Education and Wages in Brazil*

Chiara Binelli[†], Costas Meghir[‡] and Naercio Menezes-Filho[§]

First draft: May 2005. This draft: September 2008.

Abstract

This paper investigates the puzzling evidence on wages and educational attainment observed in Brazil in the 1990s: while returns to College have been rising by over one hundred per cent and graduation rates at Secondary and High School level have been steeply increasing, the proportion of students progressing to Higher education has been decreasing. First, we present evidence that the low supply at College is not due to lack of availability in positions. Then, we specify a joint model of education choices, labour force participation and wages with unobserved heterogeneity affecting all elements of the model. We find that the observed changes in the returns to schooling only marginally reflect variations in the ability composition by level of education. They are mainly the result of changes in the market value of education: the educational expansion at intermediate depressed the wage at this level dramatically. The supply increase at High School did not translate into a proportional increase at College level due to binding credit constraints. The size of the supply effect on changes in the wages depends on the substitutability between aggregate labor inputs. We quantify this effect by estimating a CES production function that allow for different elasticities of substitution between four labor inputs. We find that complementarities in production between workers with High School and College are responsible for about for about twelve (thirty one) per cent further decline (increase) in the relative return to high school (college) versus less than high school education with respect to the case of an isoelastic production function with no complementarities.

Key Words: Returns to Schooling, Educational Choices, Quality of Education.

JEL Codes: J24, J31, C14, C51.

*We thank seminar participants at the 2006 IZA/World Bank Conference on Employment and Development in Berlin, at the 2006 SOEGW Conference in Rimini and at the 2006 Latin American and Caribbean Economic Association Annual Conference in Mexico City for useful comments.

[†]Oxford University and Institute for Fiscal Studies. Corresponding author: c.binelli@ucl.ac.uk

[‡]University College London and Institute for Fiscal Studies.

[§]University of Sao Paulo, Brazil.

1 Introduction

Educational attainment in Brazil and its relationship with wages is characterized by a number of surprising phenomena that are not easy to reconcile. Since the 1930s there has been a dramatic increase in the number of individuals completing more than primary education and progressing to intermediate levels of schooling. In addition, there has been a large increase in those completing both secondary and high school, the latter of which is required to progress to college. Finally, college graduation increased steeply up until the early 1970s but has remained almost stagnant until the end of the 1990s. At the same time, relative returns to secondary and high school declined, while relative returns to college steadily increased.

While the decline in the returns to intermediate levels of education can be rationalized by the increase in the supply of this type of workers, it is hard to understand why the increase in the returns to college has not been accompanied by a large increase in the supply of college graduates. Given the apparent huge returns to college, any simple internal rate of return calculation would imply a large increase in college graduation.

Our aim in this paper is to attempt an understanding of these facts. We start by describing the trends in the Brazilian education system and we present evidence that there is no lack of availability in positions: the private sector of higher education has expanded and, most importantly, there is a persistent oversupply of positions there. It thus becomes apparent that positions are in fact abundant and that this is not a real constraint to the increase in college graduates, with one important caveat: the number of positions in public Universities, which are persistently oversubscribed, have not been increasing as much. It may be that it is the public Universities which offer higher education leading to high returns. However, the wage differentials between college and high school are very large and most college graduates actually originate from the private sector. Unless one believes that the returns to the public universities are astronomical, the data are consistent with substantial returns from both private and public Universities.

We next turn our attention to the determinants of education choices and the estimation of the actual returns to schooling, for each level. We consider four education levels - primary, secondary, high school and college and we specify a joint model of education choice, labour force participation and wages. The first component of the model aims at isolating the impact of education quality and local labour market conditions on the chosen level of education. We will examine the extent to which such factors may be limiting the participation in education. The second component seeks to estimate the impact of education on labour market attachment and to allow us to control for selection into the labour market when we estimate wages. The third component - wages, aims at estimating the wage differentials for each education group. By allowing for unobserved heterogeneity to affect all parts of the model, we can deal with the endogeneity of schooling choices and of participation and we can thus sort out composition effects on returns. This is of particular importance here, because the large change in participation in high school and college may have distorted substantially the returns by composition effects. Controlling for unobserved heterogeneity will allow us to measure the actual returns to education, which may be overstated between college and high school. We close the model with a competitive sector of firms that employ different types of labor to produce the aggregate output. We assume a CES production function that allows for different elasticities of substitution between the four types of labour in the model (primary, secondary, high school and college) and we use the education prices estimated from our model to identify the degree of substitutability between them. This will be key to assess the impact of changes in the supply of education on changes in relative returns, which is of particular relevance for Brazil as well as for many other Latin American Countries that were characterized by significant educational expansions in the 1990s.

The analysis developed in this paper is related to the literature on self selection into education and employment and their effects on changes in the returns to skills. Mainly due to the lack of data, up to now most of the papers have focused on developed countries. Among the first contributions are Heckman and Sedlacek (1985, 1990) that extend the original framework developed by Roy (1951) and analyze selection into industries and occupations. In particular, they show the importance of accounting for selection into work as the most successful extension of the Roy model in order to explain cross-section wage distributions and their evolution over time. Closer to our framework are Willis and Rosen (1979) and Taber (2001) who study selection into schooling and its impact on earnings. Carneiro and Lee (2006) also study the effect of self-selection into education on the evolution of wage inequality but they focus on marginal distributions while we focus on average parameters. On selection into employment and its impact on earnings' inequality important contributions include Blundell, Reed and Stoker (2003) and Blundell, Gosling, Ichimura and Meghir (2004). Jacoby and Skoufias (2002) is, to our knowledge, the only paper that directly addresses the issue of self selection into higher education in a Latin American country. The focus of their paper is on educational choices and they ignore the impact of self selection on labor market participation and wages. They develop a dynamic model of enrollment decisions into University and include unobserved heterogeneity as an additional factor driving educational choices. They estimate the model using data on Mexican youths in the 1990s and find weak evidence of selection bias.

Finally, the inclusion of the production function relates this paper to the literature on the estimation of aggregate production functions in models of human capital accumulation. Heckman, Lochner and Taber (1998) is a seminal contribution on the identification and estimation of the production function accounting for the endogeneity of the human capital inputs. Since then, several papers have estimated flexible specifications with many labor inputs. Among the most recent contributions, Gallipoli, Meghir and Violante (2008) estimate a CES production function allowing for different elasticities of substitution between three labor inputs. As for Latin America, to our knowledge, Binelli (2008) is the only contribution that estimates a production function in a model of human capital accumulation. She estimates a CES production function with three labor inputs using data for Mexico between 1987 and 2002. She finds that there are important complementarities in production between Intermediate and Higher education.

The remainder of the paper is organized as follows. Section 2 presents a description of Brazilian education system and the trends in availability of places, number of entrants and graduates at Higher Education. Section 3 describes the data. In section 4 we present the economic model that underlies the empirical analysis. Section 5 discusses model's identification and provides a description of the estimation procedure. In section 6 we apply the model to the study of earnings' inequality in Brazil during the 1990s. Using our estimates, we analyze the evolution of compositional changes by education group and the empirical importance of selection bias for the evolution of relative returns. Then, we run a simulation exercise to evaluate the impact of a policy intervention to increase the quality of schooling on graduation rates. The empirical session ends with the estimation of the aggregate production function. Section 7 gives some concluding remarks. Appendix A presents summary statistics by level of education.

2 Brazilian education system

Brazilian education system consists of three main cycles: fundamental (primary and secondary), high school and college. Fundamental education is compulsory for all individuals and lasts eight years; high school and college take, respectively, three and four to five years. All three cycles are offered by private and public institutions. Private institutions charge a fee, while the public system is free of charge at all levels. Most Universities are located in the South and Southeast regions of Brazil. Public Universities are administered locally by the State or nationally by the central government. Every State has at least one public federal University together with private Colleges and locally run Universities.

Mean Test Scores	*Public*	*Private*
High School	265.9	340.5
College	37.6	32.2

Table 1: Mean Test Scores at High School and College in 2003 (Source: Brazilian Ministry of Education)

Table 1 presents test scores at high school and University, by private and public. For high school Table 1 reports the mean math test scores from a representative sample of final year high school students in 2003. For college the results are the mean test scores from students attending thirteen representative University courses evaluated in 2004. Private high schools and public Universities produce the best test score results. This underlies the general belief that the best education track is the one which starts in the private high school sector, preparing one for entry exams in the public Universities, which are supposed to be the elite of the higher education system. Each University sets its own

Mean Test Scores			
Year	*Public High School*	*Private High School*	*Difference Private-Public*
2000	44.10	58.59	14.49
2001	36.56	53.57	17.01
2002	30.39	47.22	16.83
2003	45.00	64.00	19.00
2006	32.89	50.02	17.13
2007	48.30	68.72	20.42

Table 2: Mean Test Scores at the National Exam at the End of High School (Source: Brazilian Ministry of Education)

admission exam (called *Vestibular*) with a content depending on the applicant's area of interest. Over time the type and content of the entry exams did not change significantly while some changes occurred in the entrance requirements with the introduction of racial quotas in some States.

Table 2 presents the average score in the national exam (ENEM) in the final year of high school by private and public schools. In each year students in private high school do perform substantially better than students in public education. Also, the proficiency gap is increasing between 2000 and 2007. The quality gap between the education received in the public and in the private sector opens up at the high school level, if not before.[1] As reported by Waltenberg (2007), private high schools offer higher wages for teachers, principals, and staff and have better infra-structure. Parents pay tuition fees supplying schools with resources that allow some schools to operate in an environment that resembles that of good schools in developed countries (e.g. with well-paid teachers and staff, good libraries and computer facilities). Public schools, in turn, must cope with much more constrained budget sets. Also, studying in public schools is socially stigmatizing. Parents who can afford it, prefer to opt out of public schools and enroll their children in private schools, generally identified as "good schools".[2]

The situation in terms of quality and status of private versus public reverses at higher education: the best universities are supposed to be the public ones. For these reasons they are highly oversubscribed; while this varies from area to area and by year, there are about 9 applicants for each place at a public University. In the private sector this ratio has fallen from about 3.5 applicants for a place in 1980 to 1.5 in 2003 (see Figure 1 and

[1] Results between 1995 and 2005 taken from SAEB, a national exam that is run in the final year of primary and secondary education and provides a measure of proficiency that is comparable across grades and over time, show that there are constant gaps between private and public at the primary and secondary level as well.

[2] Oliveira and Schwartzman (2002) report a survey on the perception that Brazilians have about the quality of each type of school. When asked in what type of school they would prefer to enroll their children, 94.3 per cent of private school teachers said that they would choose private schools, while 69.9 per cent of public schools teachers themselves preferred private schools instead of public schools.

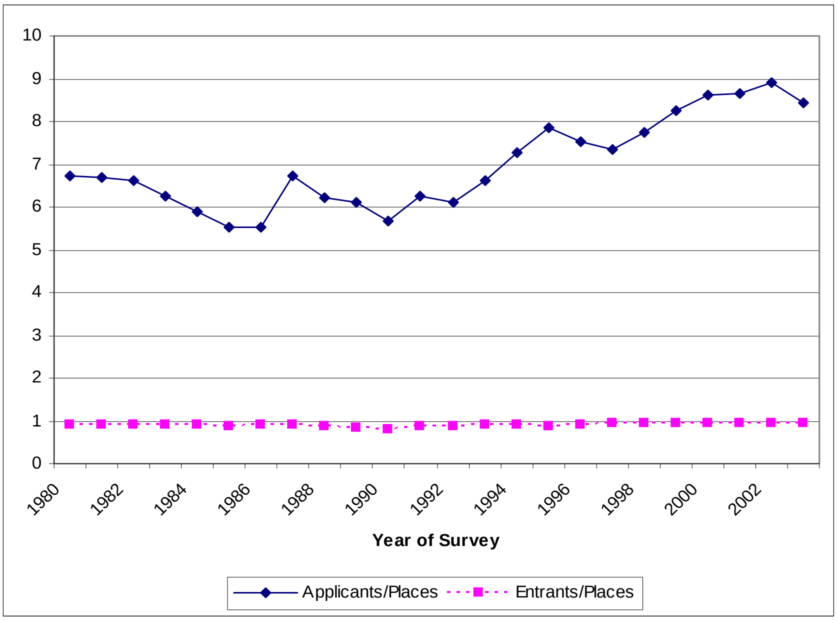

Figure 1: Ratio Applicants and Entrants Over Available Places, Public Universities (Source: Brazilian Ministry of Education)

Figure 2). What is of key interest in these figures is that the ratio of actual entrants to places for private Universities is less than one: there does not seem to be any constraint on the number of places available in the private college sector. Places at private universities are not fulfilled due to lack of demand since many applicants can not afford to pay the very expensive fees that private colleges charge.[3]

From the end of the 1990s Brazilian higher education expanded significantly through a major expansion of private Universities as well as an increase in public University admissions. As can be seen from Figure 3, between 1998 and 2002 private Universities more than doubled while the number of public Universities remained almost constant. The expansion of the higher education system has been accompanied by increasing enrolment rates at college. Figure 4 presents the evolution of the number of college entrants in the 1980s and 1990s. Enrolment rates started increasing significantly in both private and public Universities at the end of the 1990s. In 2002 the number of college entrants was almost double the one in 1992 for public Universities and almost three times the one in 1992 in the private sector. As a benchmark, the overall population grew by 42 per cent from 1980 to 2000 and 6.3 per cent from 1995 to 2000; the increase in enrollments

[3]Lack of demand accounts for most of the explanation why the number os entrants is lower than the number of official places in the private sector of education. In addition, some colleges might over-report the number of places. The autorization to increase the number of places is given by the Ministry of Education and is very bureaucratic, so some colleges could have an incentive to overstate their need in order to have room for expansion when necessary.

