of education
Age	Years of age at interview
Female	Dummy indicator with value 1 if female and 0 otherwise
White	Dummy indicator for white individuals

Source: PNAD 2013 (IBGE).

Table 2 – Descriptive statistics

Variable	Mean	S.D.	Min	Max
LnWage	3.39	0.79	1.28	6.21
YrsEduc	9.23	4.39	0	17
Age	38.64	8.55	25	55
Female	0.4310	-	0	1
White	0.4432	-	0	1
N	109,687			

Source: PNAD 2013 (IBGE).

To summarize and facilitate the understanding of our empirical strategy, the final estimation models are outlined below, in the order in which they are estimated:

$$s_i = \varphi_0 + \varphi_1 Age_i + \varphi_2 AgeSq_i + \varphi_3 Female_i + \varphi_4 White_i + \sum_{m=1}^{26} \tau_m dsb_{im} + v_i \qquad (10)$$

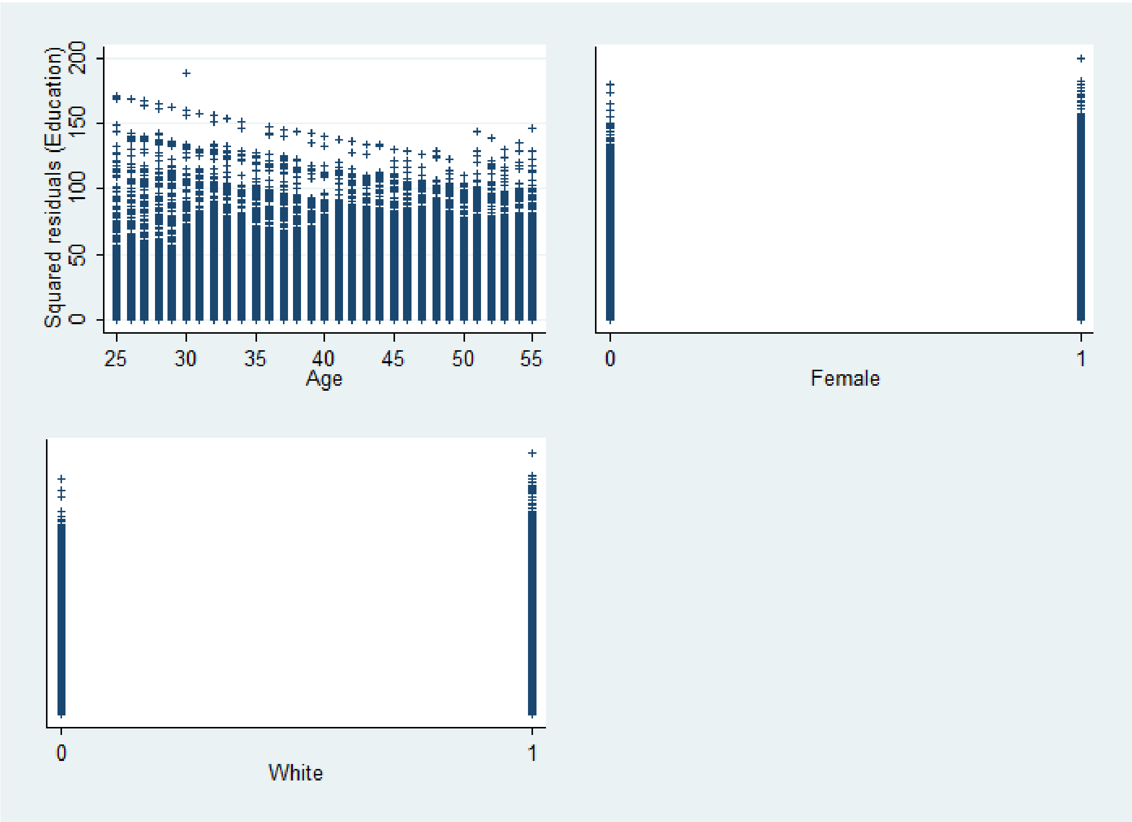
[9] We used the *DataZoom* Stata package from PUC-Rio in order to construct each year's dataset. Available at http://www.econ.puc-rio.br/datazoom.

$$H_v{}^2(x_i) = \hat{v}_i{}^2 = \exp(\theta_{v0} + \theta_{v1}Age_i + \theta_{v2}AgeSq_i + \theta_{v3}Female_i + \theta_{v4}White_i + \sum_{m=1}^{26} \psi_m dsb_{im} + \varepsilon_{vi}) \tag{11}$$

$$LnWage_i = \beta_0 + \delta s_i + \beta_1 Age_i + \beta_2 AgeSq_i + \beta_3 Female_i + \beta_4 White_i + \sum_{m=1}^{26} \gamma_m dsr_{im} + \rho \frac{\sqrt{\exp(z_i\theta_u)}}{\hat{H}_{vi}} \hat{v}_i + e_i \tag{12}$$

