Educational Attainment and Wages in Brazil

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

PRELIMINARY and INCOMPLETE

This draft: May 2006

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 compulsory school and progressing to intermediate levels of education. 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 enrolment increased up until the early 1970s but has remained stagnant since then. At the same time relative returns to Secondary and High School have declined, while relative returns to College have steadily increased.

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

[*]University College London (UCL) and Institute for Fiscal Studies (IFS)
[†]University College London (UCL) and Institute for Fiscal Studies (IFS)
[‡]University of Sao Paulo

In addition, if there is a supply constraint on College graduates and to the extent that there is some substitutability between High School and College, we would have expected to see some increase in the returns to High School relative to College.

The first aim of the paper is to study education choices and investigate the reasons for the lack of growth in the number of College graduates. We focus on the impact of local labour market opportunities and the availability of suitable high schools that can prepare individuals for the demanding College exams. [1] The decline in the quality of the High School system could explain both the inability of pupils to move to College and the decline in the returns to High School. Combining data from Brazilian National Household Survey (PNAD) in the 1990s with historical series on population, GDP, number of schools and teachers per school by year and State of birth we will examine the impact of changes in availability and quality of schooling on educational choices.

We then focus on wages, estimating a joint model of wages and education choices to understand the extent to which the observed increase in returns to higher education has been driven by changes in the composition of those who attend intermediate levels of education. A decline in the quality of the High School system due to an increase in the number of students not accompanied by a proportional increase in schools and teachers' provision could have restricted entrance into higher education to high-ability students and to those that attended high-quality intermediate schools that prepared them to seat for the competitive College exams.

Students in higher education will be a progressively selective sample of individuals and returns to unobservables will represent an important factor affecting the evolution of returns to schooling. A key component of the full education choice-wage equation model is unobserved heterogeneity that follows a discrete distribution which is non-parametrically estimated allowing for a differential impact on education choices, as

[1] The University system in Brazil is divided between public and private Universities. Public universities are free and the private universities charge yearly fees. There are exams (called "vestibular") to enter all universities. Each university has its own exam. The exams generally consist of several questions on Portuguese, Maths, History, Science, etc., depending on the applicant´s area of interest. Entrance to public Universities is generally much more competitive.

well as on labour market participation and earned wages.

The empirical results show the role of availability and quality of schooling to promote skill-upgrading and bring evidence of compositional changes as an important determinant of changes in relative wages. The estimates bring evidence of a dramatic deterioration over time in the average ability of students at intermediate levels of education and an increase in graduation among the most able at College level. For the youngest cohorts, among High School graduates there is a high proportion of low-ability individuals while most high-ability individuals have completed College.

The results have implications for the design of effective interventions to promote investments in higher education. A simulation of the effects of an increase in schooling quality on workforce composition shows how policy interventions can induce substantial compositional changes that will in turn affect changes in relative wages. Improvements in education quality at intermediate levels could be an effective device to ease constraints on the supply side and enable students to move further up in the education ladder. However, induced compositional changes should also be accounted for in order to predict the net impact of the policy intervention on relative wages. This result is in line with the findings in Cameron and Heckman (1998) that quantify the compositional changes induced by a policy intervention to increase College enrollment and graduation rates in the US.

The analysis developed in this paper is related to the literature on self selection into education and employment and their effects on changes in 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 stressing the importance of accounting for selection into work. [2] Closer to our framework are Willis and Rosen (1979) and Taber (2001) that study selection

[2] They show how accounting for the choice between market and non-market work is the most successful extension of the original Roy model in order to explain cross-section wage distributions and their evolution over time.

into schooling and its impact on earned wages. Carneiro and Lee (2005) also study the effect of self-selection into education on the evolution of wage inequality but they focus on marginal distributions while we will 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, Ichimunra 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 America country. The focus of their paper is on educational choices and they ignore the impact of self-selection on labor market participation and earned wages. They develop a dynamic model of enrollment decisions into University and include unobserved heterogeneity as an additional factor driving individuals' choices. They estimate the model using data on Mexican youths in the 1990s and find only weak evidence of selection bias.

The remainder of the paper is organized as follows. In section 2 we present the economic model that underlies the empirical analysis. Section 3 discusses model's identification and provides a description of the estimation procedure. In section 4 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. Section 5 gives some concluding remarks.

2 The model

2.1 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 level will provide our measure of availability of schooling at the State level and will account for constraints in schooling supply before College entry in a given State over and above the impact of permanent characteristics and national trends that will be captured by State and time dummies. A measure of the share of each State in national GDP will control for trends in local labour market opportunities and the number of teachers per school will provide a proxy for schooling quality. Availability and quality of schooling as well as GDP share will be measured in the year when the individual was born in order to avoid endogeneity bias.

To study schooling choice we use a simple ordered probit model that we mix with a discrete distribution of unobserved heterogeneity which allows us to relax the IIA assumption. For each education outcome s the corresponding utility level for individual i at time t will be:

$$U_{it}^{s} = X_{it}'\beta + \varepsilon_{it}^{s} \qquad\qquad s = \Pr imary, Secondary, High\ School, College \qquad (1)$$

where X is a matrix of observed characteristics and ε is the error term representing all unobserved factors affecting schooling choice s. The matrix of observables X includes individual characteristics, dummies for year and State of birth, measures of local labor market opportunities, availability and quality of schooling. Betas are assumed to be constant across transitions, which is consistent with a cost function that is stable across transitions between successive education levels.

Unobserved heterogeneity is introduced as an individual-specific shifter of return relative to costs associated with each schooling level that is observed by the individual by 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}^{s} in the following way:

$$\varepsilon_{it}^{s} = \vartheta_1 * h_i + e_{it}^{s} \qquad\qquad e_{it}^{s} \backsim N(\mu_{es}, \sigma_{es}) \qquad (2)$$

where h_i denotes unobserved heterogeneity of individual i and ϑ_1 is the coefficient associated with it.

2.2 Participation and wage equations

We model labour market participation accounting for selection into work as an additional selection mechanism that will affect observed wages. Participation decisions depend on unobserved heterogeneity and a set of observable individual characteristics and household demographics. An individual i with schooling level s at time t decides whether to participate in the labour market according to the following simple probit model:

$$I_{it}^s = Z_{it}^{'}\gamma + \xi_{it}^s \tag{3}$$

where I is an indicator function that equals one if the individual works and zero otherwise. The error term takes the following form:

$$\xi_{it}^s = \vartheta_2 * h_i + u_{it}^s \qquad\qquad u_{it}^s \backsim N(\mu_{us}, \sigma_{us}) \tag{4}$$

where, as above, h_i is unobserved heterogeneity that is assumed to be independent of the set of observable variables Z and ϑ_2 is the coefficient associated with it.

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

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

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. The error term takes the following expression:

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

The random effect assumption on h is maintained and ϑ_{3ts} gives a measure of the impact of unobserved heterogeneity on observed wages at time t for education level s. Following several previous contributions on heterogeneous returns to schooling, [3] we allow the impact of unobserved heterogeneity to vary by time and education level.

Ignoring for simplicity subscript i and using equations 5 and 6 above, observed wages for education level s will take the following expression:

$$\log w_t^s = (\log p_t^s + \zeta_s(age)) + \vartheta_{3ts} * h + v_t^s \tag{7}$$

Considering two consecutive schooling levels s and $s^{'}$ with $s^{'} > s$ and assuming $\mu_{vs} = \mu_{vs'} = 0$, observed variation in relative wages is given by:

$$\begin{aligned} E(\log w_t|S &= s^{'}) - E(\log w_t|S = s) = \left[\log p_t^{s^{'}} + \zeta_{s'}(age) - \log p_t^s - \zeta_s(age)\right] + \\ &+ \vartheta_{3ts'} * E(h|S = s^{'}) - \vartheta_{3ts} * E(h|S = s) \end{aligned} \tag{8}$$

The term in the first line is the payoff to schooling level $s^{'}$ with respect to s at time t. This will change as relative demand for different production factors changes. The term in the second line reflects heterogeneity in returns which depends on variation in two different components: variation in the payoff to h measured by changes in ϑ_{3ts} and variation in the mean of unobserved heterogeneity for a given schooling level measured by $E(h|S = s)$. Using Bayes' theorem we will compute $E(h|S = s)$ for each of the four levels of schooling we consider in the analysis. We will use this statistic to study changes in the distribution of heterogeneity among those attending different schooling levels: the change in distribution is the change in composition by unobserved returns within each education group.

[3] See Card (1999) for a detailed survey of the empirical evidence on heterogenous returns.