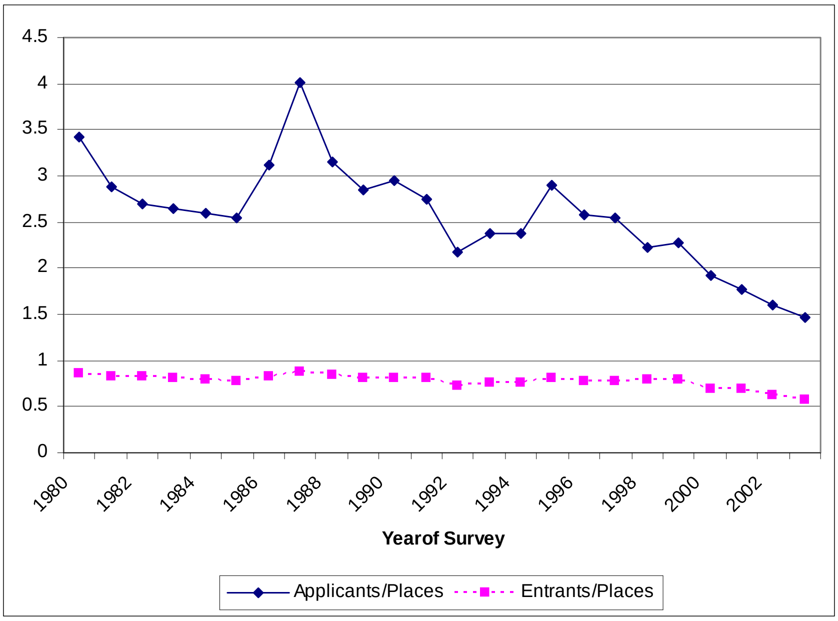

Figure 2: Ratio Applicants and Entrants Over Available Places, Private Universities (Source: Brazilian Ministry of Education)

represents a growth in participation in higher education.

Despite the increase in the number of college graduates documented in Figure 5, Brazilian higher education is characterized by relatively high drop-out rates. As shown in Figure 6, towards the end of our sample period the drop-out rate stood at 20 per cent for the public Universities and 33 per cent for the private ones. Both rates have been falling since the mid 1990s, which is consistent with the overall increase in the demand for higher education.

Since the beginning of the 1990s both the absolute number of entrants and the proportion of high school graduates over the working population significantly increased with the growth in the number of entrants almost entirely concentrated in public schools. As can be seen from Figure 7, between 1992 and 2002 the number of entrants almost tripled in the public sector while it remained stagnant in the private one. In the same years the evidence from the Brazilian National household Survey shows that the proportion of high school graduates over the adult population increased from a value of 17 per cent in 1992 to a value of around 24 per cent in 2002.

3 The data

Data on wages, employment and education come from the Brazilian National Household Survey, PNAD (*Pesquisa Nacional Por Amostra de Domicílios*), for the period 1992-2002.

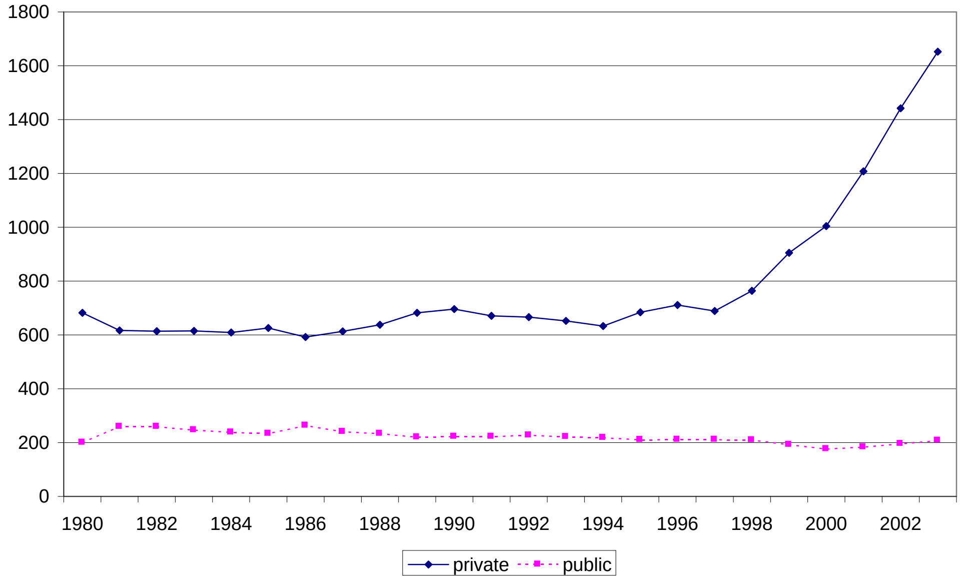

Figure 3: Number of College Institutions (Source: Brazilian Ministry of Education)

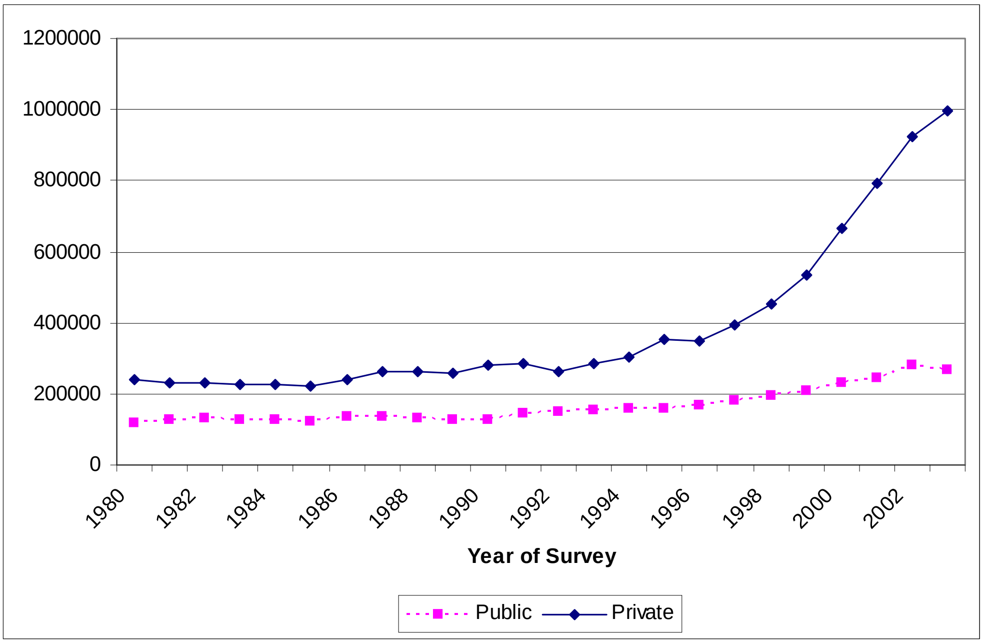

Figure 4: Number of College Entrants (Source: Brazilian Ministry of Education)

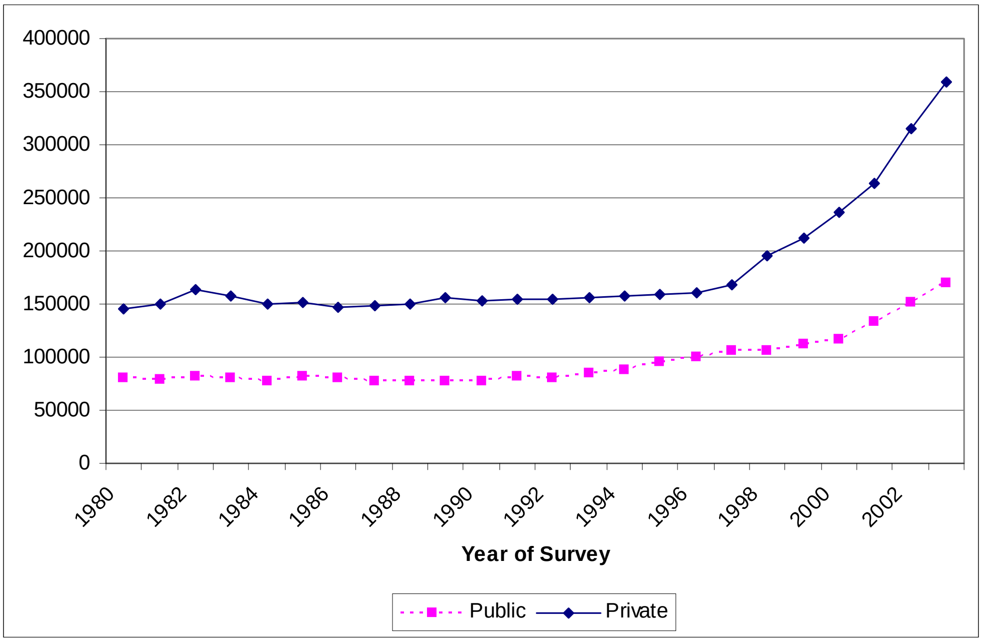

Figure 5: Number of College Graduates (Source: Brazilian Ministry of Education)

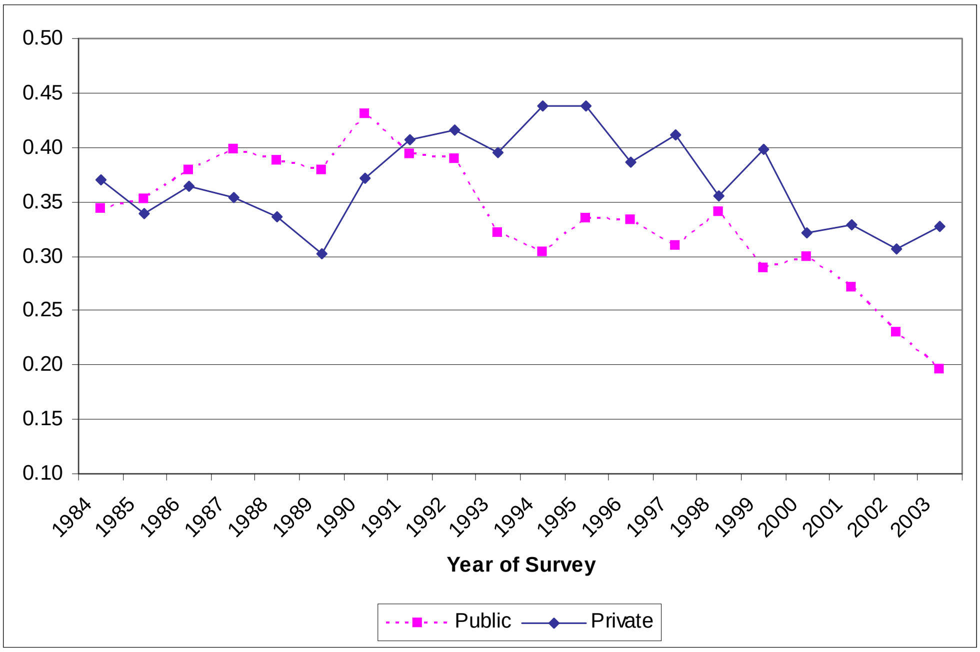

Figure 6: Drop out rate at College (Source: Brazilian National Household Survey)

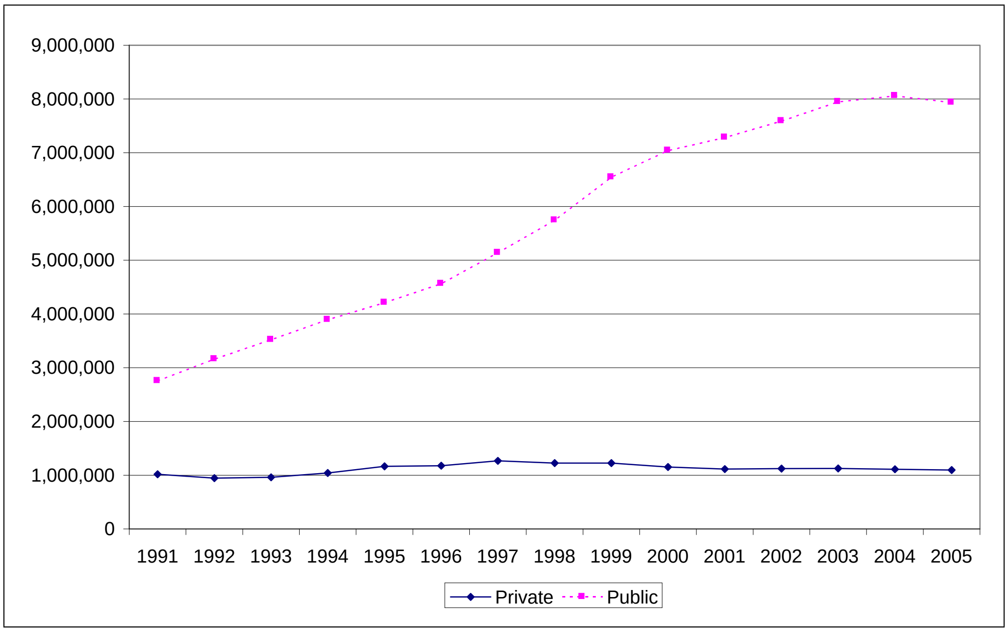

Figure 7: Number of High School Entrants (Source: Brazilian Ministry of Education)

PNAD was not conducted in 1994 due to lack of financial resources and in 2000, the year of the Brazilian population Census. PNAD is an annual, nationally representative household survey and it covers around one hundred thousand households with individual-level data on socio-demographic characteristics, labour market status, State and year of birth. Most important for our analysis, it contains accurate data on wages by level of education defined as the number of completed years of schooling. We consider the four main schooling levels in Brazil, namely primary, secondary, high school and college and accordingly construct four skill groups. We combine PNAD with data from Historical Series collected by the Brazilian statistical office[4], that provide data on the number of schools, number of teachers and teachers per school up to the high school level, population and share of national GDP by State since 1933.

Our final sample includes over one million observations on individuals aged 24-56 with matched data on availability and quality of schooling by State and year of birth. We drop data on seven States in the north of the country where household data are collected only in urban areas differently from the rest of PNAD that covers both urban and rural areas. We are left with a representative sample that includes twenty out of the twenty-seven Brazilian States.