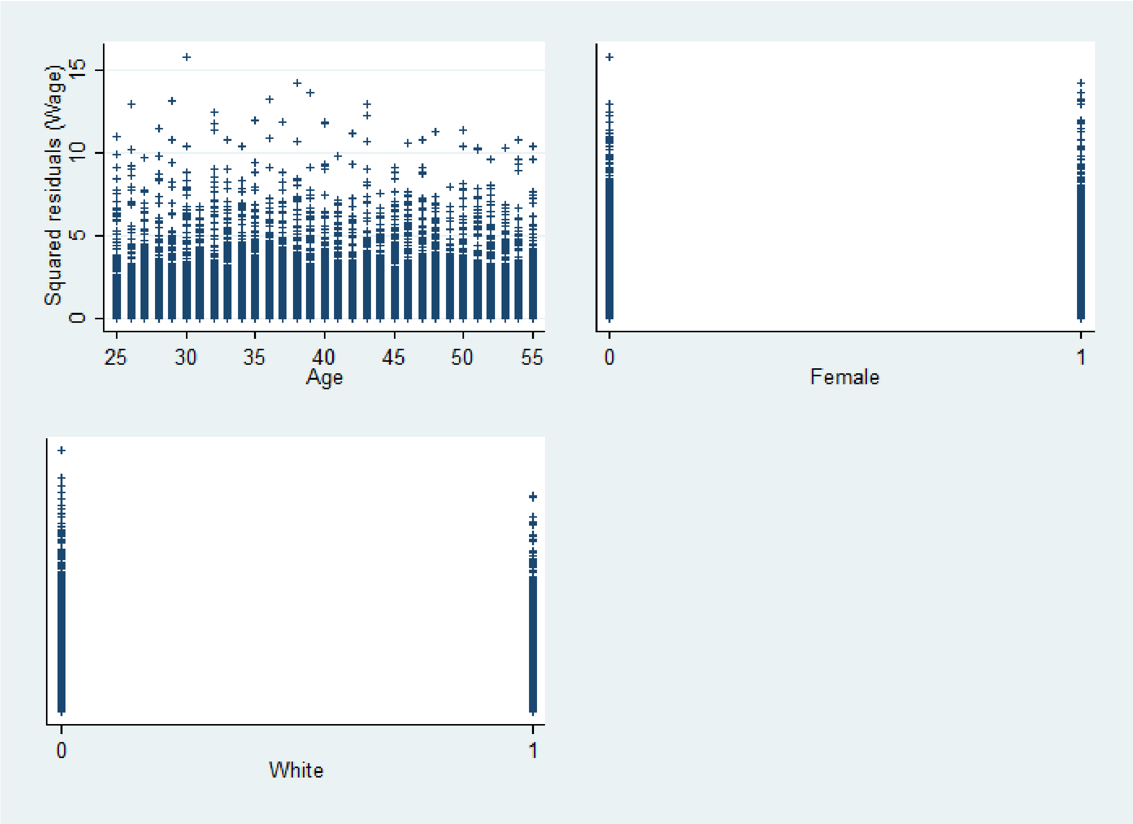
where dsb_{im} and dsr_{im} are state of birth and state of residence dummies respectively and $z_i\theta_u = \theta_{u0} + \theta_{u1}Age_i + \theta_{u2}AgeSq_i + \theta_{u3}Female_i + \theta_{u4}White_i + \theta_{u5}Female_i \cdot Age_i + \theta_{u6}Female_i \cdot AgeSq_i + \theta_{u7}White_i \cdot Age_i + \theta_{u8}White_i \cdot AgeSq_i + \theta_{u9}White_i \cdot Female_i + \sum_{m=1}^{26} \varsigma_m dsb_{im}$. Note that state of birth dummies were used in the education equation and in the $H_j{}^2(x_i)$ functions (we consider them exogenous), while state of residence were included as controls for local labor markets in the wage equation. In short, we cautiously adopted a parsimonious specification, with only a small number of strictly exogenous covariates. Equations (8) and (10) where estimated using OLS and equation (9) was estimated by fitting a Poisson regression (see footnote 6 above).

Figure 1 – Education equation: graphical analysis of heteroskedasticity



Source: Author's calculations based on PNAD 2013 (IBGE).
Note: residuals are obtained after the OLS estimation of equation (2).

Figure 2 – Wage equation: graphical analysis of heteroskedasticity



Source: Author's calculations based on PNAD 2013 (IBGE).
Note: residuals are obtained after the optimization of equation (9).

Figure 1 and Figure 2 above provide visual evidence for the presence of heteroskedasticity and, more importantly, also suggest that heteroskedasticity is not constant across x. See, for instance, that the slope of the relationship between age and the squared residuals $\hat{v}^2$ (Figure 1) and $\hat{u}^2$ (Figure 2) are different, as is the case with the female dummy. This is crucial because, as highlighted earlier, identification relies on these differences.