3 Identification and estimation procedure

3.1 Identification

The model developed in section 2 can be thought of as the reduced form specification of a two-stage decision process where individuals first choose the optimal level of schooling and then decide upon labour market participation. In order to control for different sources of selection that can affect observed wages the model allows for unobserved heterogeneity to have different impacts on schooling choice and participation in the labour market.

h can be interpreted as a persistent characteristic of the individual with different effects in the different stages of the decision process. From an econometric point of view it is an unobserved covariate that affects the outcome variable and is common to all states and time periods, while ϑ_1, ϑ_2 and ϑ_{3ts} can be seen as the regression coefficients that estimate its impact on the outcome variables. In the actual estimation, ϑ_1 will be normalized to one, so the parameters of interest will be h, ϑ_2 and the set of ϑ_3 for the different time periods and education levels.

The estimation of the model involves two main identification issues. First we need some instruments to identify and consistently estimate the skill prices p_t^s. Second, we need to be able to identify h and its distribution allowing for a differential impact of unobserved heterogeneity in the wage and participation equations.

Identification of p_t^s is achieved using the number of schools up to High School and the number of teachers per school in each State as proxies for availability and quality of schooling that affect earned wages only through education choices. Since we use a measure of local quality and availability of schooling but we do not have information on the actual State where individuals went to school, 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 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 it is likely to be realistic given the low mean education level of 6.3 completed years of schooling in the sample and the fact that our geographical units are States of large size so that long migrations would be required to change areas. The logarithm of the population of a State and the State's share in national GDP are also included in the X matrix and provide additional sources of variation to identify p_t^s.

Estimated p_t^s will be the time-varying prices to skill net out of unobserved heterogeneity which drives selection into education and work. Identification of h and its coefficients is therefore of fundamental importance to evaluate the impact of compositional changes on the evolution of wage differentials. Identification of h rests on the results in Cameron and Heckman (1998) that discuss identification of transition schooling models with unobserved heterogeneity. In particular, they show that h can be identified with no need of any assumption on the distribution of the unobservables. The only required assumptions are to model unobserved heterogeneity as a random effect with no correlation with the variables included in the matrix of the observables Xs which has to include at least one continuous variable and has to be full rank. In our model, availability and quality of schooling together with population and GDP share ensure this condition is satisfied.

The inclusion of a participation equation in the model adds an additional stage in the decision process leading to become a wage earner. Identification of ϑ_2 requires some exogenous source of variation affecting labour market participation once schooling decisions have been made. In the same way as we need some instruments to identify p_t^s, we need some variables affecting labour market participation but not education choices. Following a standard practice in the literature on labour supply, we use household demographics such as the presence and number of children and the presence of the spouse as important determinants of work decisions and effective exclusion restrictions to identify ϑ_2.

Taber (2000) shows how the effectiveness of the instruments strongly relies on the length of their support. The exclusion restrictions are powerful if the variables have a large support. In our case, while the support of the instruments to identify the returns to schooling is rather large, household demographics do vary to a much less extent among individuals so that the power of the instruments to identify ϑ_2 is clearly limited by the nature of of the exclusion restrictions.

3.2 Estimation

From the model described in section 2 we can derive the individual likelihood function and the overall likelihood that will be the object of the estimation. Denoting with $f(w, I, s|X, Z, h)$ the joint density function of earned wage w, working status I and schooling level s, the individual likelihood takes the following expression:

$$L_i = \int_h \int_{I \geq 0} f(w, I, s|X, Z, h) dh dI \tag{9}$$

Integrating over labour market participation and assuming independence between the error terms in the schooling, participation and wage equations, we can rewrite the individual joint density as it follows:

$$\begin{aligned} L_i &= \int_h \phi(w|I, s, X, Z, h) \Pr(I > 0|s, X, Z, h) \Pr(S = s|X, Z, h) dh \qquad if \qquad I > 0 \\ &= \int_h (1 - \Pr(I > 0|s, X, Z, h)) \Pr(S = s|X, Z, h) dh \qquad if \qquad I = 0 \end{aligned}$$

where $\phi(.)$ denotes the wage density which is assumed to follow a standard normal distribution.

Assuming unobserved heterogeneity follows a discrete distribution and applying the non parametric Heckman and Singer (1984) estimator, we can rewrite the individual likelihood function as a sum over a finite number of points of support:

$$
\begin{aligned}
L_i &= \sum_{m=1}^{M} \phi(w|I,s,X,Z,h)\Pr(I>0|s,X,Z,h)\Pr(S=s|X,Z,h)p_m \qquad if \qquad I>0 \\
&= \sum_{m=1}^{M} (1-\Pr(I>0|s,X,Z,h))\Pr(S=s|X,Z,h)p_m \qquad if \qquad I=0
\end{aligned}
$$

where M is the number of points of support and p_m denotes the probability attached to the mth point.

Given N individuals in the population, the overall likelihood will simply be the product of the individual contributions:

$$L = \prod_{i=1}^{N} L_i$$

4 Composition effects and the evolution of wage inequality

In this section we apply the model developed in section 2 to study the evolution of earnings' inequality in Brazil during the 1990s. The first paragraph will describe the data and present the empirical evidence on wages and supply dynamics in Brazil during the 1990s together with the evolution of our measures of quantity and quality of schooling. The second paragraph will present the results from the estimation of the full schooling choice-wage equation model and analyze the evolution of compositional changes by education group and the impact on changes in observed wage differentials.

4.1 Wage and supply dynamics in Brazil 1992-2002

Data on wages and supply by education group come from Brazilian National Household Survey, PNAD (*Pesquisa Nacional Por Amostra de Domicílios*), for the period 1992-2002. PNAD is an annual and 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 cycles in Brazilian education system, namely Primary, Secondary, High School and College and construct four skill groups including in each group all the individuals with a number of years of education smaller or equal to the number necessary to complete a given schooling cycle. We combine PNAD with data from Historical Series collected by Brazilian statistical office, IBGE (*Instituto Brasileiro de Geografia e Estatística*), that provide data on the number of schools, number of teachers and teachers per school, 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 dropped data on States in the north of the country that only cover urban areas and we are left with a representative sample that includes twenty out of the twenty-seven Brazilian States.

Table 1 and 2 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 and females always receive on average a lower wage than their male counterpart.

Figures 1 and 2 present the evolution of relative wages and supply by education between 1992 and 2002. As can be seen from the figures, relative wages earned by College graduates increased by around fifty per cent, while relative wages for both High School and Secondary decreased by around twenty per cent. However, the increase in relative wages at College level has not been accompanied by an increase in relative supply which on the contrary appears to have been decreasing over time. [4]

[4] Low and stagnant enrollment rates at University level are a characterizing feature of Latin American Countries in the 1990s. Brasil and Mexico are two leading examples of this trend.

As discussed in the introduction, a decrease in schooling quality at intermediate levels could have induced a self-selection process of high-ability students into high-education that significantly affected changes in relative returns. Figures 3 and 4 plot the evolution of our measures of availability and quality of schooling over time. As can be seen from the graphs, while there has been a significant increase in availability of schooling, schooling quality has significantly decreased over time. In each year we construct our measures dividing the number of schools and the number of teachers per school by the population of a State. As a comparison and robustness check of the quality of our measures, we construct the same proxies replacing the overall State population with the relevant schooling age population of individuals aged 10-19. 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. As can be seen from the graphs, both quantity and quality of schooling follow the same evolution as before with a significant deterioration of schooling quality over time. The estimation of the full wage equation-schooling model will further investigate the importance of schooling quality and its impact on compositional changes.

4.2 Empirical results

The estimation of the model gives two main sets of results. First, we can use the estimates of the schooling equation to identify the impact of quantity and quality of schooling on education choices and to study the evolution of compositional changes by education level. Second, we can use the skill prices estimated with the full wage and education model to study the pattern of selection and the contribution of unobserved heterogeneity to changes in relative wages. We will examine compositional changes together with changes in the price of unobserved ability to explain the difference between relative returns estimated with and without unobserved heterogeneity.

4.2.1 Quality of schooling and compositional changes

Table 3 presents the results from the estimation of the schooling choice model with unobserved heterogeneity. As expected, both availability and quality of schooling have a positive and significant impact for both males and females. The share of the State of birth in national GDP has a particularly strong positive impact showing the importance of economic prosperity and good economic prospects as incentives to skill upgrading.

Availability of high quality schools that can prepare students for the demanding College exams appears as an important determinant of entrance into higher education. Declining quality of schooling at intermediate levels might have accelerated a process of self-selection on ability that resulted in increasing (decreasing) average ability at higher (intermediate) education levels. Compositional changes will therefore have affected variations in relative returns due to changes in average ability and in its price.

We can use the estimates of the model to study compositional changes and their evolution at different levels of education. The estimated points of support of the non parametric distribution of unobserved heterogeneity give the number and values of the ability types in the population. Using Bayes' theorem, for each mth point of support we can compute the probability of the corresponding ability type conditional on achievement of a given education level j with j = Primary, Secondary, High School and College and the matrix X of observable characteristics:

$$\Pr(h = h_m|S = j, X) = \frac{\Pr(S = j|h = h_m, X) * \Pr(h = h_m)}{\Pr(S = j|X)} \tag{10}$$

where $\Pr(S = j|X)$ is the unconditional probability of being of skill level j, $\Pr(S = j|h = h_m, X)$ is the probability of being of skill level j conditional on being of ability type h_m and $\Pr(h = h_m)$ is the estimated probability of the mth point of support.