Table 9 and 10 in Appendix A report summary statistics by level of education for the main variables used in the empirical analysis separately for males and females. As can be seen from the tables, at each level of education there are fewer female than male workers

[4] IBGE (*Instituto Brasileiro de Geografia e Estatística*).

and females receive on average a lower wage than their male counterparts. Table 11 and 12 in Appendix A present the percentage of individuals in each of the four education groups by State in the first and in the last year of the sample. There is a significant cross States variation in attainment rates that is higher in magnitude at college level and did not decrease over time. The share of individuals with college education ranges between 4 and 21 percentage points in 1992 and between 5 and 25 in 2002.

3.1 Education and wages in Brazil 1992-2002

Figure 8 and 9 show the log wage differentials over time for successive education levels between 1992 and 2002 for men and women. The picture is very similar for the two and we only highlight important differences. In 1992 the difference between the average wage of college and high school graduates was around 0.80 log points and it increased to 1.0 log points in 2002. At the same time, the observed differential between high school and secondary, as well as between the latter and primary school declined substantially, particularly for women. The college premium is substantial and much higher than what we see in developed countries. For example in the UK this stands at about 30 per cent for the average 3-year degree. In the US the figure is at about 15-20 per cent per year of college.[5]

There is a general view, supported by the test scores presented earlier, that the graduates from public Universities are better than those from the private ones. The data is not informative on what type of college a person completed. However, given that 67 per cent of students graduate from private colleges it is reasonable to believe that the college premium is very high for both groups, even if it is somewhat higher in the public sector.[6] The returns to college and to high school are similar for both males and females and move together. However, the returns to secondary school are larger for men.

In Figure 10 and 11 we present the proportion of 24-56 year old that left school, respectively, at the end of secondary and at the end of high school by year of birth. The decreasing trend in Figure 10 reflects the significant educational expansion above compulsory schooling that characterized Brazil in the last decades. The percentage of individuals that leave school after completion of secondary education declines for both sexes, at an accelerating rate for the youngest cohorts and with the largest proportional decrease for women. On the contrary, Figure 11 shows that the proportion of 24-56 year old that left education after completion of high school has significantly increased over time, for both males and females. The decline in the (relative) participation in higher education is surprising given the large and increasing observed college premium.

Finally, in the schooling choice model we will use a measure of school availability by State as a proxy for direct schooling costs, and a measure of schooling quality. The

[5] See results in Carneiro, Heckman and Vytlacil (2005).

[6] For example, if the college premium was double in the public sector relative to the private one, the latter would still stand at 75 per cent.

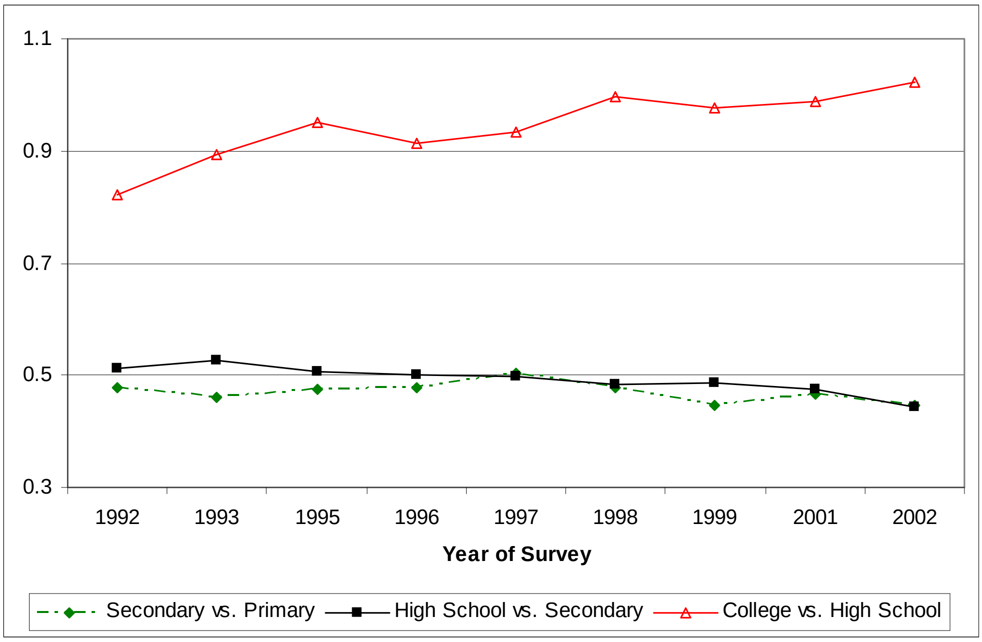

Figure 8: Log Relative Wages, Males (Source: PNAD)

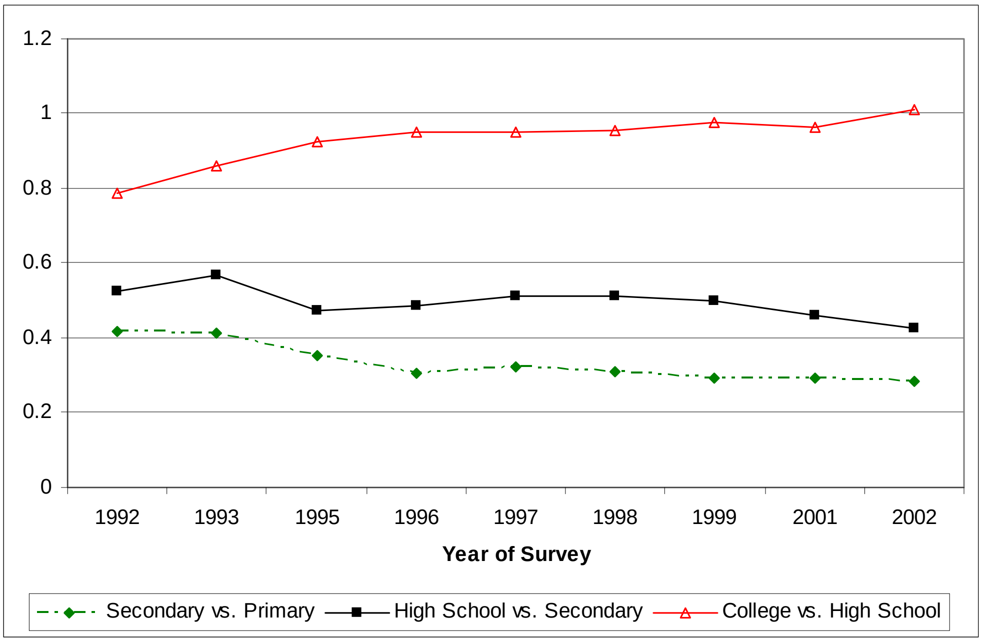

Figure 9: Log Relative Wages, Females (Source: PNAD)

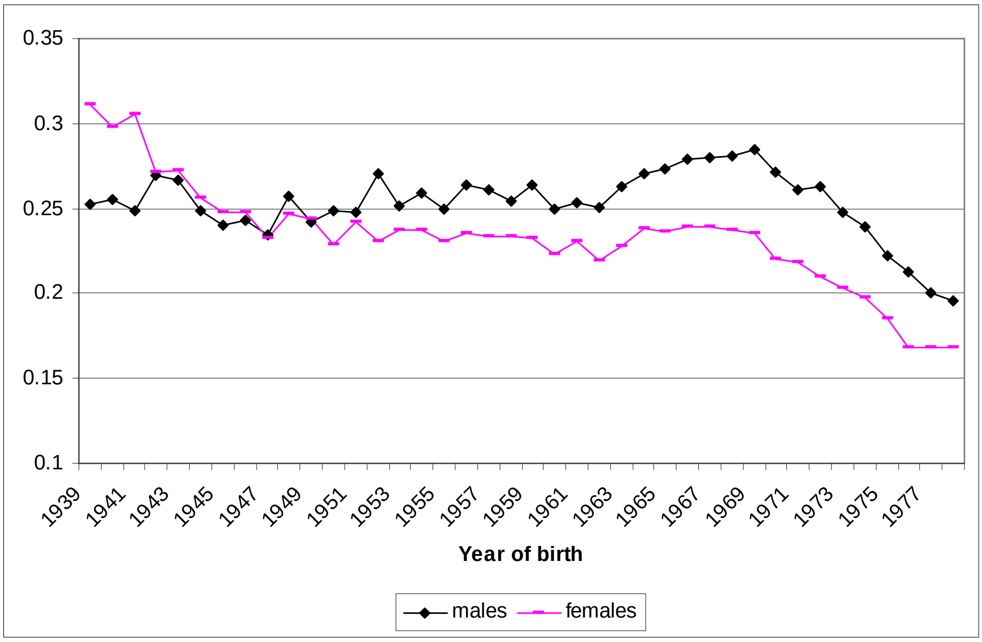

Figure 10: Proportion Secondary Graduates Not Completing Any Year of High School (Source: PNAD)

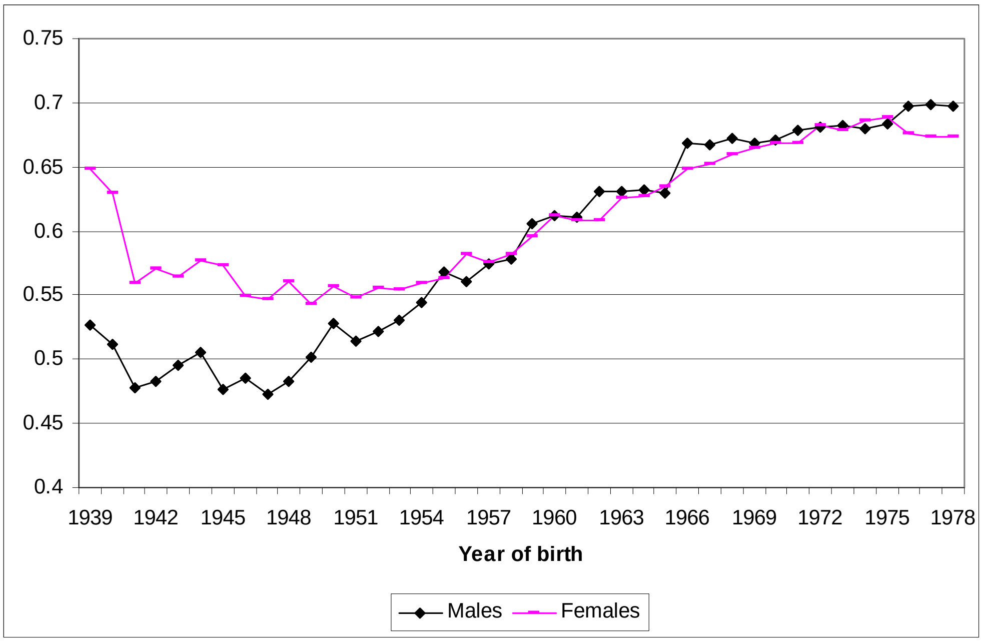

Figure 11: Proportion High School Graduates Not Completing Any Year at College (Source: PNAD)

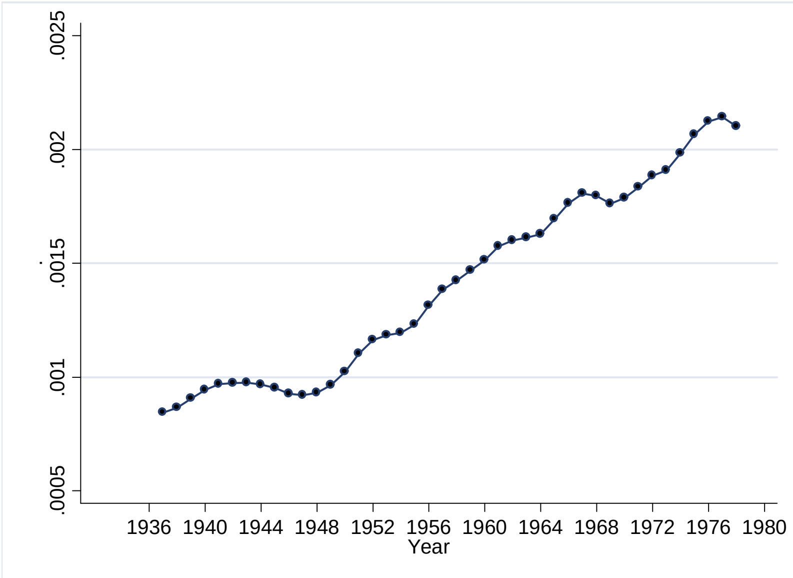

Figure 12: Number of Schools per Population (Source: Historical Series IBGE)

former we measure by the number of schools per population. The latter we measure by the number of teachers per school per population. Clearly these are imperfect measures, but they do reflect the resources devoted to education.[7] Figure 12 and 13 present these measures between 1936 and 1980 for the entire population aggregated over the whole country, to illustrate the main trends. These figures reveal a striking feature: while school availability has doubled in forty years, the resources available to each school have declined. One may infer from this that resources devoted to each pupil have declined substantially. However, this gives a mixed picture in relation to what has driven the recent increase in schooling participation: schooling costs have dropped because of the increased availability but schools may not be as effective due to the decline in quality, thus compromising the ability of pupils to qualify for college.

As a comparison, we construct the same proxies replacing the overall State population with the relevant schooling age population of youth aged between 10 and 19. Data for this variable are available only for each year at the start of a decade, so we interpolate data points between decades and look at the trend over time. The alternative proxies for both quantity and quality of schooling follow the same evolution as before with a significant deterioration of schooling quality over time. Results from the estimation of the full wage equation-schooling model will be used to investigate the importance of schooling quality

[7] An alternative measure of education quality is the annual expenditures per student by level of education. Disaggregated data by State are available from the Brazilian National Treasury Secretariat for the years 2005 and 2006. By taking the average over the Brazilian States, the correlation between annual expenditures and teachers per student is in the range of 0.40 in both years.