4. Results

The three-step procedure outlined in section 2 yields the results displayed in Table 3 and Table 4. We begin displaying in Table 3 the results for the estimation of the education equation, which yielded estimates for $\hat{v}$. The results show that older people are less educated than younger people in 2013. This is a consistent finding for all the years and greatly reflects the fact that younger cohorts were more exposed to the expansion of the public school system in the end of the second half of the 20th century. Moreover, older and non-white individuals are less educated. Also of interest is the formal test for the presence of heteroskedastic errors in the education equation, displayed in the Table A.1. in the appendix. The test strongly rejects the hypothesis of homoscedastic errors, which is in favor of the identification strategy adopted in this paper. The estimates for $\hat{\theta}_v$ are also displayed in the appendix in the Table A.2.

Table 3 – OLS Estimates: Education equation (2013)

Variable	β_{OLS}
AGE/100	-11.3409***
	(1.4290)
$(\text{AGE}/100)^2$	2.0272
	(1.8029)
FEMALE	1.4820***
	(0.0249)
WHITE	1.5377***
	(0.0273)
CONSTANT	14.1465***
	(0.2887)
N	109,687
Adjusted R^2	0.1342

Source: Authors' calculations based on PNAD (IBGE). Standard errors in parentheses.

Note: OLS estimates for equation (8). State of birth dummies are also included in the regression.

** significant at 10%; ** significant at 5%; *** significant at 1%.*

Next, we turn to the wage equation. Table 4 shows the differences between the OLS and the control function estimates for β, with standard errors displayed in parenthesis. As for the coefficients, we see that age affects wages in a linear fashion in the CF setting and that men and white workers are expected to earn more. The coefficients are, in most cases, smaller in the CF setting.

The main result of this paper is displayed in the first row, where we find an average return of 9.2% for each extra year of education in the OLS case and a 11.4% in the controlled setting, meaning that the OLS model is thus biased downward by 2.2 percentage points. The estimates for $\hat{\theta}_u$ are displayed in the appendix in the Table A.3. The other coefficients have none or small changes in the controlled function setting, with the exception of the age coefficient, which is higher in the OLS estimate.

The same exercise is replicated for the years of 1995 to 2012. Remember that the KV (2010) approach allows to calculate the returns to schooling at any period (independent of any instrument) and for all individuals, not only compliers. The KV (2010) estimator, as we have argued before, measures the ATE, while any IV strategy yields a LATE.

For simplicity, results are displayed graphically in Figure 3[10]. Each year represents the equivalent to the first row of Table 4, which depicts the estimates for 2013. A clear pattern of declining returns to education emerges, either in the OLS or the CF setting. Moreover, when comparing 1995 to 2013 one finds similar declines in return rates: 3.0 and 2.8 respectively.

[10] All reported coefficients are significant at the 1% level.

Table 4 – OLS and Control function estimates: Wage equation (2013)

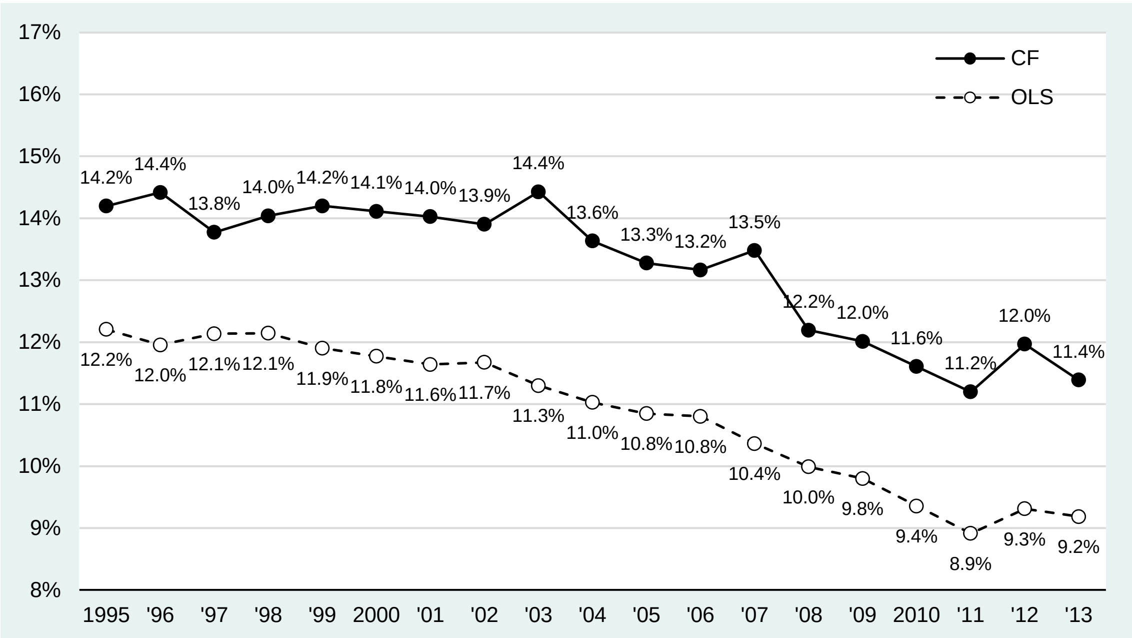
	β_{OLS}	β_{CF}
YrsEduc	0.0919***	0.1139***
	(0.0005)	(0.0010)
Age/100	4.6322***	3.4823***
	(0.2257)	(0.2253)
$(\text{Age}/100)^2$	-4.0935***	-2.3851
	(0.2847)	(0.2838)
Female	-0.2571***	-0.2879***
	(0.0040)	(0.0042)
White	0.1423***	0.1067***
	(0.0044)	(0.0045)
Constant	1.8217***	1.7953***
	(0.0451)	(0.0464)
ρ		-0.0686***
		(0.0289)
N	109,687	109,687
Adjusted R^2	0.3377	0.3437