The evolution of the probability in (10) provides a description of changes in unobserved ability over time. From the estimation of the model we were able to identify two points of support. Their values and corresponding probabilities are reported in table 5 together with the estimated overall distribution of unobserved heterogeneity. The results are presented in the graphs in panels 1 to 4 for males and females for each of the four schooling levels we are considering. Together with changes in the probability of being a high ability type conditional on the matrix of observables Xs, the graphs report the unconditional probability $\Pr(S = j|X)$ and the probability of having achieved a given education level conditional on being high ability type by year of birth.

As expected, the probability of having completed a schooling level higher than Primary conditional on the observables Xs has significantly increased over time, with a dramatic decrease in the probability of having stopped at Primary from an average of around 74 (77) percent for the oldest cohort to around 26 (21) percent for the youngest cohort for males (females).

At the same time, the probability that an individual from the upper half of the ability distribution achieved a given schooling level has been decreasing for all schooling levels but College. For Primary and Secondary education from already low levels of 23 (24) and 7 (10) per cent respectively for males (females), it quickly decreased to zero at Primary and almost to zero at Secondary level. At higher levels of education, while for the oldest cohort around 52 (50) per cent of the high ability group attended High School and only around 17 (15) per cent attended College among males (females), for the youngest cohort the proportion declines to 23 (16) percent at High School while it increases to 75 (83) per cent at College level for males (females). Therefore, over time, for both males and females, there has been a significant increase in College graduation among the most able.

The decline over time in the average ability of High School graduates is evident from the fast downsloping trend of $\Pr(h = high|S = High\ School, X)$ in panel 2. The probability of being high ability type dramatically decreased for High School graduates from a value slightly above 86 (88) per cent to a value around 14 (8) per cent for males (females). Increasing enrollment rates also decreased average ability at College level, but to a much lower extent. From a value of almost 100 per cent for the oldest cohort, the probability of being high ability decreased to a value of around 85 (76) per cent for males (females) for the youngest cohort. For Primary school graduates it has always been practically zero. For Secondary graduates it almost reached zero starting from a value of around 31 (39) per cent for the oldest cohort for males (females).

4.2.2 Selection on unobservables and returns to schooling

Together with selection into education, the model accounts for selection into work. This is likely to be rather important in the Brazilian context where unemployment rate has been varying substantially over time and States. [5] Table 4 presents the estimates of the participation equation for males and females. The equation includes dummies for the level of education with Primary being the excluded category, dummies for race, age, a dummy for the presence of children and of a spouse, the number of children and the age of the youngest child as important household demographics that affect labour market participation. All explanatory variables are significant with the probability of working increasing with age and the level of skill and decreasing with presence and number of children. The estimated coefficient of unobserved heterogeneity is reported in table 5 together with the estimated overall distribution of unobserved heterogeneity. Unobserved heterogeneity has a positive impact on participation decisions which is significant for both males and females but much higher in magnitude in the males' sample.

We can use the skill prices estimated from the joint wage education model to compute the returns to schooling once we have accounted for unobserved heterogeneity.

Applying the decomposition of changes in observed wages from equation (8) to the College-High School wage differential, we can write:

$$
\begin{aligned}
E(\log w_t|S &= C) - E(\log w_t|S = HS) = \left(\log p_t^C - \log p_t^{HS}\right) + (\zeta_C(age) - \zeta_{HS}(age)) + \\
&+ \vartheta_{3tC} * E(\eta|S = C) - \vartheta_{3tHS} * E(\eta|S = HS)
\end{aligned}
$$

[5] Between 1992 and 2002 average unemployment rate increased from 5 to almost 7 percentage points for males and from 8 to more than 11 percentage points for females with significant variation among States.

The observed wage gap is given by changes in the market price for College and High School graduates once accounted for education-specific experience and changes in the pattern of selection. The contribution of unobserved heterogeneity is given by the two terms in the second line and consists of two different components: changes in the price of unobserved heterogeneity reflected in changes in ϑ_{3t} for each schooling level and changes in average ability of those completing a given education level s given by changes in $E(\eta|S = s)$. In order to capture the evolution of the pattern of selection while maintaining a sufficient number of observations to perform the non parametric estimation, we have divided our nine years period into three three-years time intervals and ϑ_3 has been allowed to have a differential impact by schooling level in each of the three periods. [6]

We can compare skill prices estimated from the full education and wage model with and without unobserved heterogeneity. As can be seen from the graphs in panel 5, for the female sample, relative returns at both High School and College estimated from the model that includes unobserved heterogeneity are significantly higher than the ones computed ignoring unobserved heterogeneity, while for males differences are much smaller and relative returns to College are lower between 1996 and 1998 and higher between 1999 and 2002 in the model with unobserved heterogeneity with respect to the estimates obtained without controlling for selection. In addition, for both males and females, coherently with the role of unobserved ability as an important determinant of wages at the top of the income distribution, accounting for unobservabed heterogeneity does not seem to matter to explain changes in relative returns to Secondary education.

The gap between the two lines in the figures in panel 5 is explained by the terms in the second line of the above decomposition, which is a combination of average ability at a given education level and the time varying, skill-specific coefficients of unobserved heterogeneity. The results discussed in the previous paragraph document an average

[6] The choice of dividing the sample into three main periods reflects the evolution of relative supply in the 1990s that shows major changes in enrollment rates at High School and College level between rather than within three-years intervals.

ability level at High School that was always much lower than at College level and steeply declining over birth cohorts for both males and females. Expected values of ability levels are multiplied by ϑ_{3tS} measuring variation in the impact of unobserved heterogeneity over time and by education level. The estimated values for each education level in the three different time intervals together with their standard errors are reported in Table 5.

As can be seen from the table, for both males and females, ϑ_3 is always positive and significant at High School level while at College level it is significant only in the last period for males and in the second and third period for females. At Secondary level it is still significant and positive for all periods for males, while it is significant at the ten per cent level only in the first two periods in the female sample. At Primary level it is insignificant in the last period for males and in the last two periods for females. It is interesting to note how the largest differences in the magnitude of the estimated ϑ_3 in the male with respect to the female sample appear at Secondary and High School levels.

4.2.3 Composition effects and policy interventions

The analysis in the previous paragraphs has shown the role of variations in schooling quality and induced compositional changes to explain the evolution of wage dynamics. Improving quality of education at intermediate levels could ease the supply constraints and enable more students to sit successfully for the competitive University entry exam. However, at the same time it will induce compositional changes that will lower the average ability level of College graduates and in turn affect the evolution of earnings' inequality.

In order to evaluate the importance of composition effects and gain some idea of the magnitude of the dilution of quality, we can examine the impact of an increase in the quality of schooling on enrolment rates and the distribution of unobservables. The graphs in figures 5 and 6 show the effects of a ten per cent increase in the number of teachers per school on College and High School graduates by birth cohort for males and females. Table 6 summarizes the results for the most recent cohort.

The quality improvement increases enrollment rates to a maximum of 2.8 per cent at College and of around 10 per cent at High School level. For the most recent cohort there is a probability of around 9 (6) per cent that an individual from the upper half of the ability distribution graduates from College and of around 10 (7) per cent from High School for the males (females) sample. However, together with high-ability students, there will be many low-ability new entrants so that the net impact on average ability will depend on the relative proportion of the two ability groups. Looking at the youngest cohort, while the probability of being a high ability type increases at High School level from a value of 13.8 (7.8) per cent before the policy is implemented to a value of 17.5 (11.6) per cent after the intervention for males (females), at College level from a value higher than 85 (75) per cent before the increase in schooling quality it decreases to a value of around 59 (38) per cent for males (females).

The decrease in average ability at College level will affect the level of wages received by the newly induced College graduates. Given their lower ability level, their wages will be lower on average than the mean wage of the original College graduates. This result is in line with the findings in Cameron and Heckman (1998) that evaluate the impact of a ten per cent increase in family income on College enrollment and graduation rates in the US.

Unobserved heterogeneity can be interpreted as unobserved ability, intelligence or aptitude that is an innate and permanent characteristic observed by the individual by not by the analyst. However, only part of h is properly attributed to ability, since it also includes all unobserved and idiosyncratic costs related to availability/quality of schooling. The estimates of the policy impact are therefore an upper-bound of the effect of changes in policy on the composition of students enrolled in school. However, at the same time, h is the component of ability that is uncorrelated with observables; to the extent that ability is correlated with some of the observables, the estimates will understate the effect of the policy intervention on the composition of ability because all observables are held constant in the simulation. As noted by Cameron and Heckman

(1998), a more precise analysis would require breaking unobserved heterogeneity into components due to costs and ability.