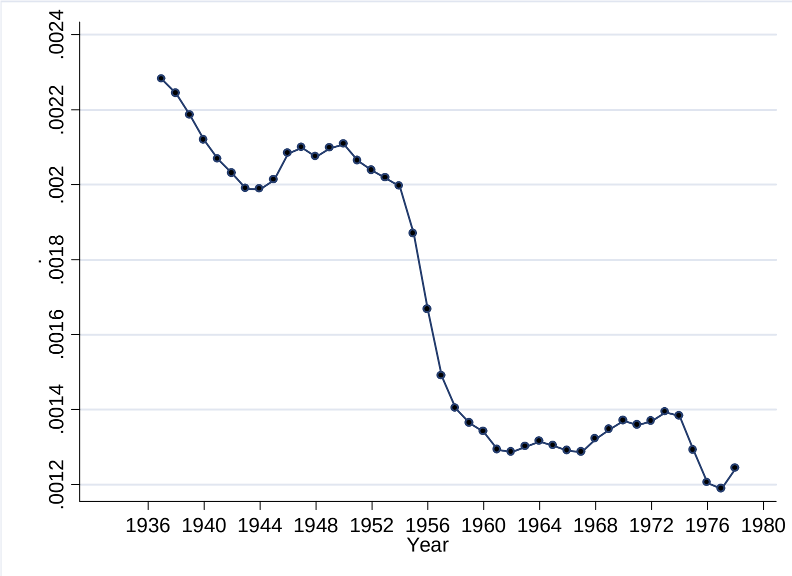

Figure 13: Number of Teachers per School per Population (Source: Historical Series IBGE)

and its impact on education choices.

4 The model

It is quite clear from the description above that there is no obvious constraint on the availability of places at college. Although the elite public Universities are heavily oversubscribed, private colleges are not. Moreover, at least at first glance, the returns to the private colleges must themselves be very high, given average returns of over 100 per cent and given that nearly 70 per cent of graduates have been educated in private colleges. In what follows we specify a model of wages and education choices to understand some of the factors influencing college attendance and to provide estimates of the returns to education net out of unobserved heterogeneity.

The model we estimate is not structural. As such we do not model expectations explicitly and we do not solve for equilibrium prices. Rather we condition on the prices we observe and on cohort and regional characteristics, which determine expectations among other things. What we estimate is akin to a Roy model, where the individual chooses sector (education) and by doing so obtains the equivalent stochastic income stream.

From this model we expect to learn the role of schooling costs and availability and economic conditions when young in determining schooling outcomes. We will also obtain a series for education returns at each level, cleaned up for composition effects; this in turn will uncover the actual evolution of returns.

4.1 Schooling choice

Individuals choose between four education levels: primary, secondary, high school and college. We use a reduced form utility where each schooling level depends on costs and availability of schooling as well as on local labour market opportunities and quality of schooling. The number of schools up to high school in a given State in the year of birth divided by the population of the State in that year will provide our measure of availability of schooling. A measure of the share of each State in national GDP relative to the population share of the State in the year of birth will control for local labour market opportunities and the number of teachers per school over the population of a State in the year of birth will provide a proxy for schooling quality at the State level by birth cohort. To allow for expectations of future returns and other factors affecting choices we also include State and cohort dummies.

Lifetime utility of schooling net of schooling costs for individual i at time t will be:

$$U_{it} = X_{it}'\beta + \varepsilon_{it} \tag{1}$$

where X is a matrix of observable characteristics and ε_{it} is the error term representing all unobserved factors affecting utility. The matrix of observables X includes individual characteristics, dummies for year and State of birth and our proxies for local labor market opportunities, availability and quality of schooling by birth cohort. The beta coefficients are assumed to be constant across transitions, which is consistent with a cost function that is stable between successive education levels. Permanent unobserved characteristics are introduced as an individual-specific shifter of returns relative to costs associated with each schooling level that is observed by the individual but not by the analyst. It is assumed to be independent of X and to enter the relative cost function in an additive and separable way. We can therefore rewrite ε_{it} in the following way:

$$\varepsilon_{it} = \vartheta_1 h_i + e_{it} \qquad\qquad e_{it} \backsim N(\mu_e, \sigma_e^2) \tag{2}$$

where h_i denotes unobserved heterogeneity of individual i and ϑ_1 is the coefficient associated with it. The decision rule associated with this simple model is given by

$$\text{choose schooling level } S = s \qquad \text{if} \quad a^s < U_{it} \leq a^{s+1}$$

Cameron and Heckman (1998) provide conditions for this ordered choice model to be a reduced form of a dynamic education choice model.

4.2 Labor market participation

When modelling wages there are two endogenous variables. The first is education choice and we control for this by specifying the model described above. However, it is as important to control for endogenous labour supply/participation decisions. Not doing so can bias both the skill prices and the returns to education. The decision to work is modelled as a reduced form equation which depends on a set of observable individual characteristics, household demographics and unobserved heterogeneity. An individual i with schooling level s at time t decides whether to become a wage earner according to the following single threshold crossing model:

$$P_{it}^{*} = G_{it}^{'}\gamma + \xi_{it} \tag{3}$$

where P_{it}^{*} is a latent indicator function; when this is positive the individual works. The error term takes the following form:

$$\xi_{it} = \vartheta_2 * h_i + u_{it} \qquad\qquad u_{it} \backsim N(\mu_u, \sigma_u^2) \tag{4}$$

where, as above, h_i summarizes unobserved characteristics that are assumed to be independent of the set of observable variables G and ϑ_2 is the coefficient associated with it. Participation status will be denoted by $P = 1$ $(P^* > 0)$ for workers and $P = 0$ $(P^* \leq 0)$ otherwise.

4.3 Wages

We assume that labour markets are competitive. Individuals are price takers. If working, an individual i with schooling level s at time t is paid a wage:

$$w_{it}^{s} = p_t^s \exp(\zeta_s(age_{it}) + \omega_{it}^{s}) \tag{5}$$

where the function $\zeta_s(age)$ reflects the growth of wages with experience which here is proxied by age and is education specific. The term p_t^s is the price of education level s at time t. We allow the relative price of human capital to differ by education group, allowing for the possibility that the skill categories may not be perfectly substitutable in production. The error term takes the following expression:

$$\omega_{it}^{s} = \vartheta_{3ts} h_i + v_{it}^{s} \qquad\qquad v_{it}^{s} \backsim N(\mu_{vs}, \sigma_{vs}^2) \tag{6}$$

By allowing for the common factor h_i we allow for correlation between the education and participation choices and wages. The factor loading ϑ_{3ts} varies by education and time to allow for the impact of unobserved heterogeneity to change over time as in much of the literature on wage inequality. By allowing for schooling specific factor loadings we also allow the returns to education to be heterogeneous.[8]

The observed wage differential between two schooling levels s and $s^{'}$ is given by:[9]

[8] Card (1999) provides a detailed summary of the evidence on heterogenous returns to schooling in non-strucural model and Belzil (2006) revises the empirical evidence in structural models.

[9] For simplicity we are ignoring the composition effects related to working in this example (but not in the actual empirical analysis).

$$
\begin{aligned}
&E(\log w_t|S=s',\ age)-E(\log w_t|S=s,\ age)= \\
&\Big[\log p_t^{s'}-\log p_t^{s}+(\zeta_{s'}(age)-\zeta_s(age))\Big] \\
&+(\vartheta_{3ts'}-\vartheta_{3ts})E(h|S=s') \\
&-\vartheta_{3ts}(E(h|S=s)-E(h|S=s'))
\end{aligned}
\tag{7}
$$

The term in the first line is the average payoff to schooling level s' with respect to schooling level s at time t at some given age. The term in the second line reflects heterogeneity in returns and shows how the return differs for those who choose schooling level s'. This will change as the composition of individuals changes over time; however, it is still interpretable as part of the average return for the sub-population who at a point in time decided to obtain that level of schooling. The last term is the "selection bias" and reflects the differing ability composition of the education groups.

4.4 Firms

All firms are assumed homogeneous and competitive in the labour market. The production function in year t is given by:

$$Y_t = Z_t K_t^{\alpha} HH_t^{1-\alpha} \tag{8}$$

where Y_t denotes aggregate output, K_t is physical capital and HH_t is aggregate human capital. Capital-skill complementarities are ruled out. This is consistent with the near-constancy of the share of physical capital in Brazilian production in the 1990s.[10] We assume no adjustment costs for capital and we price out K at the international interest rate. Z_t is the technology factor and it is normalized to one in all years.

We consider four types of human capital corresponding to the four levels of education that individuals can complete, namely primary, secondary, high school and college. The aggregate stock of human capital j in year t, $H_{j,t}$, is the sum of the individual supplies of skill j, $h_{j,i,t}$, over each cohort a and individual i active in the labor market in year t:

$$H_{j,t} = \sum_a \sum_i h_{j,i,t}(a) \qquad j = 1,2,3,4 \tag{9}$$

where $h_{j,i,t}(a)$ is the supply of skill level j of individual i who is of age a in year t.

We choose a specification for the aggregate human capital that allows for different elasticities of substitution (ES) between human capitals' pairs. We group primary and secondary into unskilled and high school and college into skilled human capital

[10] National account data show that the share of physical capital in Brazil in the 1990s is stable at around 0.43.

and we specify the aggregate human capital HH as a nested CES over skilled and unskilled:

$$HH_t = \left\{\delta_{s,t} H_{s,t}^{\rho} + \delta_{u,t} H_{u,t}^{\rho}\right\}^{\frac{1}{\rho}} \qquad \delta_{u,t} = (1 - \delta_{s,t}) \tag{10}$$

where $H_{s,t}$ and $H_{u,t}$ are, respectively, the human capital for skilled and unskilled labour at time t. We assume that the skilled (unskilled) human capital is a CES combination of the two top (bottom) education groups:

$$H_{s,t} = [\delta_{4,t} H_{4,t}^{\theta} + \delta_{3,t} H_{3,t}^{\theta}]^{\frac{1}{\theta}} \qquad \delta_{3,t} = (1 - \delta_{4,t}) \tag{11}$$

$$H_{u,t} = [\delta_{2,t} H_{2,t}^{\gamma} + \delta_{1,t} H_{1,t}^{\gamma}]^{\frac{1}{\gamma}} \qquad \delta_{1,t} = (1 - \delta_{2,t}) \tag{12}$$

The time-varying and skill-specific factors δ denote the shares of the human capital types in production. Changes in δ reflect variations in the productivity and in the demand of the different H inputs. The parameter ρ determines the ES between skilled and unskilled labor, which is given by $ES_{s,u} = \frac{1}{1-\rho}$, while θ and γ determine, respectively, the ES between college and high school and the one between secondary and primary, which are given by $ES_{4,3} = \frac{1}{1-\theta}$ and $ES_{2,1} = \frac{1}{1-\gamma}$.[11]

The demand for human capital j at time t is given by the solution to the market clearing conditions:

$$p_t^j = \frac{\partial Y_t}{\partial H_{j,t}} \qquad j = 1, 2, 3, 4 \tag{13}$$

where p_t^j is the price of human capital of type j in year t.

5 Identification and estimation

5.1 Identification

The model is fully parametric. However, it is useful to understand what drives identification, other than the obvious functional form assumptions. In particular, we are controlling for the endogeneity of both participation and education choices in the wage equations. Education choice depends on variables reflecting school quality, and availability as well as labour market conditions when the student was young. These variables are assumed to play no direct role on wages. Perhaps the most tenuous of these assumptions is that school quality does not have a direct effect on wages, other than through the level of education. They are also assumed to play no role in the employment choice equation, conditional on education, which identifies the impact of the latter on the decision to work. Participation

[11] There are different ways of measuring the ES when the aggregate output is produced with more than two inputs. We use the definition of the direct ES. Two other commonly used definitions are the Allen and the Hicksian elasticity of substituion.

depends on marital status and the presence of children, which are also excluded from the wage equation, driving the identification of the selection into work effect on wages. This may be a strong assumption and one way to relax it would be to include in the participation equation local child cost variables, including local health infrastructure, at the values they had at some critical age for the individual, such as at the age of 20 or 25; however, we have not pursued this here as we do not think it is central to the issues we are concerned with.

Finally, we use a measure of local quality and availability of schooling in the year of birth but we do not have information on the actual State where individuals attended school. Then, a key identifying assumption is that labour markets clear across States while the education market is constrained. As in Behrman and Birdsall (1983), we assume that quality of education varies across geographical areas and individuals do not move across areas in response to quality differentials, while they can move in post schooling years in response to geographical wage differentials. The assumption of constrained education markets is untestable with our data but is likely to be realistic given that our geographical units are States of large size so that long migrations would be required to change areas.

5.2 Estimation

The joint density of wages, probability of schooling level $S = \{1, 2, 3, 4\}$, and probability of work status $P = \{1, 0\}$ is

$$f(w, S, P|Z, h) = f_w(w|Z, h)F_s(S = s|Z, h)F_p(P = p|Z, h)$$

with $f_w(.)$ denoting the density function of wages, assumed to be normal, $F_s(S = s|Z, h)$ is defined by an ordered probit for education choice and $F_p(P = p|Z, h)$ is defined by a probit for participation. The dependence of these three outcomes is driven by the overlap in observables and by the single permanent unobserved factor h, which is assumed independent of Z. Let $F(h)$ denote the distribution of this unobservable. Then the contribution of individual i to the likelihood function is:

$$\begin{aligned} L_i &= \int_h f_w(w_i|Z_i, h)F_s(S = s_i|Z_i, h)F_p(P_i = 1|Z_i, h)dF(h) \qquad if \qquad I = 1 \text{ (workers)} \\ &= \int_h F_s(S = s_i|Z_i, h)F_p(P_i = 0|Z_i, h)dF(h) \qquad if \qquad I = 0 \text{ (non-workers)} \end{aligned}$$

Following Heckman and Singer (1984), we approximate the distribution of unobserved heterogeneity by a discrete distribution with M points of support, with M being determined by the data. The sample log-likelihood function for N individuals is:

$$\log L = \sum_{i=1}^{N} P_i \log \sum_{m=1}^{M} f_w(w_i|Z_i, h) F_s(S = s_i|Z_i, h) F_p(P_i = 1|Z_i, h) p_m$$

$$+ (1 - P_i) \log \sum_{m=1}^{M} F_s(S = s_i|Z_i, h) F_p(P_i = 0|Z_i, h) p_m$$

where p_m denotes the probability attached to the m_{th} point of support.