Source: Authors' calculations based on PNAD (IBGE). Standard errors in parentheses.

Note: OLS estimates for equation (1) and CF estimates for equation (10). State of birth dummies are also included in the regression.

** significant at 10%; ** significant at 5%; *** significant at 1%.*

Figure 3 – Returns to education in Brazil: 1995-2013



Source: Authors' calculations based on PNAD (IBGE).

Note: Each line depicts the evolution of δ in equations (1) for OLS and (9) for the control function estimates. Estimations are done separately for each year. All coefficients are significant at the 1% level in all years. In the last years of each decade, IBGE does not carry PNAD: values for 2000 and 2010 are simple linear interpolations.

Menezes-Filho, Fernandes and Picchetti (2006) and Tavares and Menezes-Filho (2011) have already documented the decline in returns to schooling in the past decade and its importance to the decline in earnings inequality, even though they are unable to properly account for endogeneity. Fernandes and Menezes-Filho (2012) present evidence that this phenomenon might be related to the sharp increase in the relative supply of medium and high-skilled workers. This also has been found by Manacorda et al. (2010). It seems a reasonable explanation, but arguably further research is needed, especially to identify the reasons for the recent increase in the returns.

The results presented thus far are not directly comparable to other studies that attempt to measure the causal impact of education on wages in Brazil. Menezes-Filho, Fernandes and Pichetti (2006) report returns of 14% in 1997 and Tavares and Menezes-Filho (2011) report declining returns between 1995 and 2009, when they reached slightly less than 12%. Both studies rely on repeated cross-sections to estimate average returns to schooling. Teixeira and Menezes-Filho (2012), on the other hand, use an IV approach to estimate a much lower return of 5.5% per year of education, leading to a conclusion of an upward bias in OLS estimates. The instruments they use are the number of schools in the state and year when the individual was born and an educational law passed in 1971, which unified primary education in Brazil. While their results seem to contradict the ones presented in this paper, they can be reconciled taking into account their LATE interpretation. As argued by Imbens and Angrist (1994), the coefficients estimated by means of IV represent the causal effect only for the subsample of compliers (the individuals who were *actually* affected by variations in the instrument). It seems sound to assume that for Teixeira and Menezes-Filho (2012) the compliers are individuals with lower levels of education, so their estimates are not directly comparable to the ones presented in this section. In order to make results comparable, we would need to be able to identify the same compliers in our sample and then estimate the returns to education only for that subsample.

The results presented in this paper also seem reasonable when compared internationally. Psacharopoulo and Patrinos (1994) argue that, overall, the international average of the Mincerian return to schooling is 10% and that it is higher in middle and low income countries due to diminishing marginal returns. Interestingly, three recent studies apply similar approaches to ours. Klein and Vella (2009) find a return of 10% for Australian workers and Saniter (2012) estimated and average return of 8.5% in Germany (both studies estimate H_j semiparametrically). Farré, Klein and Vella (2013) employ the same parametric methodology we used in this paper and find an average return of 11.2% for the US using the NLSY79 database. All three studies find that the OLS estimates are biased downwards, as our results also suggest. As extensively documented in the literature, it seems reasonable that in Brazil, a middle-income country, returns to schooling are higher than those in Australia, Germany and the US. Less obvious, but also of interest, is the fact that the OLS estimates are biased downwards in these studies and in our results as well.