Even if a complete assessment of policy changes should take into account general equilibrium effects, [7] the results show the importance of accounting for unobserved heterogeneity and its conditional distribution to ex-ante evaluate the impact of interventions that promote schooling attendance and are likely to induce compositional changes that will interact with the policy targets.

5 Conclusion

The analysis developed in the paper provides a unified framework to study schooling attainment and the evolution of relative returns and offers a more complete picture on how returns to education have been changing in Brazil during the 1990s.

In the last decades Brazil has been characterized by a low rate of educational progress with respect to other Latin American Countries. While the number of individuals with completed High School has significantly increased, graduation rates at College level have been stagnant. The significant increase in relative returns to College did not induce more individuals to enroll into University.

The paper studies educational choices and investigates the reasons for the lack of growth in the number of College graduates. The decline in the quality of the High School system could explain both the inability of pupils to move to College and the decline in the returns to High School. A decline in the quality of the High School system due to an increase in the number of students not accompanied by a proportional increase in schools and teachers' provision could have restricted entrance into higher education to high-ability students and to those that attended high-quality intermediate

[7] The analysis assumes a partial equilibrium framework where skill prices are fixed and do not respond to supply variations. However, increases in the enrollment rate of one schooling level will correspond to a decrease in relative wages of the corresponding schooling group. Therefore, a complete assessment of the impact of policy changes should take into account the interaction between changes in the market price of skills and variations in relative supplies.

schools that prepared them to seat for the competitive College exams. Students in higher education will be a progressively selective sample of individuals and returns to unobservables will represent an important factor affecting the evolution of returns to schooling.

The empirical results show the role of availability and quality of schooling to promote skill-upgrading and bring evidence of compositional changes as an important determinant of changes in relative wages. The estimates bring evidence of a dramatic deterioration over time in the average ability of students at intermediate levels of education and an increase in graduation among the most able at College level. For the youngest cohorts, among High School graduates there is a high proportion of low-ability individuals while most high-ability individuals have completed College.

The results have implications for the design of effective interventions to promote investments in higher education. A simulation of the effects of an increase in schooling quality on workforce composition shows how policy interventions can induce substantial compositional changes that will in turn affect changes in relative wages. Improvements in education quality at intermediate levels could be an effective device to allow more students to move further up in the education ladder. However, induced compositional changes should also be accounted for in order to predict the net impact of the policy intervention on changes in earnings' inequality.

The paper can be extended in three main directions. First, unobserved heterogeneity could be allowed to follow a multiple-factor distribution with different factors affecting schooling choice and labor market outcomes. Secondly, returns to schooling could be separately estimated for public and private schools using the information on public/private school attendance which is available in PNAD since the year 2001. Finally, the framework could be completed with the estimation of a production function to model the way the prices of the different types of human capital have been evolving over time and assess the extent to which the various education levels are substitutable for each other. The estimated prices from the wage equation could be used to obtain estimates of the parameters of the production function directly.

References

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

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

[3] 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

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

[5] 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

[6] Carneiro P., Lee S. (2005) "Ability, Sorting and Wage Inequality", Cemmap Working Paper CWP16/05

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

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

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

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

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

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

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

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

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

Table 1

Variable	**MALES** Mean	SD	**FEMALES** 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
GDP share	0.076	0.093	0.079	0.095
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
GDP share	0.098	0.112	0.094	0.110
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 2

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
GDP share	0.105	0.119	0.097	0.114
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
GDP share	0.137	0.132	0.128	0.129
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

Figure 1

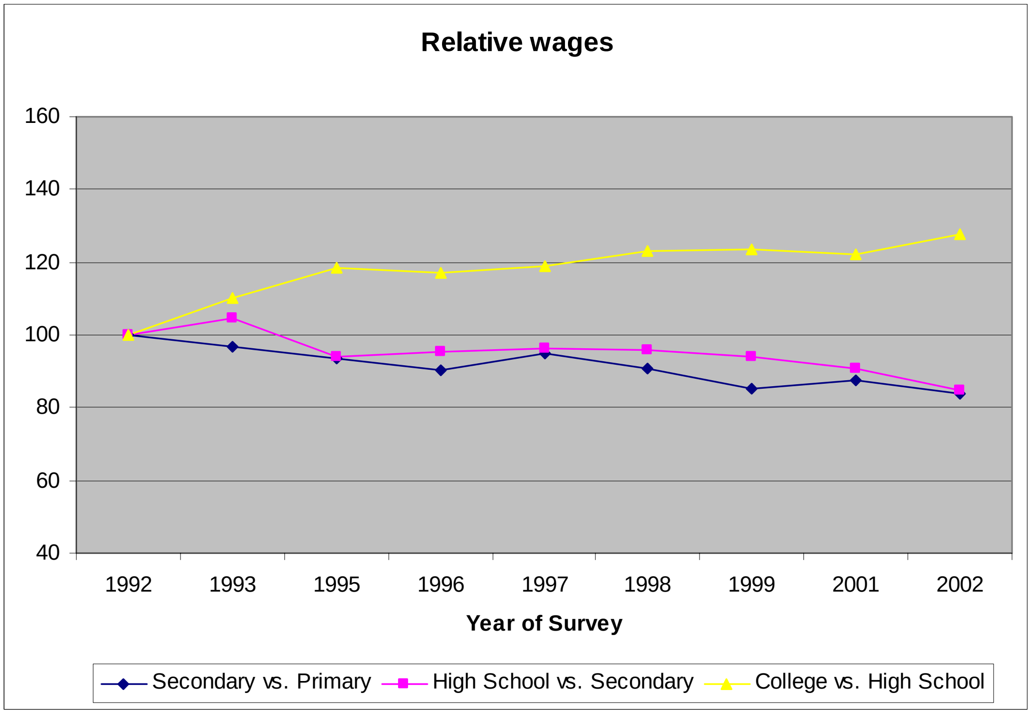


Figure 2

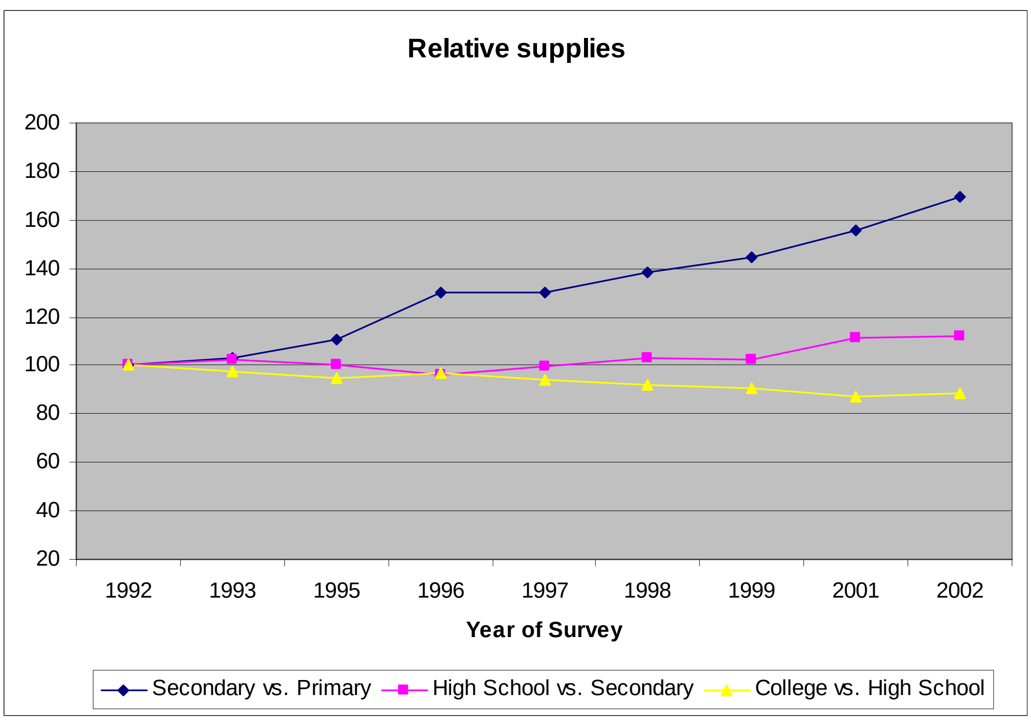
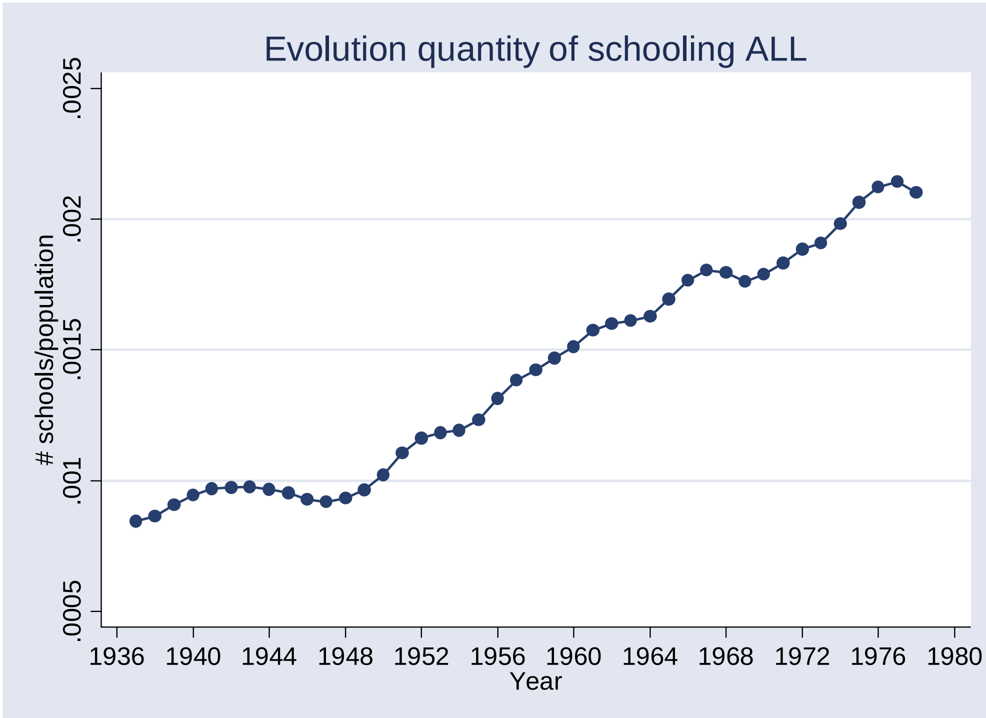