Finally, note that not all variables in Z enter all equations. There are exclusion restrictions which we discussed above. They are not shown explicitly to simplify the notation. We estimate all the equation jointly using a full MLE.

6 Empirical results

6.1 Educational attainment

Table 3 presents the results from the estimation of the ordered probit model for educational attainment allowing for unobserved heterogeneity. The regression includes State of birth and cohort dummies, controlling for general trends and permanent regional differences. Identification of the reported effects works as of differential changes in the economic position of the States and resources available for schooling.

The State share in national GDP relative to the population share of the State where the person was born is a significant factor to explain education choices: a 10 per cent increase of this measure for males translates into an increase of 0.022 in the probability of graduating from college for both males and females and into an increase of 0.032 in the probability of graduating from high school for males and of 0.039 for females at sample means. Another way of looking at the the role of the GDP is considering the effect of this variable on the educational gap between the poorest and the richest State. The probability of graduating from college in the State with the highest per capita share in national GDP (per population share) is 0.135 for males and 0.148 for females. The figure for the probability of graduating from high school is 0.252 for males and 0.281 for females. Over the same period (1992 and 2002) the probability of graduating from college in the State with the lowest per capita share in national GDP is 0.081 for males and 0.089 for females. The figure for the probability of graduating from high school is 0.164 for males and 0.185 for females.

Thus, the effects are quite substantial. At college level they account for around 41 per cent of the educational gap for males and for around 36 per cent of the gap for females. At high school level, they account for around 37 per cent of the educational gap between the poorest and richest State for males and for around 41 per cent of the gap for females.

This can reflect a number of factors: first, it may reflect economic opportunities in the labour market, both for unskilled workers and for graduates; second, it can reflect

Parameter	*Males*	*Females*
State Share in National GDP	0.2408	0.2403
	(0.0222)	(0.021)
Population of the State	0.3246	0.2518
	(0.0384)	(0.0366)
Quantity Education	0.0581	0.0752
	(0.0221)	(0.0209)
Quality Education	0.2060	0.1779
	(0.0309)	(0.0292)
White	0.6387	0.6351
	(0.0097)	(0.0094)
Natives	1.5928	1.4795
	(0.0278)	(0.0267)
Mulatos	0.0787	0.0879
	(0.0099)	(0.0096)
N. Obs	568477	618255
Log-Lik	1492565.2	1494913.6

Table 3: Schooling Choice Equation, standard errors in parenthesis

resources put into schooling, not captured by our other measures; finally, it can reflect liquidity constraints. All these factors imply a positive effect of GDP per head on educational attainment, except for the possible correlation of GDP with the unskilled wage, which reflects the opportunity cost of education and would tend to push education downwards. However, the positive effect of local income clearly dominates and almost certainly reflects liquidity constraints, given the large returns to education.

Race appears as an important determinant of schooling attainment. There are four main ethnicity groups in Brazil: white, natives, blacks and mulatos. With respect to the excluded category, being native and white has a particularly strong positive association with skill upgrading.[12]

As expected, for both males and females, both availability and quality of schooling have a positive and significant impact on educational attainment, with the availability effect being larger for women. These variables are measured at the date and State of birth of the individual and the identification of the effect works as of differential change in the position of the State.

We run a simulation exercise to evaluate the impact of a policy intervention to increase the quality of schooling on graduation rates. Table 4 presents the implied impact of a 10 per cent change in school availability and teachers per school on graduation rates for high school and college.

[12] Leite (2005) shows that in Brazil there are persistent educational inequalities between races, which in turn explain a significant part of the racial wage differentials.

	Males		*Females*	
	High School	College	High School	College
10% increase school per population	0.007	0.005	0.011	0.006
10% increase teachers per school	0.027	0.018	0.028	0.016

Table 4: Impact of a Ten Per Cent Increase in the Number of School and Teachers per School on Graduation Rates

The implication of these numbers is that resources are a very important factor for attainment, with the number of teachers per school having the strongest influence. Also, for both males and females the impact of an increase in education quality on graduation rates is much higher at high school than at college level. The large drop in schooling resources, shown in Figure 13, which is of the order of 30 per cent, would have exerted a large negative influence on educational attainment. Thus, although attainment has been improving, the decline in resources has held back growth. This may be one of the key reasons why the level of college graduation may be so low despite the large returns.

6.2 Unobserved heterogeneity

We have employed a one factor distribution for unobserved heterogeneity having found that a two factor model did not improve the fit. The distribution of unobserved heterogeneity necessary to fit this data well has just two points of support, with about 17 per cent of men and 16 per cent of women having a relatively high propensity for schooling. Richer distributions of unobservables did not improve the likelihood and did not change the estimates of the parameters of interest. The impact of unobservables on the various parts of the model is discussed in the sections below.

A useful exercise is to trace how the composition of the education groups in terms of this unobserved propensity/ability changed over time. Using Bayes' theorem we can compute the probability of the corresponding ability type conditional on achievement of a given education level and the matrix X of observable characteristics. Figure 14 and 15 present this probability for males and females. Among men the proportion of college graduates with high ability was virtually 100 per cent for those born in 1940 and dropped to just under 90 per cent for the youngest individuals. For women it dropped from 100 per cent to 80 per cent. Thus the quality of college graduates has been dropping steadily as higher education expanded. The effects have been much more dramatic for secondary and high school: the proportion of high ability males attaining secondary education fell from about 40 per cent in our earlier cohorts to less than 5 per cent for the youngest one. For females the drop is from 40 per cent to zero. For high school the proportion of high ability males and females fell from 90 per cent in our earlier cohorts to below 20 per cent for the youngest one. Thus, there has been much closer sorting by ability and educational attainment recently, with those attaining the lower levels being dominated by lower abilities and thus leading to increasing earnings' inequality. Section 6.4 will quantify

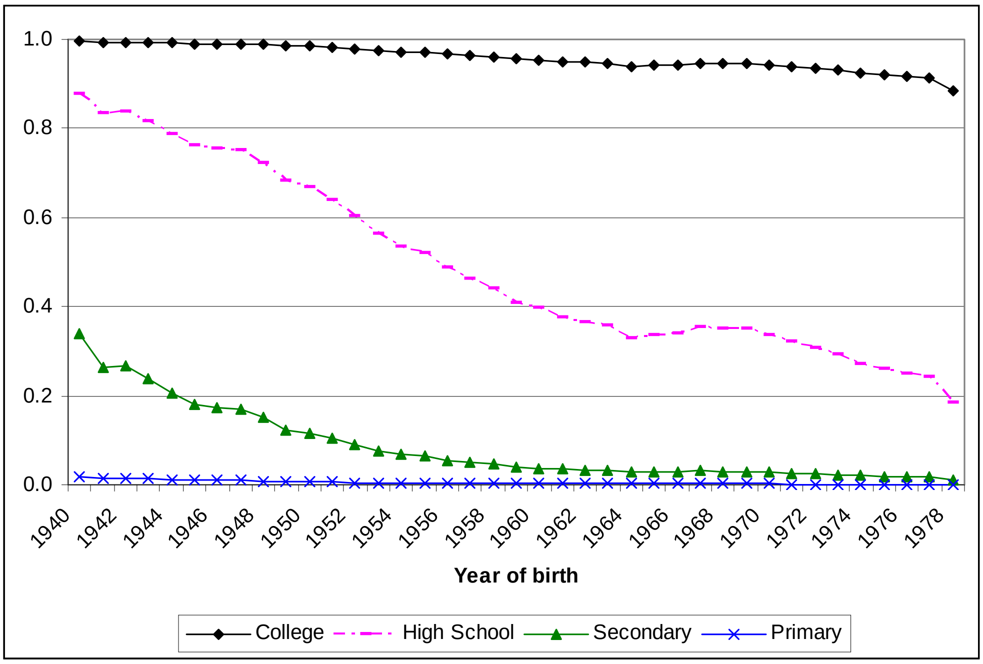

Figure 14: Proportion if High Ability by Level of Education, Males

these effects.

6.3 Participation

Table 5 presents the results from the estimation of the participation equation. The regression includes race dummies, State of birth and cohort dummies, the latter controlling for general trends and permanent regional differences. There is a very strong impact of education on participation. The effect of increasing educational attainment is stronger for men than for women. Being married and having young children makes a woman less likely to work and a man more likely. As children grow, the size of the impact declines. All in all the labour market participation results are very similar to what has been obtained in many other countries.

The unobserved factor (whose effect on schooling is normalized to 1) has a coefficient of 0.126 in the participation equation for males, which, given the values of the points of support, implies that there is a 0.34 increase in the index for those with the higher value. For the normal distribution at the mean this implies an increase in participation of 13 percentage points for the higher ability group. For women the effect is much smaller. The coefficient on the unobserved factor is 0.015 and the estimates of the points of support are very similar. Thus the higher ability women have an increase in the index of just 0.04 implying an increase of participation at the mean of the normal (0 index) of 1.6 percentage points. Thus unobserved ability is not much of a factor in explaining female

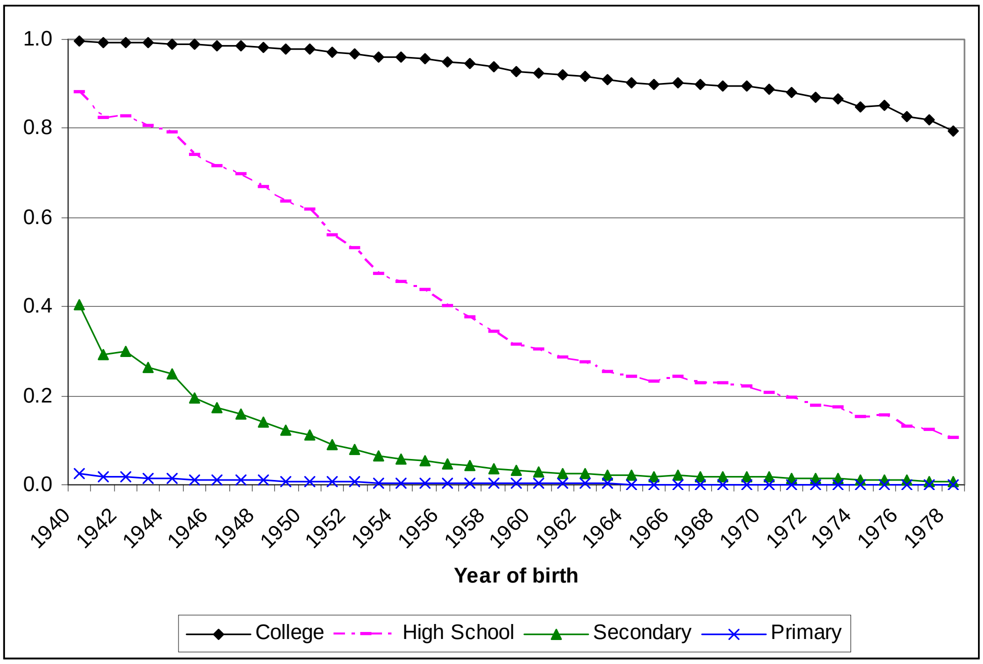

Figure 15: Proportion of High Ability by Level of Education, Females

participation, implying that observables account for differences between working and non-working women. It also implies that education is not endogenous for female labour force participation.

6.4 Wages and returns to schooling

We estimate a Roy model for wages with four sectors. The stochastic link between these and the rest of the model is unobserved heterogeneity. Table 6 shows the impact of the unobserved factor on wages as it changes over time. For men the unobserved factor is very important at all educational levels. The impact remains effectively unchanged over time, other than for primary school, where the impact has declined to almost a third of its original impact. This group may now be involved in the most menial tasks as the better jobs are now taken by better educated individuals, who will be more productive. Another remarkable aspect of these results is that the ability premium is lower among college graduates, where the pay rates for the two ability groups differ by about 21 per cent. At the other extreme for men with primary education the ability premium was 156 per cent in the early part of our sample and declined to 48 per cent in the later part, but still more than twice that for the college students. The fact that unobserved heterogeneity is quantitatively important implies that both education and participation are endogenous for wages. The results for women follow the same patterns with small

Parameters	*Males*	*Females*
Secondary	0.1339	0.2093
	(0.0061)	(0.0047)
High School	0.3605	0.5403
	(0.0128)	0.0079
College	0.6189	1.0625
	(0.0243)	(0.0193)
Age	3.9017	4.0810
	(0.0787)	(0.0661)
Age Squared	-1.9043	-2.1462
	(0.0374)	(0.0317)
Couple with Children	-0.3074	-0.0977
	(0.0074)	(0.0063)
Number of Children	-0.0511	-0.0181
	(0.0015)	(0.0013)
Children aged 0 to 1	0.6429	-0.3546
	(0.0084)	(0.0069)
Children aged 2 to 6	0.5942	-0.1089
	(0.0067)	(0.0053)
Children aged 7 to 15	0.4121	0.0559
	(0.0059)	(0.0047)
Spouse is present	0.2164	-0.4872
	(0.0059)	(0.0041)
Constant	-0.8868	-1.7815
	(0.0473)	0.0474
N. Obs	568477	618255

Table 5: Labor Market Participation

quantitative differences. Here again unobserved heterogeneity is important and education is endogenous for wages.