5. Discussion: the OLS-IV gap

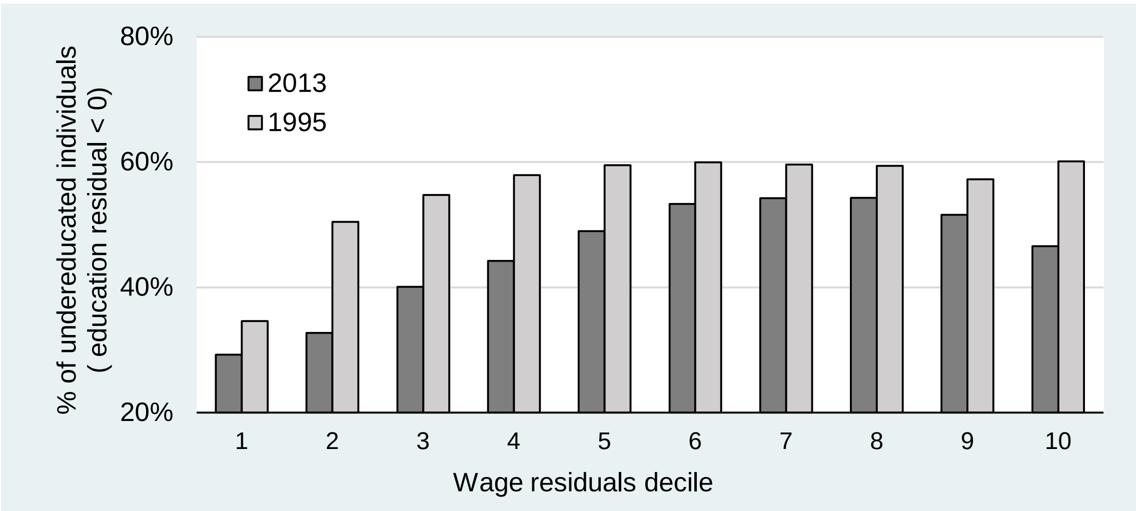
In this section we shed some light on the direction of the OLS bias that was found and reported earlier. The negative estimate for ρ implies that the OLS estimate will be smaller than the one obtained in the controlled function setting.

The OLS-IV gap has been extensively discussed in the literature. Some early papers suggest that the omitted variable bias arises because of unaccounted ability, which in turn would produce an upward bias in OLS estimates. Other authors have claimed that measurement errors in the education variable would produce downward biased OLS estimates (see Angrist and Krueger 1991; Card 1995, 1999; Cameron and Taber 2004).

A third trend in the literature, however, claims that aside from ability, the error component of the education equation captures other factors, such as motivation, that would lead individuals to obtain what Vella and Gregory (1996) call 'over-education'. This over achievement would, in turn, yield lower returns to schooling because returns to over education are lower than the average returns to education (Dolton and Vignoles 2000; Groot and van den Brink 2000; Rubb 2002; Farré, Klein and Vella 2013), yielding what has been called the 'over education penalty'. This penalty is an interpretation for the negative value of ρ. If the correlation between u and v is negative, than if one has a higher education than the one predicted by the education model, she will likely have a lower than expected wage. In a developing country context, however, it might be useful to look at the other half of the coin.

If the claim that there is an over-education penalty is valid, then for the same reasons one might expect an 'under-education premium'. This means that if the education of one individual is lower than expected, she probably earns more than what is predicted by the wage model.

Figure 4 – Share of Undereducated Individuals by Wage Residual Deciles (1995 and 2013)



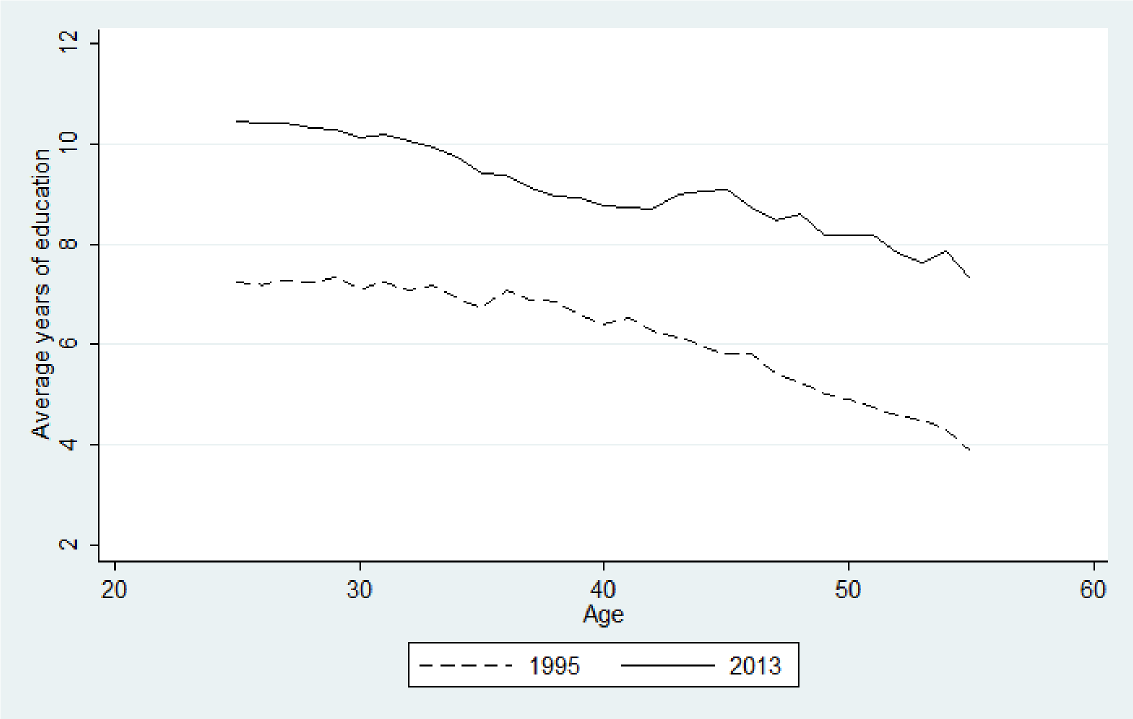
Source: Authors' calculations based on PNAD (IBGE). Education residuals are the $\widehat{v's}$; the wage residuals are given by $\hat{u}$, which is calculated as the difference between observed and predicted log wages predicted with the estimate of δ.