Figure 3


Evolution quantity of schooling ALL
schools/population
.0005
.001
.0015
.002
.0025
1936
1940
1944
1948
1952
1956
1960
1964
1968
1972
1976
1980
Year

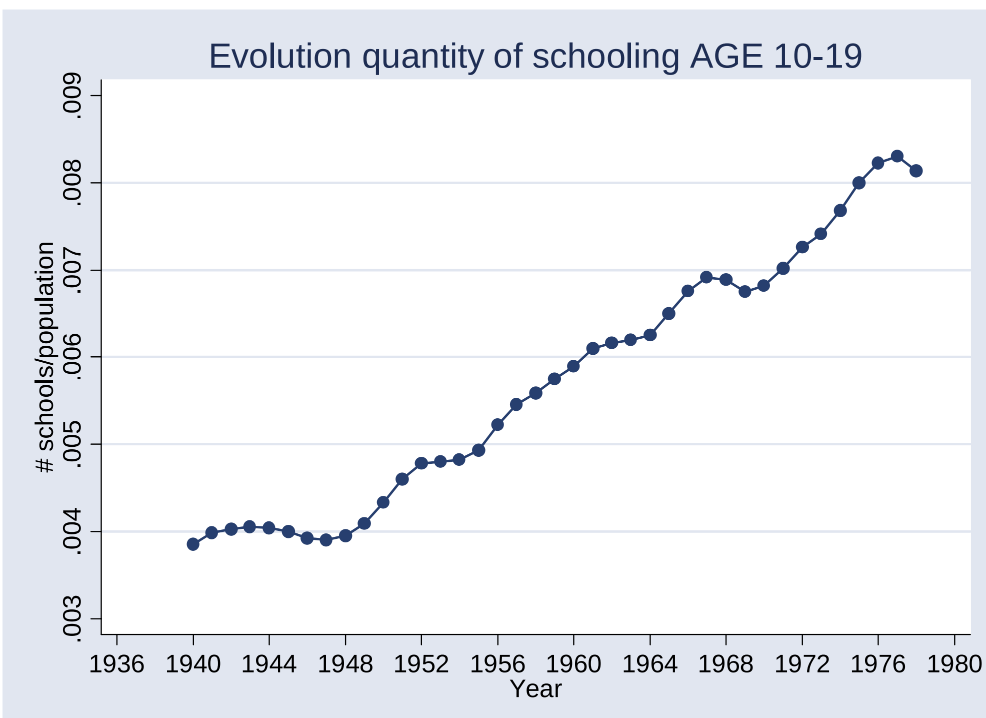
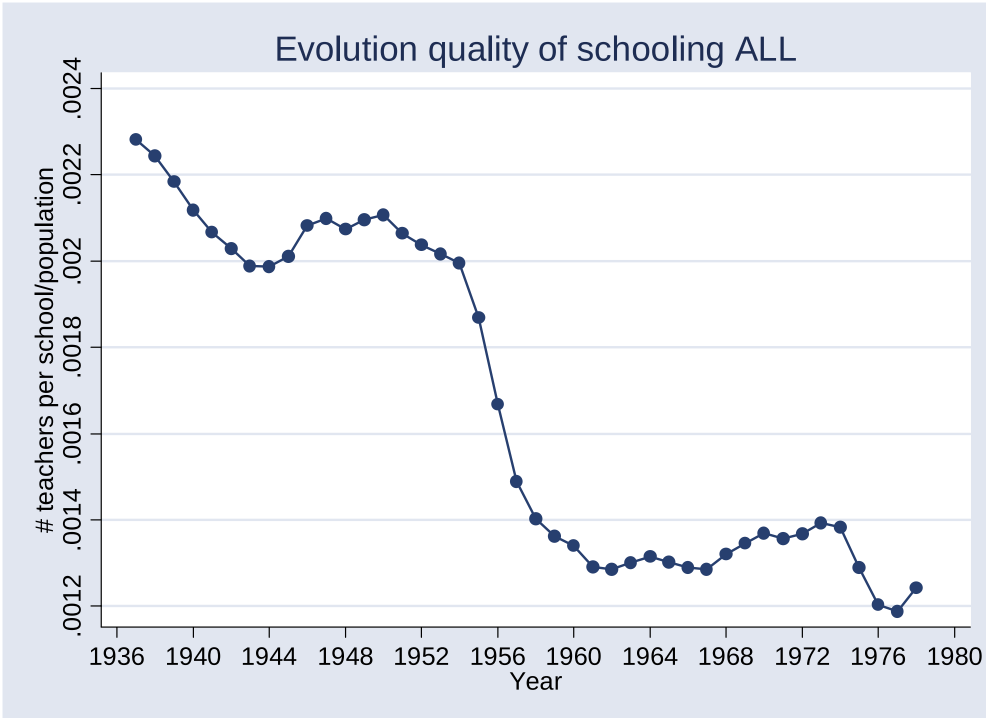

Evolution quantity of schooling AGE 10-19
schools/population
.003
.004
.005
.006
.007
.008
.009
1936
1940
1944
1948
1952
1956
1960
1964
1968
1972
1976
1980
Year

Figure 4


Evolution quality of schooling ALL
teachers per school/population
.0012
.0014
.0016
.0018
.002
.0022
.0024
1936
1940
1944
1948
1952
1956
1960
1964
1968
1972
1976
1980
Year

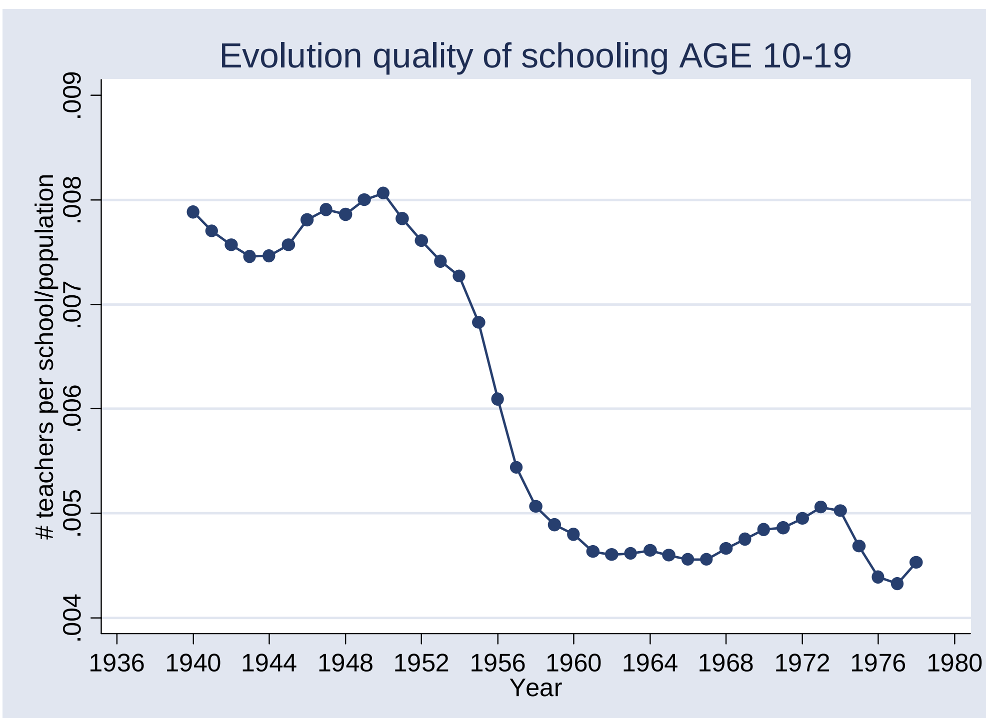

Evolution quality of schooling AGE 10-19
teachers per school/population
.004
.005
.006
.007
.008
.009
1936
1940
1944
1948
1952
1956
1960
1964
1968
1972
1976
1980
Year