In Figure 16 we present the returns to college relative to high school and the way these have changed over time. In the figure we show how the returns evolve if we ignore unobserved heterogeneity (No UH in the graph) and if we control for it. In interpreting these results, note that the composition of college graduates in the population changes only slowly, because of the coexistence of many cohorts. It is noteworthy that the upward bias in the returns when we ignore UH becomes smaller and eventually negative, which reflects the way the ability composition of college and high school graduates have been changing over time. Entrance into college of high ability people initially resulted into a positive selection bias that eventually became negative due to the entrance of low-ability people as the high ability ones were already all in. Overall, the returns for males have

Males

θ_3	*1992-1995*	*1996-1998*	*1999-2002*
Primary	0.5922	0.2354	0.1845
	(0.023)	(0.1087)	(0.1952)
Secondary	0.0991	0.1210	0.1145
	(0.0299)	(0.0323)	(0.0392)
High School	0.1094	0.1259	0.1130
	(0.0121)	(0.0117)	(0.0117)
College	0.0753	0.0625	0.0779
	(0.0466)	(0.0428)	(0.0337)

Females

θ_3	*1992-1995*	*1996-1998*	*1999-2002*
Primary	0.6831	0.1083	0.0925
	(0.0415)	(0.3151)	(0.4728)
Secondary	0.0787	0.0917	0.0932
	(0.0445)	(0.0547)	(0.0667)
High School	0.1054	0.0984	0.0933
	(0.0163)	(0.0168)	(0.0170)
College	0.0666	0.0754	0.0754
	(0.0456)	(0.0377)	(0.0299)

Table 6: Unobserved Heterogeneity - Wage Equations

increased from about 80 per cent to just over 100 per cent.

Turning to women, the returns to higher education and how they have changed over time are shown in Figure 17. Controlling for composition effects the returns increase from about 80 per cent to 107 per cent. Interestingly and differently than for men, the bias in the returns when we ignore unobserved heterogeneity is always negative: the returns are biased downwards by the presence of a percentage of low ability women among college graduates. The negative bias increases over time because of the worsening ability mix among the college graduates with an increase that is more pronounced than for men. Almost all the rise in the returns happened between 1992 and 1996, which is a different pattern than for men. However, what stands out in both cases is the fact that the already large returns increased substantially and now stand by over 100 per cent for both groups.

Quite a different picture emerges when we consider the returns to high school (versus secondary) and to secondary (versus primary). These returns have been steadily declining as shown in Figure 18, 19, 20 and 21. Composition effects always bias the returns negatively and for males the decline in the returns to secondary appear lower if selection bias is not taken into account. The negative bias is the result of the presence of a large proportion of low ability workers with intermediate and primary education in each year of the sample (see section 6.2).

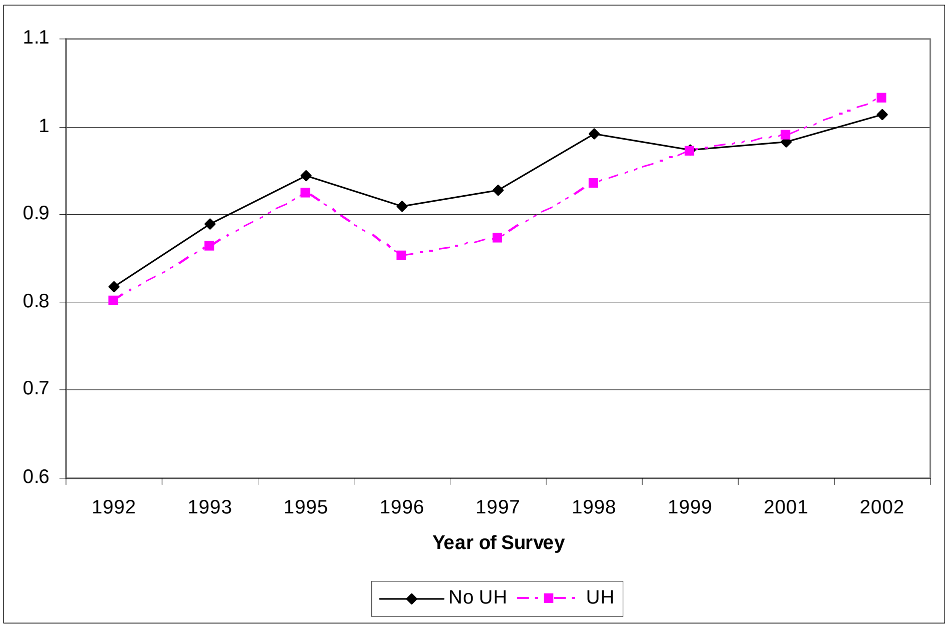

Figure 16: Log Relative Wages College versus High School, Males

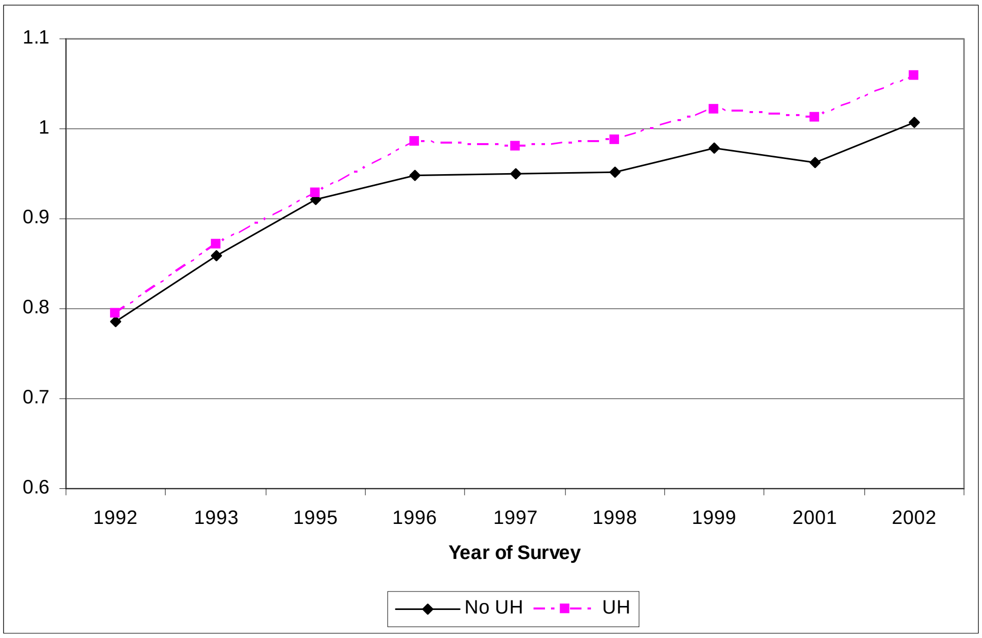

Figure 17: Log Relative Wages College versus High School, Females

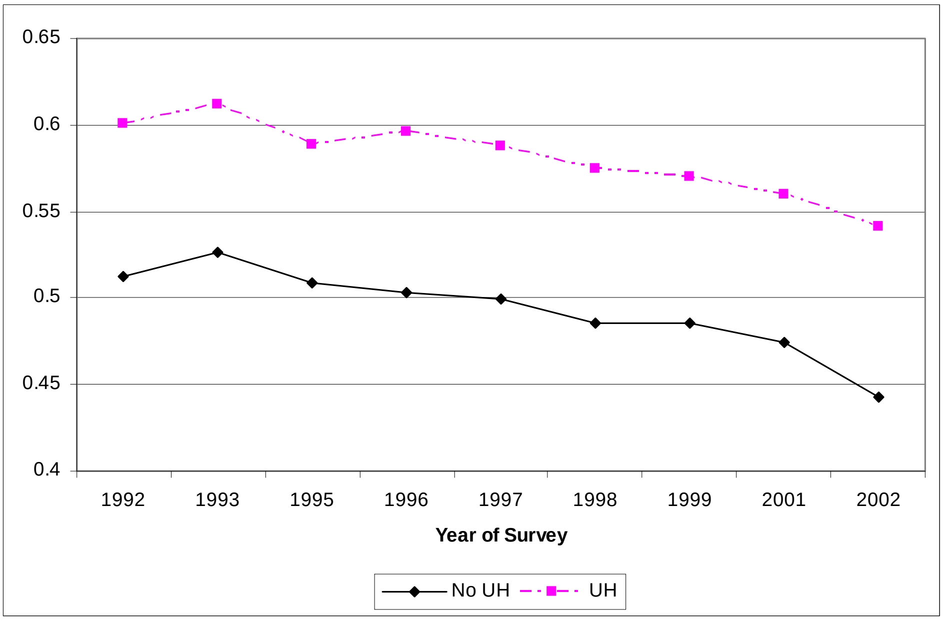

Figure 18: Log Relative Wages High School versus Secondary, Males

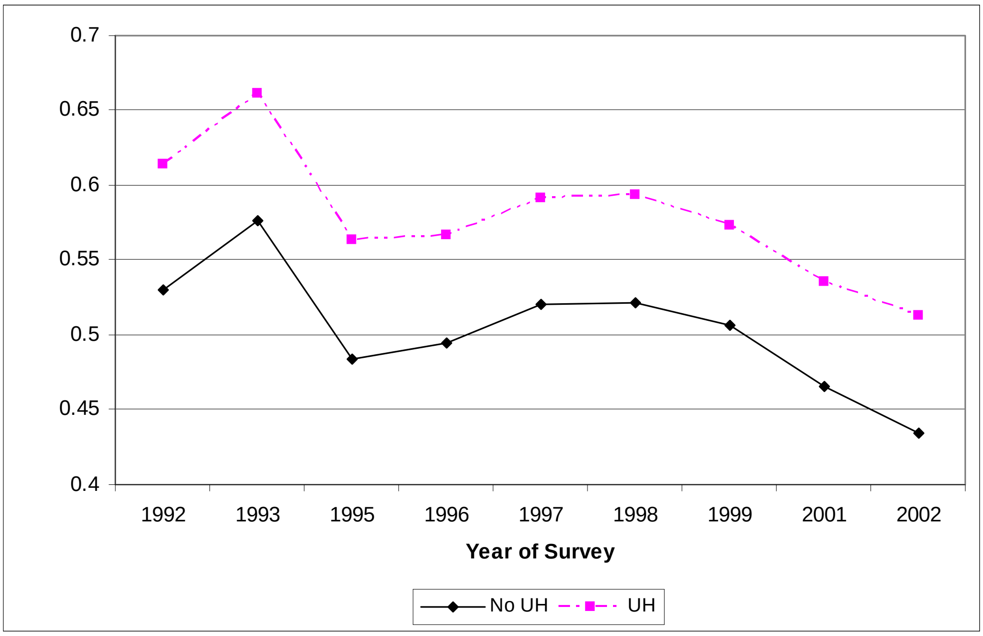

Figure 19: Log Relative Wages High School versus Secondary, Females

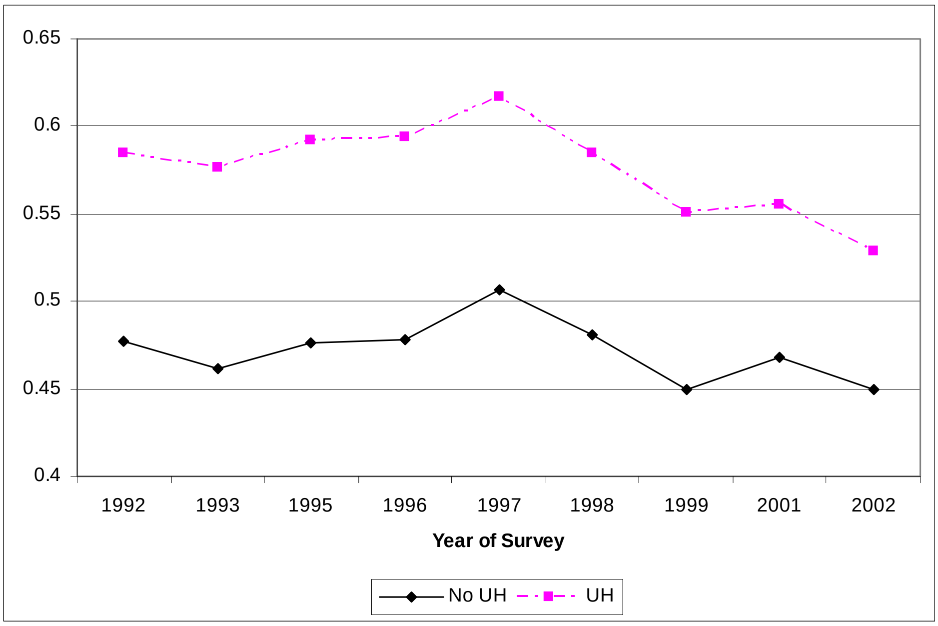

Figure 20: Log Relative Wages Secondary versus Primary, Males

Thus, the overall picture is of large increases in the returns to higher education and large declines in the returns to secondary and high school. These last two events are consistent with the large increase in the supply of secondary and high school graduates and the decline in those completing just primary school. However, the returns to higher education have been accompanied by an increase in supply, indicating that demand outstripped the gradual supply increase at this level.

Finally, we can use the decomposition in equation 7 above to estimate how the change in educational composition affects the observed changes in the returns to education. This can be viewed in two ways. First, we can consider the observed returns to a level of education in the entire population with that level of attainment - this will include the contribution of many cohorts and is precisely what we showed in the graphs above. Considering the first and last year of our sample and all alive cohorts in these years, in the male (female) sample the difference in relative returns to college versus high school is around seventeen (nineteen) per cent higher when the model does account for selection.

Second, we can carry out the same exercise for the returns of a particular cohort. Considering the youngest cohort in our sample, the changes in composition show a bigger impact on the measurement of the returns. However, the impact remains in the range of the one estimated as the mean among all alive cohorts and the returns are still very high.