In fact, we do observe in our data that individuals with education levels above the predicted levels (the undereducated ones) are the ones with greater unpredicted wages. A negative (positive) residual means that the expected education or wage is higher (lower) than what is observed. Over- (under-) education means that $\mathrm{v} > 0$ ($\mathrm{v} < 0$), and a wage premium (penalty) means that $\mathrm{u} > 0$ ($\mathrm{u} < 0$). Figure 4 above presents the share of undereducated individuals by wage residual deciles for 1995 and 2013. The lower deciles have a disproportionally greater share of overeducated whereas the higher deciles have a disproportionally share of undereducated individuals.

As KV (2013), we interpret u_i^* and v_i^* as measures of unobserved abilities. The contribution of the unobserved abilities to wages and schooling depends on the individual´s socioeconomic characteristics. However, after conditioning on the socioeconomic factors the returns to unobserved ability is constant. These assumptions are plausible in developing country contexts. There is a large body of empirical literature for developing countries and Brazil in particular showing that families specialize the time allocation of their children across activities such as working and schooling (e.g., Emerson and Souza (2007, 2008), Horowitz and Souza (2010), among others). These findings are consistent with models of intrahousehold decisions among poor families. Dahan and Gaviria (2003) construct a model that shows that families may treat their children unequally even when they are identical if returns to human capital increase with the level of human capital, and parent's decisions are based on efficiency consideration. Horowitz and Wang (2004) develop a model of heterogeneous children. They show that when the ability differences across children are great, families may reverse specialize such that the more talented child is allocated in the labor market earlier and the less talented one goes to school. In contexts such as this, one may observe "over education penalties" and "under education premiums" as presented in the Figure 4 above.

Interestingly, the change in the shares of undereducated ones by wage residuals deciles from 1995 to 2013 may be explained by the schooling attainment expansion observed across the cohorts as shown in Figure 5 below.

Figure 5 – Average years of Education by Age (1995 and 2013)



Source: Authors' calculations based on PNAD (IBGE).

6. Robustness check

In this section we replicate the estimation of the returns to schooling using different approaches to examine the robustness of the findings presented in previous sections. Two different tests are performed and results are displayed in Table 6.

The estimates presented in the second columns use a less rigid form for H_v and H_u, replacing the age and age squared variables by 5-year interval dummies. In the third column, we relax the multiplicative structure imposed by equation (6) by estimating a different version of equation (8), $H_j^2(x_j) = \exp(z\theta_j) + \vartheta$, which can be estimated by OLS instead of Poisson regression. In this alternative setting, however, we lose the previous interpretation of ρ.

These tests are relevant because the identification strategy employed earlier in the estimation requires that the variables used in each specification of the equations generate enough and consistent heteroskedasticity. Thus, changing the functional form and the set of variables used in each heteroskedasticity function are key tests for the robustness of our findings.

The estimates for these alternate specifications are 10.8% and 11.3%, slightly below the 11.4% estimate obtained under the benchmark specification. Note, however, that the difference between all these estimates are within the 95% confidence interval of each other. Thus, we can state that the results presented earlier are unaffected by the use of different functional forms for heteroskedasticity.

Table 6 - Control function estimates with alternative specifications: Wage equation (2013)

	$\boldsymbol{\beta_{CF}}$ **(bechmark especification)**	$\boldsymbol{\beta_{CF}}$ **(dummies specification)**	$\boldsymbol{\beta_{CF}}$ **(exponential heteroskedasticity)**
YRSEDUC	0.1139***	0.1081***	0.1129***
	(0.0010)	(0.0041)	(0.0010)
ρ	-0.0686***	-0.0279	-0.0704*
	(0.0289)	(0.0433)	(0.0374)
N	109,687	109,687	109,687
Adjusted R^2	0.3437	0.3070	0.3436

Source: Authors' calculations based on PNAD (IBGE). Standard errors in parentheses.

Note: Column 1 reports estimates for equation (10). Results reported in column 2 uses a more flexible specification for H_j with only dummy variables, while in column 3 H_j is modeled as an exponential function of age, age squared, gender and race. Standard errors in parentheses.