Table 3*

Schooling choice

	MALES	FEMALES
raced1	-0.9589	-0.8521
	(0.0261)	(0.0250)
raced2	-1.5968	-1.4867
	(0.0278)	(0.0267)
raced4	-1.5186	-1.3987
	(0.0265)	(0.0254)
pop_b	0.1797	0.0963
	(0.0393)	(0.0373)
GDP_b	2.1104	2.6802
	(0.2004)	(0.1946)
school_b	0.0744	0.0910
	(0.0217)	(0.0205)
prof_b	0.1896	0.1622
	(0.0309)	(0.0291)
	# obs = 568477	# obs = 618255

***** Standard errors in parentheses. The model includes dummy variables for State and year of birth. Dependent variable is schooling level. The model is estimated with two points of support.

<u>Explanatory variables</u>:
raced: race dummy (1=white; 2=black; 3=natives; 4=mulatos)
pop_b: log(population) by State and year of birth
GDP_b: share of State in national GDP at year of birth
school_b: log(# schools/population) by State and year of birth
prof_b: log(# teachers per school/population) by State and year birth

Table 4*

Labor market participation

	FEMALES	MALES
Secondary	0.2093 (0.0047)	0.1336 (0.0061)
High School	0.5403 (0.0078)	0.3596 (0.0128)
College	1.0623 (0.0193)	0.6177 (0.0243)
raced1	0.1369 (0.0277)	0.0389 (0.0089)
raced2	0.3567 (0.0287)	-0.1676 (0.0350)
raced4	0.1998 (0.0280)	0.0181 (0.0087)
age	4.0811 (0.0661)	3.9012 (0.0787)
agesq	-2.1463 (0.0317)	-1.9042 (0.0374)
child	-0.0977 (0.0063)	-0.3074 (0.0074)
nchildr	-0.0181 (0.00013)	-0.0511 (0.0015)
spouse	-0.4872 (0.0041)	0.2164 (0.0059)
childrd01	-0.3546 (0.0069)	0.6429 (0.0084)
childrd26	-0.1089 (0.0053)	0.5942 (0.0067)
childrd715	0.0557 (0.0047)	0.4121 (0.0059)
const	-1.7811 (0.0474)	-0.8865 (0.0473)
	# obs = 618255	# obs = 568477

* Standard errors in parentheses. The model includes dummy variables for State of birth and year of Survey. Dependent variable is binary indicator working/not working. The model is estimated with two points of support.

raced: race dummies (1=white; 2=black; 3=natives; 4=mulatos)
age, agesq: age, age^2
child: dummy for having a child
nchildr: # children in the hh
spouse: dummy for whether there's a spouse in the hh
childrd: dummy children's age group (excluded category 16 and over)

Table 5 - Distribution of unobserved heterogeneity[1]

MALES

UH low-ability = 0 UH high-ability = 2.6311** (0.0159)
Prob(low-ability)= 0.8287 Prob(high-ability)= 0.1713

Participation equation

theta2 = 0.1253** (0.0089)

Wage equations

theta3	1992-1995	1996-1998	1999-2002
Primary	0.6065** (0.0230)	0.2390** (0.1062)	0.1868 (0.1927)
Secondary	0.0998** (0.0300)	0.1209** (0.0324)	0.1143** (0.0392)
High School	0.1093** (0.0121)	0.1257** (0.0117)	0.1126** (0.0117)
College	0.0749 (0.0466)	0.0625 (0.0425)	0.0784** (0.0333)

FEMALES

UH low-ability = 0 UH high-ability = 2.5737** (0.0149)
Prob(low-ability)= 0.8389 Prob(high-ability)= 0.1602

Participation equation

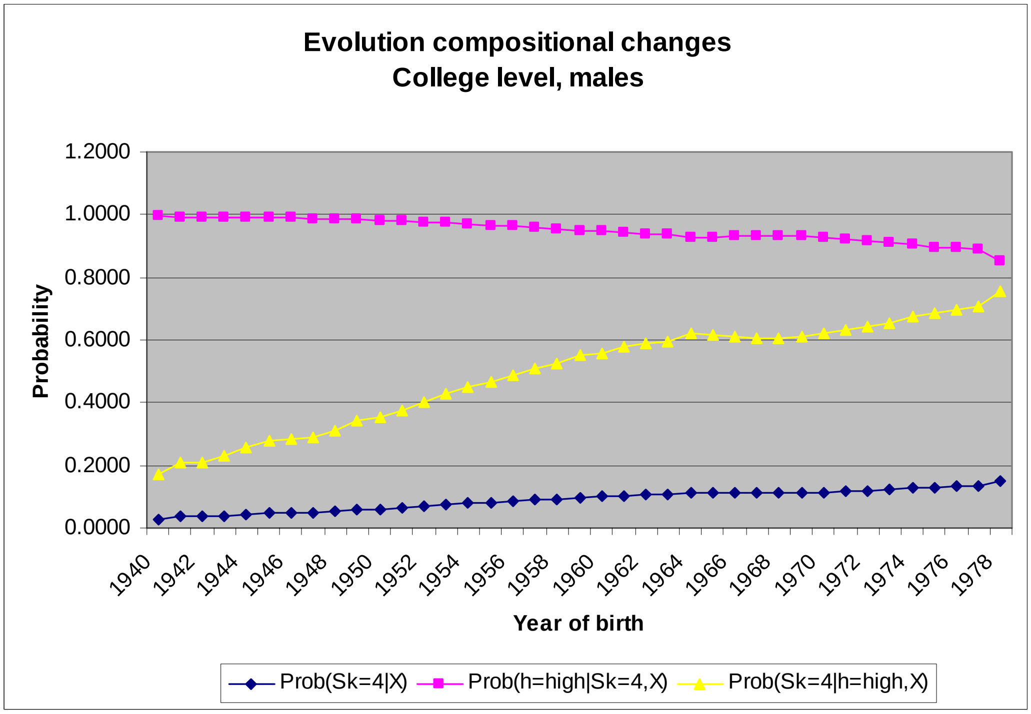
theta2 = 0.0145** (0.0078)

Wage equations

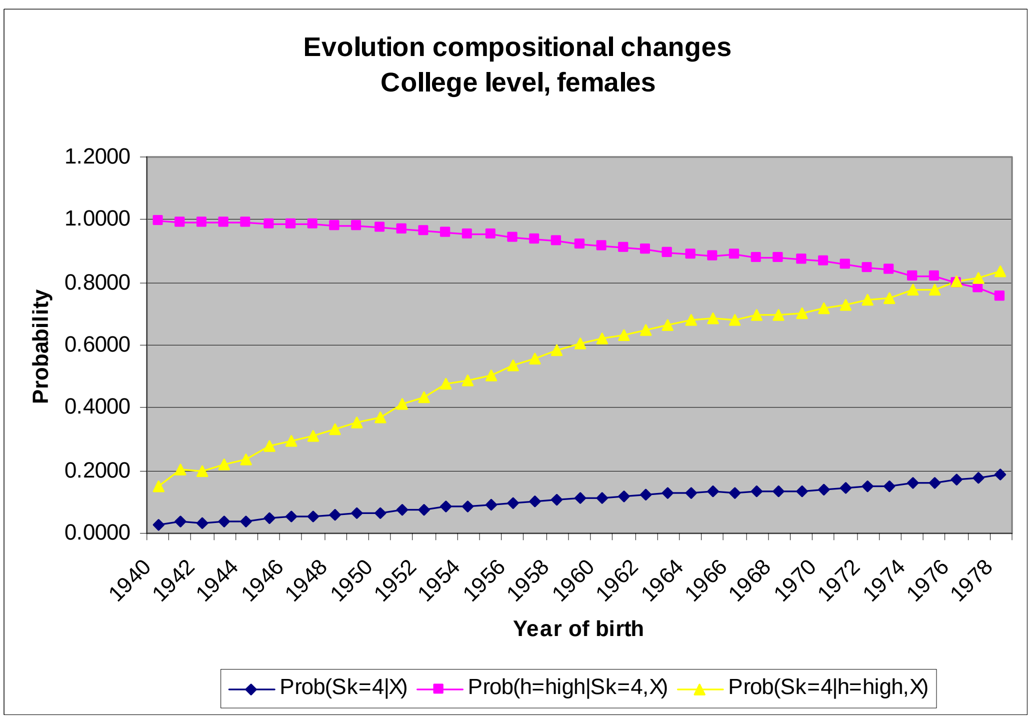
theta3	1992-1995	1996-1998	1999-2002
Primary	0.6902** (0.0416)	0.1096 (0.3147)	0.0932 (0.0670)
Secondary	0.0790* (0.0447)	0.0919* (0.0552)	0.0932 (0.4740)
High School	0.1053** (0.0164)	0.0984** (0.0168)	0.0930** (0.0171)
College	0.0674 (0.0453)	0.0768** (0.0371)	0.0766** (0.0295)