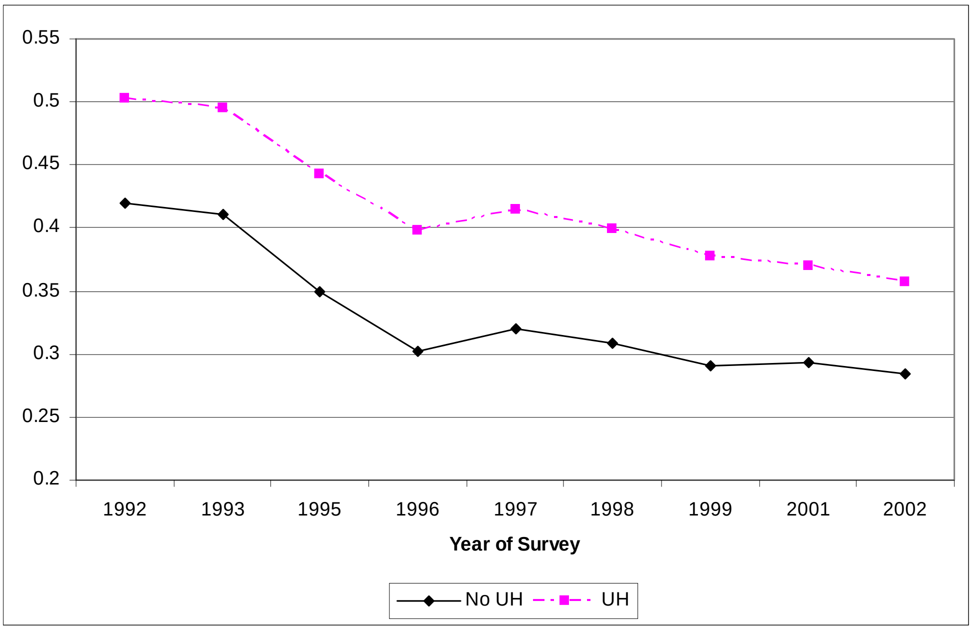

Figure 21: Log Relative Wages Secondary versus Primary, Females

6.5 Production function

We estimate an aggregate production function to understand the substitutability between the different types of human capital. The elasticity of substitution (ES) between production inputs is an important parameter to pin down since it determines the size and the direction of the general equilibrium effects of changes in the supply of skills on changes in their prices.

By having an estimate of the log prices for each level of education, we can identify the ES between aggregate labor inputs from the variation of log relative prices and log relative supplies.[13] Since we estimated the model separately for males and females, in each year of the sample and for each skill level we compute the log price as a weighted average of the log price estimated in the male and in the female sample using as weights the proportions of males and females by skill group in the data.

Let us consider the estimation of the elasticity of substitution between college and high school, $ES_{4,3}$. We solve for the market clearing conditions 2.13 for college and high school to derive an expression for p_t^4 and p_t^3. By taking the ratio between the two and log

[13] With our model we have controlled for the endogeneity of both education choices and labor market participation. Therefore, our results do not rely on the assumption that wages equal prices, which is typicaly made when the supply of skills is not explicitly modelled (see, for example, Katz and Murphy (1992) and Manacorda, Sanchez-Paramo and Schady (2006)).

linearizing, we have:

$$(\log p_t^4 - \log p_t^3) = [\log \delta_{4,t} - \log \delta_{3,t}] + (\theta - 1) * (\log H_{4,t} - \log H_{3,t}) \quad (14)$$

where the $\log \delta_{j,t}$ are the time series of the relative demand for the jth input measured in log quantities units and $H_{j,t}$ denotes the supply of education level j in year t. Variations in $\delta_{4,t}$ and $\delta_{3,t}$ reflect skill-biased technological change as well as changes in the productivity of workers with college and high school education. Given the significant shifts in education attainment in the 1990s, we express the log of the share parameters as the sum of a constant and a time-varying component:

$$\log \delta_{j,t} = \phi_{0,j} + \phi_{1,j} * t + e_{j,t} \qquad j = 1, 2, 3, 4 \quad (15)$$

where $\phi_{0,j}$ is a skill-specific constant, t denotes a linear time trend and $e_{j,t}$ is a normally distributed i.i.d. shock at time t for skill level j.

Combining equations 2.14 and 2.15, the value of the parameter determining the ES between college and high school, θ, can be estimated from a regression of the log prices differential between college and high school on the log supply differential, a linear trend and a constant. We control for the possible endogeneity of the human capital inputs in the production function through an IV approach with lagged regressors.

We can use the same procedure to estimate the elasticity of substitution between primary and secondary, $ES_{2,1}$. Then, we use equations 2.11 and 2.12 above to construct a measure of skilled and unskilled human capital. To do so, we need an estimate of the log factor shares $\delta_{j,t}$. By simple substitution of the expression for $\log \delta_{4,t}$ and $\log \delta_{3,t}$ from equation 15 into 14, it is easy to see that the coefficient of the time trend is the difference of the time trend for college and high school $(\phi_{1,4} - \phi_{1,3})$, which gives an estimate of the yearly relative demand-shift for college with respect to high school. Given $\delta_{3,t} = (1 - \delta_{4,t})$, we have that $\log \left[\frac{\delta_{4,t}}{(1-\delta_{4,t})} \right] = (\beta_0 + \beta_1 * t)$, where $\beta_0 = (\phi_{0,4} - \phi_{0,3})$ and $\beta_1 = (\phi_{1,4} - \phi_{1,3})$. Therefore, $\delta_{4,t} = \frac{\exp(\beta_0 + \beta_1 * t)}{(1+\exp(\beta_0+\beta_1 * t))}$. Following the same procedure, we can use the equation of the log prices differential between secondary and primary to obtain an estimate of $\delta_{2,t}$.

Finally, we can estimate a regression of the log prices differential between skilled and unskilled on the log supply differential, a linear trend and a constant to obtain an estimate of the parameter determining the ES between skilled and unskilled. We compute the skilled (unskilled) log price as a weighted average of the log price for high school and college (primary and secondary) using as weights the proportions of workers with high school and college (primary and secondary) in the data. In order to improve efficiency, for each skill level we linearly interpolate the values for the supply and log price for the two missing years in the sample, 1994 and 2000.

Table 7 presents the results of the estimation for secondary versus primary. The elasticity of substitution between primary and secondary is not statistically different from one. We then take $\gamma = 1$ and define unskilled labour as the sum of the supply of workers with primary and secondary education.

Parameter	*Secondary versus Primary*
Difference log supply	-0.0743
	(0.1039)
Time trend	-0.0131
	(0.0057)
Constant	26.7745
	(11.5576)
	$\gamma = 0.926$
Implied ES	13.51

Table 7: Estimation of the Production Function - Secondary versus Primary, standard errors in parenthesis

Parameter	*College versus High School*	*Skilled versus Unskilled*
Difference log supply	-0.7941	-0.1937
	(0.2357)	(0.0829)
Time trend	0.0091	-0.0112
	(0.0032)	(0.0038)
Constant	-18.3411	22.5549
	(6.3879)	(7.7809)
	$\theta = 0.206$	$\rho = 0.806$
Implied ES	1.26	5.16

Table 8: Estimation of the Production Function - College versus High School and Skilled versus Unskilled, standard errors in parenthesis

Table 8 presents the estimates obtained for college versus high school and for skilled versus unskilled labour. The estimates of ρ and θ are consistent with the presence of complementarities between high school and college: $\rho > \theta$.[14] This result confirms the findings for Mexico. Using Mexican data between 1987 and 2002, Binelli (2008) estimates a CES production function with three labor inpupts and finds that $ES_{4,3} = 4.4$ and

[14] Manacorda, Sanchez-Paramo and Schady (2006) use a cross section of LACs that includes Argentina, Brazil, Chile, Colombia and Mexico in the 1980s and 1990s and find a value of around 4.5 for the ES between higher and intermediate education and a value of around 2.5 for the ES between skilled and unskilled. Therefore, they find a lower ES between skilled and unskilled than between higher and intermediate. They allow for a different ES between age groups and they pool five LACs together. The difference with respect to our results could be due to any of these features.

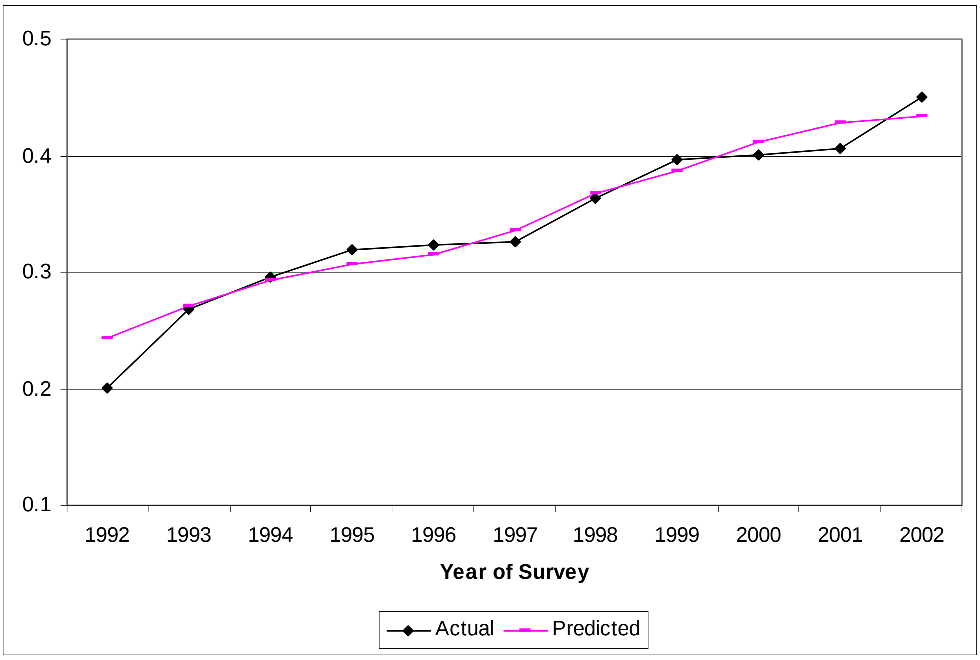

Figure 22: Actual and Predicted Log College Wage Premium

$ES_{s,u} = 7.1$.[15]

A joint estimation of the system of equations to test for the equality of the coefficients of the log relative supplies confirms that ρ and θ are highly statistically significantly different. The test gives a value of chi-squared of 31.69 with a P-value of 0.000. We also test for the assumption of equality between $ES_{4,u}$ and $ES_{3,u}$, which is a restriction imposed by the symmetry of the CES operator. The test gives a value of chi-squared of 0.93 with a P-value of 0.334. Therefore, the test can not reject the null hypothesis of equal coefficients.

Figure 22 and 23 present the actual and fitted log wage premium for college versus high school and skilled versus unskilled computed from estimating equation 14 above and the equivalent regression for skilled versus unskilled. The model does explain the movements in the college and skill wage premium rather well: changes in relative supply have been an important determinant of changes in the relative prices of education in the 1990s.

The impact of changes in the supply of education on changes in the returns depend on the degree of substitutability/complementarity between production inputs. The presence of complementarities between high school and college reinforces the effect of an increase in the supply of high school on changes in relative returns. By taking the ratio of the

[15] The prices used to estimate the production function are generated by our model. In order to correct for this, we are working on the computation of the bootstrapped standard errors. Preliminary results show that the coeffcients of all the explanatory variables in Table 2.8 remain highly statistically significant.

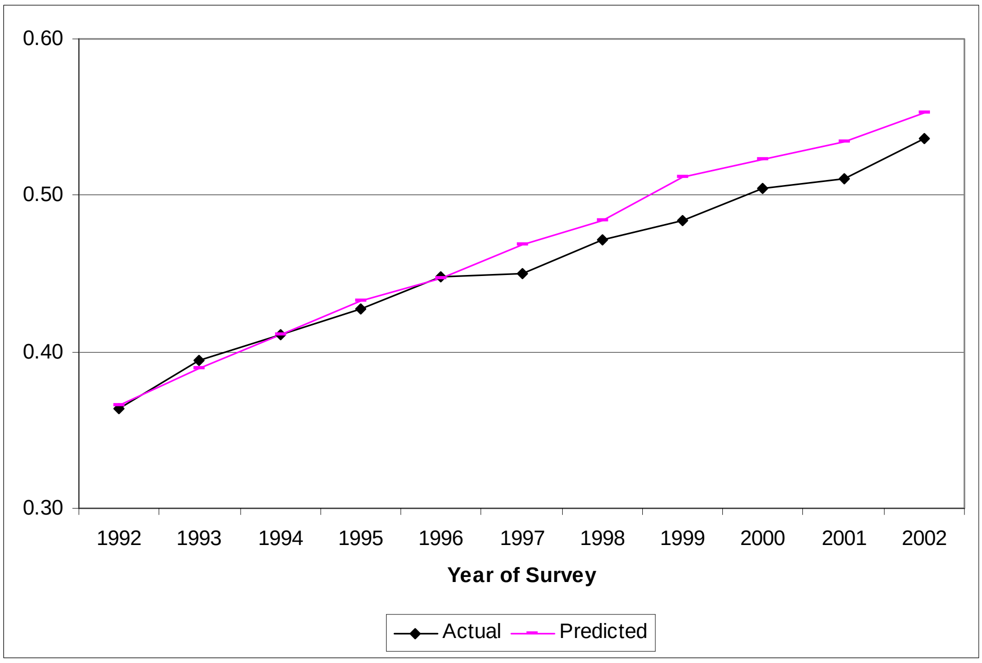

Figure 23: Actual and Predicted Log Skill Wage Premium

demand of human capital j and $(j+1)$, we can derive the expressions for the relative returns to schooling:

$$\frac{p_{3,t}}{p_{u,t}} = \frac{\delta_{s,t}}{(1-\delta_{s,t})}(1-\delta_{4,t})\left(\frac{H_{u,t}}{H_{3,t}}\right)^{1-\rho}\left\{(1-\delta_{4,t})+\delta_{4,t}\left[\frac{H_{4,t}}{H_{3,t}}\right]^{\theta}\right\}^{\frac{\rho-\theta}{\theta}} \tag{16}$$

$$\frac{p_{4,t}}{p_{3,t}} = \frac{\delta_{4,t}}{(1-\delta_{4,t})}\left(\frac{H_{4,t}}{H_{3,t}}\right)^{\theta-1} \tag{17}$$

$$\frac{p_{4,t}}{p_{u,t}} = \frac{\delta_{s,t}}{(1-\delta_{s,t})}\delta_{4,t}\left(\frac{H_{u,t}}{H_{4,t}}\right)^{1-\rho}\left\{\delta_{4,t}+(1-\delta_{4,t})\left[\frac{H_{3,t}}{H_{4,t}}\right]^{\theta}\right\}^{\frac{\rho-\theta}{\theta}} \tag{18}$$

An increase in the supply of workers with high school has two different effects: a standard supply effect (SE) and a complementarity effect (CE). The standard SE is clear from the ratio in round brackets in equation 16 and 17. An increase in H_3 decreases the relative return to high school with respect to unskilled workers and increases the relative return to college with respect to high school. The CE is given by the ratio in curly brackets in equations 2.16 and 2.18. If $\rho > \theta$, that is if college and high school are more complementary than college and unskilled (or high school and unskilled), an increase in H_3 further decreases the relative returns to high school with respect to unskilled and increases the relative return to college versus unskilled.