** significant at 10%; ** significant at 5%; *** significant at 1%.*

7. Conclusion

This paper estimates the causal returns to education for the Brazilian population during the period 1995-2013. The naïve OLS regression of earnings on years of schooling yields estimates with the well-known endogeneity bias. Klein and Vella (2010) developed a control function setting in which heteroskedasticity provides identification without the need for exclusion restrictions. The key advantage of this approach is the fact that estimation can be done for the entire population at any point in time, allowing for a more general interpretation than the IV's LATE.

One possible drawback of the KV (2010) method is the computational demands that arise due to their semiparametric estimators. Farré, Klein and Vella (2013) propose a fully parametric approach that allows for the implementation of the KV (2010) estimator in practice. We apply this parametric approach and find that the average return to education have declined in Brazil from 15.6% in 1995 to 11.7% in 2011 and since then has bounced back to 12.7% in 2013. These estimates are higher than the OLS estimated coefficients suggest, pointing to a downward bias in the OLS estimation. We interpret this bias as a sign of under-education premiums that are likely to occur in environments where the more talented ones are dropped from school and moved into the labor market earlier in life. Finally, we also find a decline in the returns to schooling during the period. It seems to be associated with the well-documented increase in the supply of more educated workers observed in the past two decades in Brazil.

References

Angrist J. and Krueger, A. B. (1991) "Does compulsory school attendance affect schooling and earnings". Quarterly Journal of Economics 106:979–1014

Azevedo, J.P. & Inchaust, G. & Sanfelice, V. (2013). Decomposing the Recent Inequality Decline in Latin America. World Bank Policy Research, Working Paper n. 6715.

Barros, R. and Reis, J. G.A. (1991). Wage inequality and the distribution of education : A study of the evolution of regional differences in inequality in metropolitan Brazil. Journal of Development Economics, 36(1), 117-143.

Cameron, S. and Taber, C. (2004) "Estimation of educational borrowing constraints using returns to schooling". Journal of Political Economy 112:132–182.

Card, D. (1995) "Earnings, schooling and ability revisited". In: Polachek, S. W. (ed) Research in labor economics. JAI Press, Greenwich, Connecticut, pp 23–48

Card, D. (1999) "The causal effect of education on earnings". In: Ashenfelter O, Card D (eds) Handbook of labor economics, vol 3A, Chap. 30. Elsevier Science/North-Holland, Amsterdam.

Carneiro P, Lee S (2008) "Changes in college enrolment and wage inequality: distinguishing price and composition effects", UCL Working Paper.

Chen, S. H. (2008) "Estimating the variance of wages in the presence of selection and unobserved heterogeneity". Review of Economics and Statistics, 90(2):275–289.

Daha, M.; Gaviria, A. (2003). "Parental Actions and Siblings' Inequality". Journal of Development Economics 72, October 2003: 281-297.

Dolton P, Vignoles A (2000) "The incidence and effects of overeducation in the UK graduate labour market". Econ Educ Rev 19(2):179–198

Duflo E (2001) "Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual policy experiment". American Economic Review, 91(4):795–813.

Emerson, P.; Souza, A.P. (2008). "Birth Order, Child Labor, and School Attendance in Brazil". World Development, v. 36, p. 1647-1644.

Emerson, P.; Souza, A.P. (2007). Child Labor, School Attendance and Intra-Household Gender Bias in Brazil. The World Bank Economic Review (Print), v. 21, p. 301-316.

Farré, L;. Klein, R. and Vella, F. (2013). "A Parametric control function approach to estimating the returns to schooling in the absence of exclusion restrictions: an application to the NLSY". Empirical Economics 44:111-113.

Figueirêdo, Erik, Nogueiray, Lauro and Santanaz, Fernanda Leite. (2014) "Igualdade de Oportunidades: Analisando o Papel das Circunstâncias no Desempenho do ENEM". Rev. Bras. Econ., Rio de Janeiro , v. 68, n. 3, p. 373-392.

Griliches, Zvi (1977). "Estimating the Returns to Schooling: Some Econometric Problems". In: Econometrica 45.1, pp. 1-22.

Groot W, van den Brink HM (2000) “Overeducation in the labor market: a meta-analysis”. Econ Educ Rev 19(2):149–158

Heckman J, Urzua S, Vytlacil E (2006) “Understanding instrumental variables in models with essential heterogeneity”. Rev Econ Stat 88(3):389–432.

Horowitz, A. and Wang, J. (2004). “Favorite son? Specialized child laborers and students in poor LDC households”. Journal of Development Economics, 73: 631-642.

Horowitz, A.; Souza, A. P (2010). “The impact of parental income on the intra-household distribution of school attainment?A measurement strategy and evidence”. The Quarterly Review of Economics and Finance, v. 51, p. 1-18.

Ichino, Andrea and Rudolf Winter-Ebmer (2004). “The Long-Run Educational Cost of World War II”. In: Journal of Labor Economics 22.1.

Imbens, Guido W. and Joshua D. Angrist (1994). “Identification and Estimation of Local Average Treatment Effects”. In: Econometrica, pp. 467-475.