[1] ** (*) indicates significance at 5(10) per cent level

Panel 1

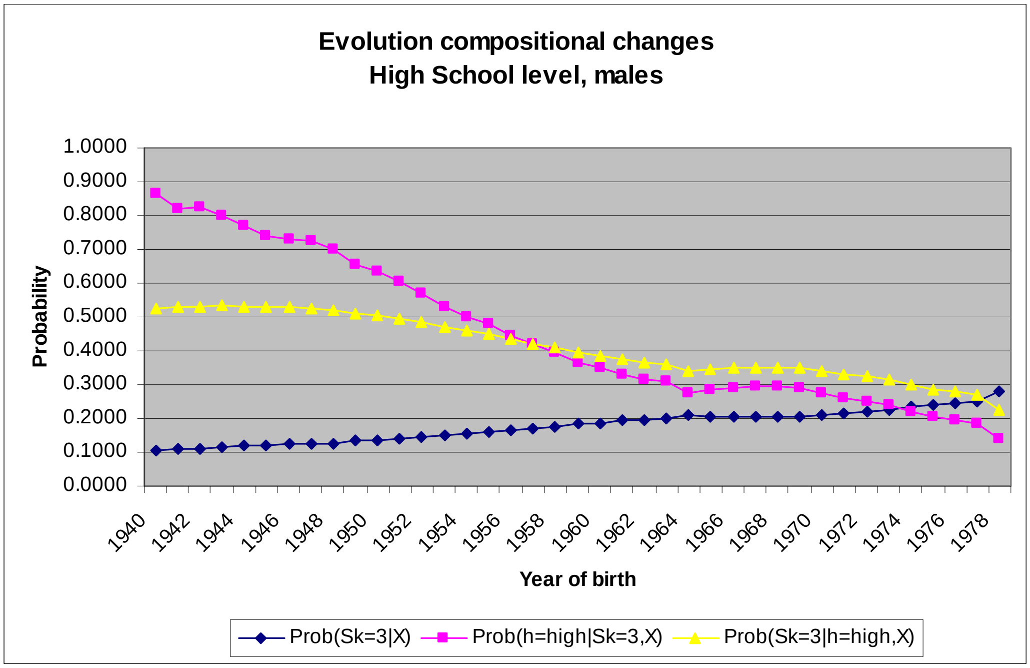


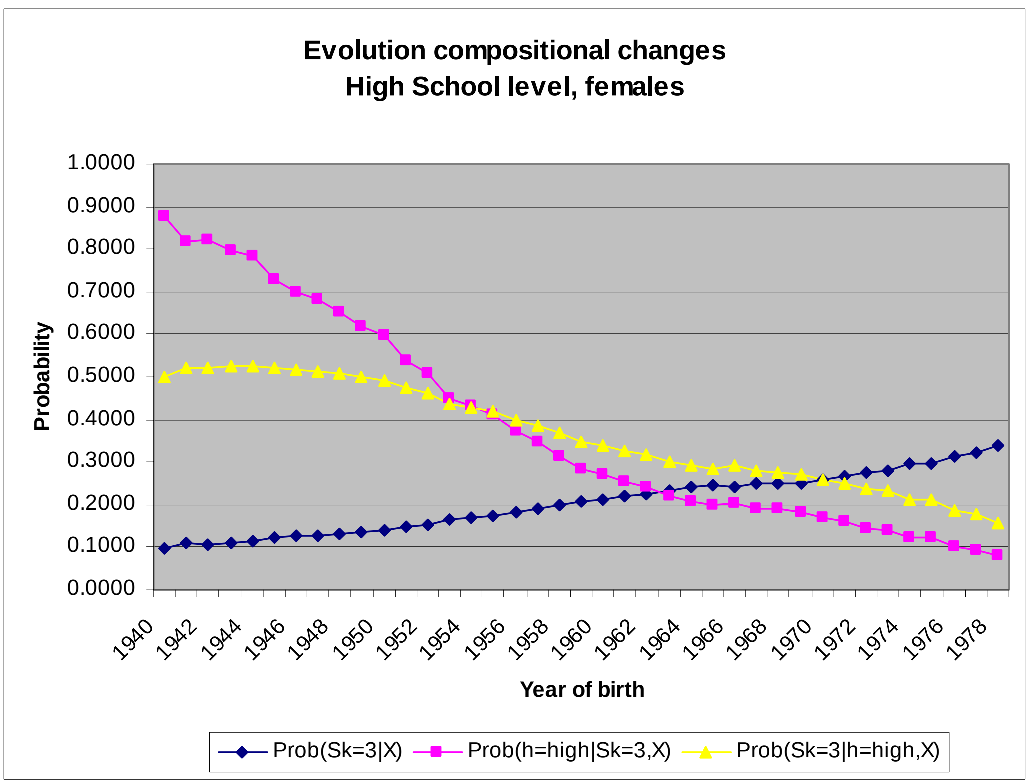
Evolution compositional changes
College level, males
Probability
1.2000
1.0000
0.8000
0.6000
0.4000
0.2000
0.0000
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
Year of birth
Prob(Sk=4|X)
Prob(h=high|Sk=4,X)
Prob(Sk=4|h=high,X)



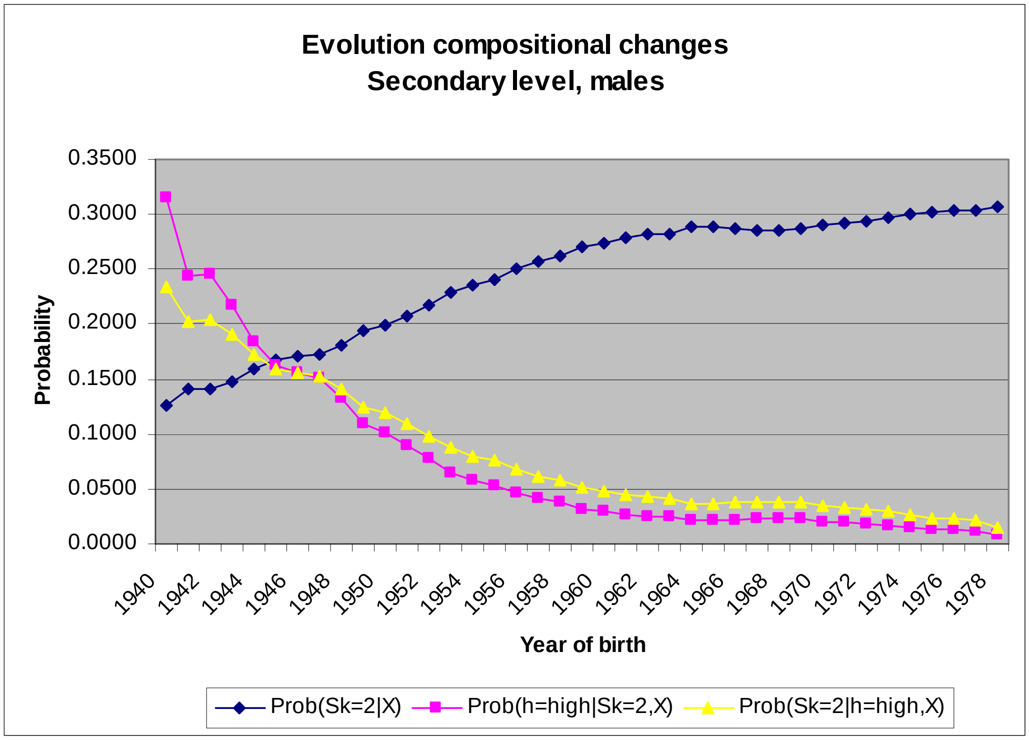
Evolution compositional changes
College level, females
Probability
1.2000
1.0000
0.8000
0.6000
0.4000
0.2000
0.0000
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
Year of birth
Prob(Sk=4|X)
Prob(h=high|Sk=4,X)
Prob(Sk=4|h=high,X)