By substituting the values of the factor shares and the education supplies estimated from the data, we can use equation 16 to compute the relative returns to high school versus unskilled in the presence ($\rho > \theta$) and in the absence ($\rho = \theta$) of complementarities. Complementarities in production are responsible for around twelve (thirty-one) per cent further decline (increase) in the relative return to high school (college) versus unskilled labour with respect to the case of $\rho = \theta$.

7 Conclusion

In the 1990s Brazilian education expanded to quite an extraordinary extent due to a significant increase in graduation rates at secondary and high school levels. On the contrary, the relative proportion of students progressing to higher education decreased and college education remained at a low level. This is quite surprising given that returns to college relative to high school have been increasing by over one hundred per cent in the decade.

The challenge is to understand what prevented greater participation. First, we show that there is no lack of availability in positions: the private sector of higher education has expanded significantly and there is a persistent oversupply of positions there. If availability of places is not an issue, there are two main possible explanations for the high returns to college: first, the observed returns could reflect a premium to ability rather than changes in the market value of education. This is an appealing explanation since the educational expansion might have resulted into significant changes in the ability composition of workers with intermediate and higher education. Second, individuals might face credit constraints to access and complete higher education.

We specify and estimate a joint model of education choices, labour force participation and wages where ability enters as an unobserved factor affecting all parts of the model. We find that compositional changes have been substantial. The increase in graduation rates resulted into a deterioration of the ability level at all levels of education, particularly at secondary and high school. However, the changes in composition can only explain about a fifth of the increase in the college premium. Then, given the large increase in high school graduation and the huge returns to college, individuals must be constrained when facing the decision to enter college. Our analysis provides some hints in this direction.

First, in Brazil, as in many other Latin American Countries, there are two main education sectors, the public and the private one, which differ significantly in the quality of education they offer and in the fees they charge. The quality gap opens up at intermediate levels (if not earlier) with private expensive high school providing higher quality education. Then, the differential reverses at college with the top Universities being public and free of charge. There is a very competitive exam to enter public Universities. It then becomes crucial to have attended high quality expensive intermediate schools to be able to access free-of-charge high quality public Universities, the alternative being the very expensive private colleges.

Second, we show that in the 1950s and 1960s there was an effective reduction of resources per pupil, which has discouraged a further increase in educational attainment. Also, our simulations show that resources are a very important factor for schooling attainment, with the number of teachers per school having the strongest influence.

Third, we find a very large positive impact of relative GDP on schooling progression. The large dropout rates in the private sector of higher education and the fact that the public sector is many times oversubscribed are further indicators of liquidity constraints being an important determinant of educational attainment.

Overall, our results show that the observed changes in the returns to schooling only marginally reflect variations in the ability composition by level of education. They are mainly the result of changes in the market value of education due to a bottleneck effect at intermediate: the educational expansion at intermediate did not translate into a proportional increase at college and therefore depressed the wage at this level dramatically. The size of the supply effect on changes in the wages depends on the substitutability between production inputs. We quantify this effect by estimating a CES production function that allow for different elasticities of substitution between four labor inputs. We find that there are important complementarities in production between workers with high school and college. These complementarities are responsible for about twelve (thirty one) per cent further decline (increase) in the relative return to high school (college) versus less than high school education with respect to the case of an isoelastic production function with no complementarities.

References

[1] Behrman J. R., Birdsall N., (1983) "The Quality of Schooling: Quantity Alone is Misleading", *American Economic Review*, Vol. 73, No. 5, pp. 928-946.

[2] Belzil C., (2007) "The Return to Schooling in Structural Dynamic Models: A Survey of the Literature", *European Economic Review*, 51 (5), 1059-1105.

[3] Belzil C., Hansen J., (2001) "Heterogeneous Returns to Human Capital and Dynamic Self-Selection", IZA Discussion Paper No. 272.

[4] Belzil C., Hansen J., (2002) "Unobserved Ability and the Return to Schooling", *Econometrica*, 70 (2), 2078-2091.

[5] Binelli C., (2008) "Returns to Education and Increasing Wage Inequality in Latin America", Mimeo UCL.

[6] Blundell R., Reed H., Stoker T., (2003) "Interpreting Aggregate Wage Growth: The Role of Labour Market Participation", *American Economic Review*, 93:1114-1131.

[7] Blundell R., Gosling A., Ichimura H., Meghir C., (2004) "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition using Bounds", IFS Working Paper, W04/25.

[8] Cameron S., Heckman J., (1998) "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males", *Journal of Political Economy*, Vol. 106, No. 2, pp. 262-333.

[9] Card D., (1999) "The causal effect of education on earnings", chapter 30 in O. Ashenfelter & D. Card (ed.) Handbook of Labor Economics, pp. 1801-1863.

[10] Card D., Lemieux T., (2001) "Can Falling Supply Explain the Rising Return to College for Younger Men? A Cohort-Based Analysis", *Quarterly Journal of Economics*, Vol. 116, Issue 2, 705-746.

[11] Carneiro P., Heckman J., Vytlacil E., (2005) "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education", Mimeo.

[12] Carneiro P., Lee S., (2006) "Changes in College Enrollment and Wage Inequality: Distinguishing Price and Composition Effects", Mimeo.

[13] Fallon P. R., Layard P. R. G., (1975) "Capital-Skill Complementarities, Income Distribution, and Output Accounting", *Journal of Political Economy*, Vol. 83, No. 2, p. 279-301.

[14] Gallipoli G., Meghir C., Violante G., (2007) "Equilibrium Effects of Education Policies: a Quantitative Evaluation", Mimeo.

[15] Jacoby H., Skoufias E., (2002) "Jacoby H. G., Skoufias E., (2002) "Financial Constraints on Higher Education: Evidence from Mexico", Mimeo.

[16] Heckman J., Sedlacek G., (1990) "Self-Selection and the Distribution of Hourly Wages", *Journal of Labor Economics*, Vol. 8, No. 1, S329-S363.

[17] Heckman J., Sedlacek G., (1985) "Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market", *Journal of Political Economy*, Vol. 93, No. 6, pp. 1077-1125.

[18] Heckman J., Singer B., (1984) "A method for minimizing the impact of distributional assumptions in economic models for duration data", *Econometrica*, 52, 271-320.

[19] Heckman J., Singer B., (1984) "The Identifiability of the Proportional Hazard Model", *Review of Economic Studies*, 51, No. 2, 231-241.

[20] Katz L., Murphy K., (1992) "Changes in Relative Wages, 1963-1987: Supply and Demand Factors", *Quarterly Journal of Economics*, Vol. 107, Issue 1, 35-78.

[21] Manacorda M., Sanchez-Paramo C., Schady N., (2006) "Off and Running? Technology, Trade, and the Rising Demand for Skilled Workers in Latin America", Mimeo.

[22] Oliveira J. B. A., Schwartzman S., (2002) "A Escola vista por dentro". Belo Horizonte, Alfa Educativa Editora.

[23] Roy A. D., (1951) "Some thoughts on the distribution of earnings", *Oxford Economic Papers*, 3, 135-146.

[24] Taber C., (2001) "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?", *Review of Economic Studies*, Vol. 68, No. 3, 665-691.

[25] Taber C., (2000) "Semiparametric identification and heterogeneity in discrete choice dynamic programming models", *Journal of Econometrics*, 96, 201-229.

[26] Waltenberg F. (2007) "Benchmarking of Brazil's Education Performance Using PISA 2003 Datasets", in *The Quality of Education in Brazil*, IADB

[27] Willis R., Rosen S., (1979) "Education and Self-Selection", *Journal of Political Economy*, Vol. 87, No. 5, pp. 7-36.

Appendix A - Data: Descriptive Statistics

Variable	MALES		FEMALES	
	Mean	SD	Mean	SD
Primary (S=1)	N = 266804 Workers = 219829		N = 280179 Workers = 105661	
Log wage	-6.306	2.802	-6.593	2.726
Age	39.556	9.119	40.137	9.056
Race	5.038	2.894	4.899	2.897
Log number schools	-6.601	0.397	-6.619	0.394
Log number profs	-14.099	0.674	-14.105	0.670
Log population	15.187	0.779	15.193	0.775
Per capita GDP share	0.765	0.512	0.784	0.523
Number children	2.359	1.784	2.443	1.773
Age youngest child	11.003	9.728	11.604	9.161
Year of birth	1957.46	9.450	1956.819	9.318
Secondary (S=2)	N = 137055 Workers = 117476		N = 140693 Workers = 66229	
Log wage	-6.106	2.553	-6.599	2.446
Age	34.669	8.017	34.772	8.057
Race	4.446	2.852	4.434	2.860
Log number schools	-6.498	0.379	-6.492	0.377
Log number profs	-14.129	0.705	-14.124	0.691
Log population	15.423	0.782	15.391	0.778
Per capita GDP share	0.909	0.564	0.883	0.561
Number children	1.953	1.429	2.064	1.368
Age youngest child	10.649	9.835	10.529	9.210
Year of birth	1962.818	8.268	1962.727	8.286

Table 9: Summary Statistics Primary and Secondary (Source: PNAD)

Variable	MALES		FEMALES	
	Mean	SD	Mean	SD
High School (S=3)	N = 109228 Workers = 95482		N = 131702 Workers = 79399	
Log wage	-5.635	2.567	-6.124	2.457
Age	35.222	8.190	34.939	8.159
Race	4.070	2.773	4.094	2.785
Log number schools	-6.546	0.394	-6.532	0.396
Log number profs	-14.106	0.772	-14.103	0.770
Log population	15.419	0.816	15.391	0.812
Per capita GDP share	0.945	0.595	0.897	0.586
Number children	1.765	1.309	1.813	1.308
Age youngest child	11.956	10.319	12.601	10.332
Year of birth	1962.397	8.583	1962.725	8.549
College (S=4)	N = 55390 Workers = 49283		N = 65681 Workers = 50948	
Log wage	-4.630	2.589	-5.092	2.508
Age	37.866	8.629	36.772	8.381
Race	3.073	2.235	3.092	2.254
Log number schools	-6.613	0.386	-6.585	0.386
Log number profs	-14.133	0.747	-14.127	0.750
Log population	15.483	0.811	15.480	0.815
Per capita GDP share	1.119	0.629	1.071	0.619
Number children	1.571	1.168	1.563	1.172
Age youngest child	12.986	10.505	14.479	11.195
Year of birth	1959.59	9.006	1960.805	8.719

Table 10: Summary Statistics High School and College (Source: PNAD)

1992				
State	**Share Primary**	**Share Secondary**	**Share High School**	**Share College**
Maranhao	0.638	0.158	0.162	0.042
Piaui'	0.661	0.149	0.149	0.041
Ceara'	0.605	0.183	0.157	0.055
Rio Grande Norte	0.618	0.177	0.146	0.058
Paraiba	0.638	0.161	0.128	0.073
Pernambuco	0.540	0.213	0.167	0.079
Alagoas	0.659	0.144	0.144	0.053
Sergipe	0.608	0.175	0.152	0.066
Bahia	0.611	0.171	0.173	0.045
Minas Gerais	0.580	0.196	0.143	0.081
Espirito Santo	0.582	0.190	0.168	0.060
Rio de Janeiro	0.375	0.237	0.240	0.148
Sao Paulo	0.404	0.230	0.199	0.167
Parana'	0.527	0.221	0.158	0.095
Santa Catarina	0.542	0.226	0.159	0.073
Rio Grande Sul	0.463	0.246	0.168	0.123
Mato Grosso	0.466	0.264	0.174	0.096
Mato Grosso Sul	0.440	0.312	0.171	0.078
Goias	0.514	0.236	0.185	0.065
Distrito Federal	0.127	0.240	0.426	0.207

Table 11: Education Shares by State in 1992 (Source: PNAD)

2002				
State	**Share Primary**	**Share Secondary**	**Share High School**	**Share College**
Maranhao	0.493	0.240	0.221	0.046
Piaui'	0.536	0.213	0.186	0.065
Ceara'	0.467	0.253	0.205	0.075
Rio Grande Norte	0.459	0.252	0.209	0.080
Paraiba	0.551	0.203	0.167	0.078
Pernambuco	0.425	0.248	0.232	0.095
Alagoas	0.581	0.214	0.157	0.048
Sergipe	0.475	0.224	0.229	0.072
Bahia	0.473	0.215	0.251	0.061
Minas Gerais	0.428	0.252	0.217	0.103
Espirito Santo	0.407	0.262	0.238	0.094
Rio de Janeiro	0.259	0.262	0.300	0.178
Sao Paulo	0.239	0.252	0.290	0.219
Parana'	0.370	0.266	0.250	0.114
Santa Catarina	0.380	0.281	0.217	0.123
Rio Grande Sul	0.315	0.301	0.229	0.155
Mato Grosso	0.322	0.319	0.242	0.118
Mato Grosso Sul	0.347	0.295	0.256	0.102
Goias	0.369	0.280	0.252	0.099
Distrito Federal	0.074	0.231	0.455	0.240

Table 12: Education Shares by State in 2002 (Source: PNAD)