Klein, Roger and Vella, Francis (2010). “Estimating a class of triangular simultaneous equations models without exclusion restrictions,” Journal of Econometrics, Elsevier, vol. 154(2), pages 154-164, February.

Lopez-Calva, L. F., & Lustig, N. (2010). Declining inequality in Latin America: A decade of progress?. Washington, DC: Brookings Institution and UNDP.

Lustig, N.; Lopez-Calva, L. & Ortiz-Juarez, E. (2013). Declining Inequality in Latin America in the 2000s: The Cases of Argentina, Brazil, and Mexico. World Development, 44, 129-141.

Manacorda, M., Sanchez-Paramo, C., & Schady, N. (2010). Changes in returns to education in Latin America: The role of demand and supply of skills. Industrial and Labor Relations Review, 63(2), 307–326.

Menezes-Filho, Naercio Aquino, Fernandes, Reynaldo, and Picchetti, Paulo. (2006). “Rising human capital but constant inequality: the education composition effect in Brazil”. Revista Brasileira de Economia, 60(4), 407-424.

Mincer, Jacob (1974). “Schooling, Experience, and Earnings”. New York: Columbia University Press.

Nocedal, J. (1996). Conjugate gradient methods and nonlinear optimization. Linear and Nonlinear Conjugate Gradient-Related Methods, 9-23.

Oreopoulos, Philip (2007). “Do Dropouts Drop Out Too Soon? Wealth, Health and Happiness from Compulsory Schooling”. In: Journal of Public Economics 91.11-12, pp. 2213-2229.

Psacharopoulos, George, and Harry Anthony Patrinos*. "Returns to investment in education: a further update." Education economics 12.2 (2004): 111-134.

Saniter, Nils, 2012. “Estimating Heterogeneous Returns to Education in Germany via Conditional Heteroskedasticity,” IZA Discussion Papers 6813, Institute for the Study of Labor (IZA).

Rubb S (2002) “Overeducation in the labor market: a comment and re-analysis of a meta-analysis”. Economics of Education Review 22(6):621–629

Schroeder, Elizabeth (2010). “The Impact of Microcredit Borrowing on Household Consumption in Bangladesh”. Department of Economics, Georgetown University.

Tavares, P. A. and Menezes-Filho, N. A. (2011). Human Capital and the Recent Decline of Earnings Inequality in Brazil. Brazilian Review of Econometrics, 31(2), 27.

Teixeira, Wladimir Machado and Menezes-Filho, Naércio Aquino. (2010). “Estimando o retorno à educação do Brasil considerando a legislação educacional brasileira como um instrumento”. Revista de Economia Política (Impresso), v. 32, p. 479-496, 2012

Appendix

Table A.1 – Heteroskedasticity test: Education equation (2013)

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity H_0: Constant variance		
Variable	**Chi^2**	**p**
AGE/100	2179.33	0.0000
$(AGE/100)^2$	2064.07	0.0000
FEMALE	102.19	0.0000
WHITE	20.47	0.0000
Simultaneous	3144.31	0.0000

Source: Authors' calculations based on PNAD 2013 (IBGE). State of birth dummies were also included in the regression.

Table A.2 – Heteroskedasticity test: Wage equation (2013)

Breusch-Pagan/Cook-Weisberg test for heteroskedasticity H_0: Constant variance		
Variable	**Chi^2**	**p**
YRSEDUC	739.12	0.0000
AGE/100	699.67	0.0000
$(AGE/100)^2$	670.53	0.0000
FEMALE	0.30	0.5861
WHITE	239.84	0.0000
Simultaneous	1929.10	0.0000

Source: Authors' calculations based on PNAD 2013 (IBGE). State dummies were also included in the regression.

Table A.3 –Heteroskedasticity functions: education and wage equations (2013)

	$\boldsymbol{\theta_v}$	$\boldsymbol{\theta_u}$
AGE/100	9.4979***	-0,0002***
	(0.4597)	(0,0000)
$(AGE/100)^2$	-8.9398***	0,0000
	(0.5646)	(0,0000)
FEMALE	-0.0804***	-0,0856
	(0.0078)	(0,3107)
WHITE	0.0234***	-0,4219***
	(0.0085)	(0,0235)
CONSTANT	-0.3762***	0,7054***
	(0.1008)	(0,0737)
N	109,687	109,687
Pseudo R^2	0.0381	0.3352

Source: Authors' calculations based on PNAD 2013 (IBGE). Standard errors in parentheses.

Note: Poisson regression estimates for θ_v (equation (9)) are depicted in column (i) and Nonlinear Least Squares for θ_u (equation (10)) are depicted in column (II). State of birth dummies were included in the regression.

** significant at 10%; ** significant at 5%; *** significant at 1%.*

Escola de Economia de São Paulo da Fundação Getulio Vargas FGV-EESP
www.fgvsp.br/economia

WORKING PAPER 6/2015 • DEC, 2015 •