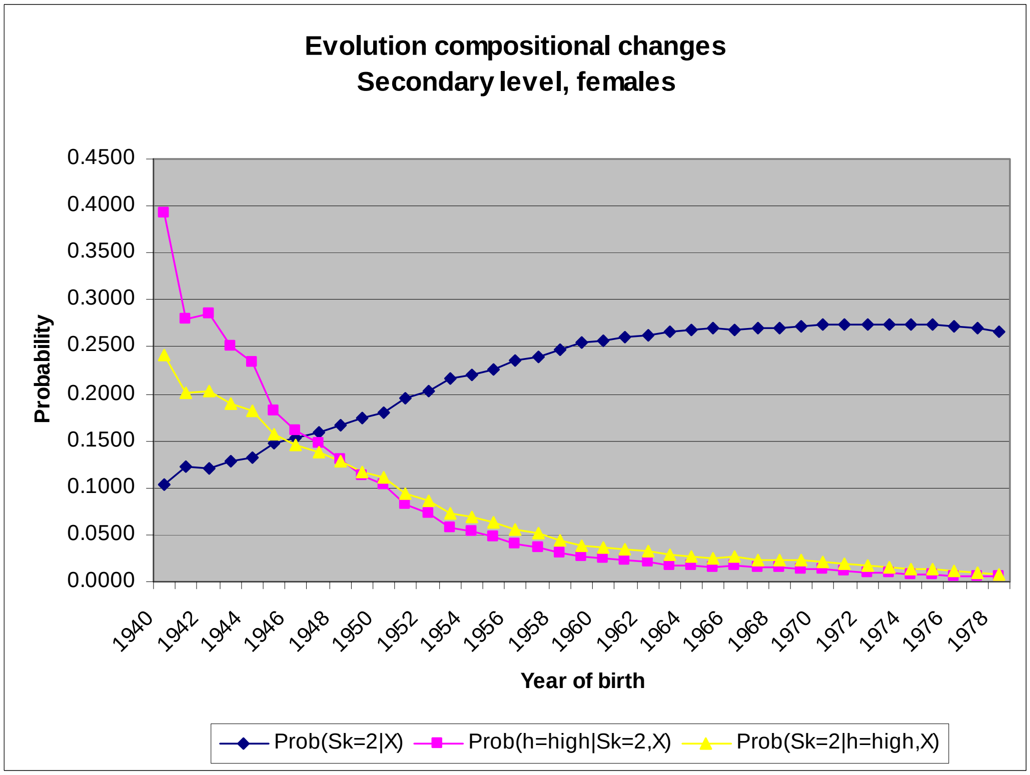
Panel 2


Evolution compositional changes
High School level, males
Probability
1.0000
0.9000
0.8000
0.7000
0.6000
0.5000
0.4000
0.3000
0.2000
0.1000
0.0000
1940 1942 1944 1946 1948 1950 1952 1954 1956 1958 1960 1962 1964 1966 1968 1970 1972 1974 1976 1978
Year of birth
Prob(Sk=3|X)
Prob(h=high|Sk=3,X)
Prob(Sk=3|h=high,X)

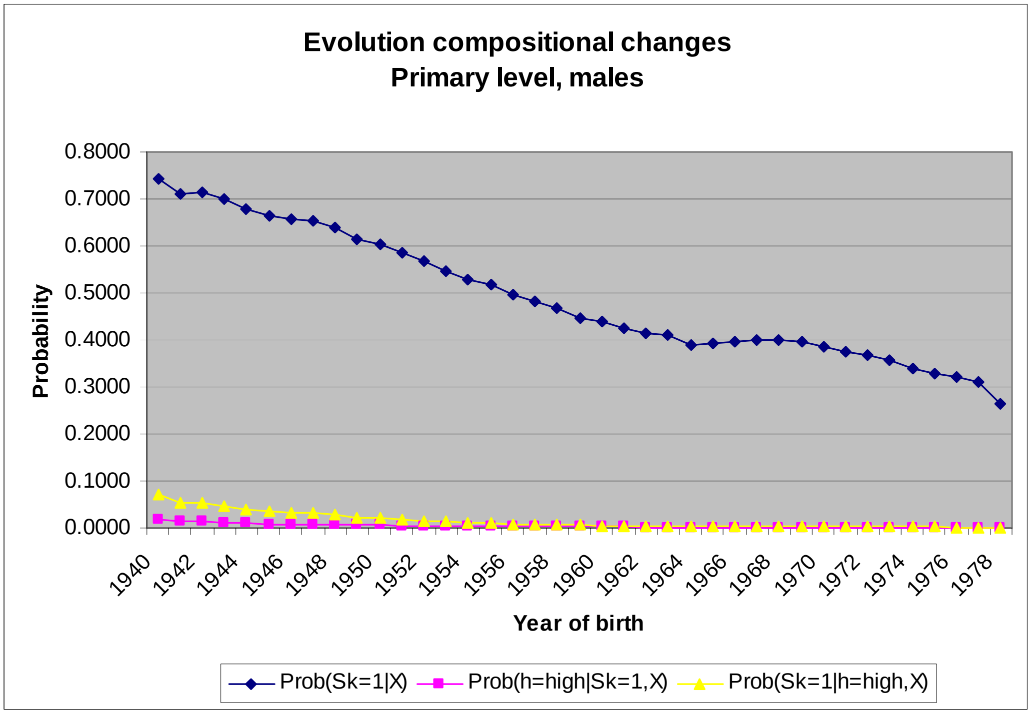

Evolution compositional changes
High School level, females
Probability
1.0000
0.9000
0.8000
0.7000
0.6000
0.5000
0.4000
0.3000
0.2000
0.1000
0.0000
1940 1942 1944 1946 1948 1950 1952 1954 1956 1958 1960 1962 1964 1966 1968 1970 1972 1974 1976 1978
Year of birth
Prob(Sk=3|X)
Prob(h=high|Sk=3,X)
Prob(Sk=3|h=high,X)

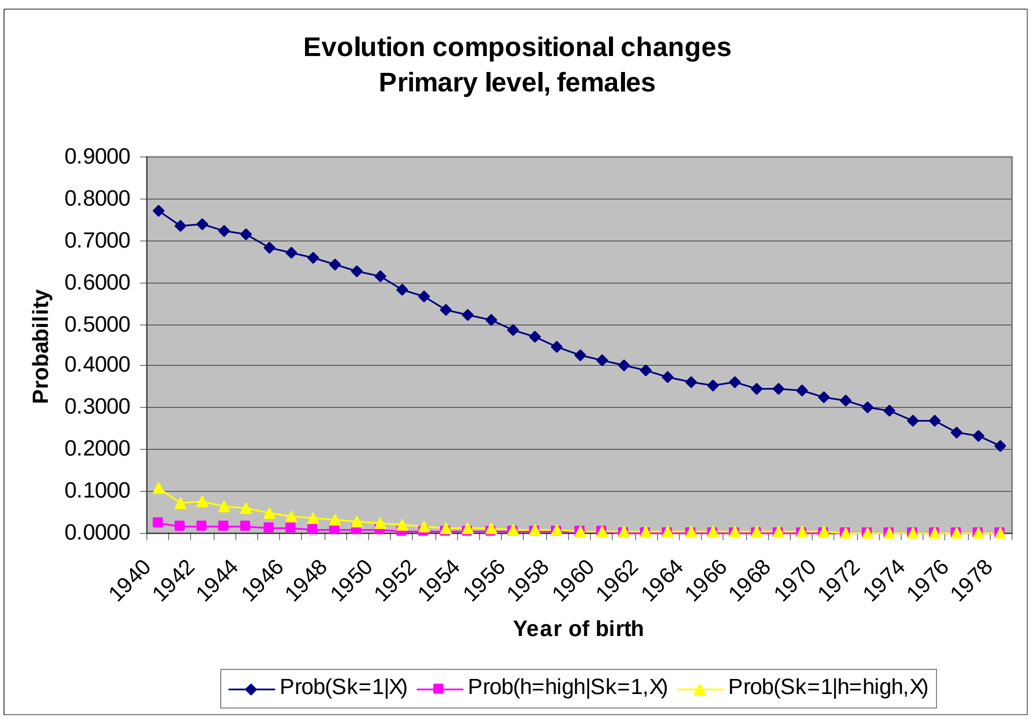
Panel 3


Evolution compositional changes
Secondary level, males
Probability
0.3500
0.3000
0.2500
0.2000
0.1500
0.1000
0.0500
0.0000
1940 1942 1944 1946 1948 1950 1952 1954 1956 1958 1960 1962 1964 1966 1968 1970 1972 1974 1976 1978
Year of birth
Prob(Sk=2|X)
Prob(h=high|Sk=2,X)
Prob(Sk=2|h=high,X)

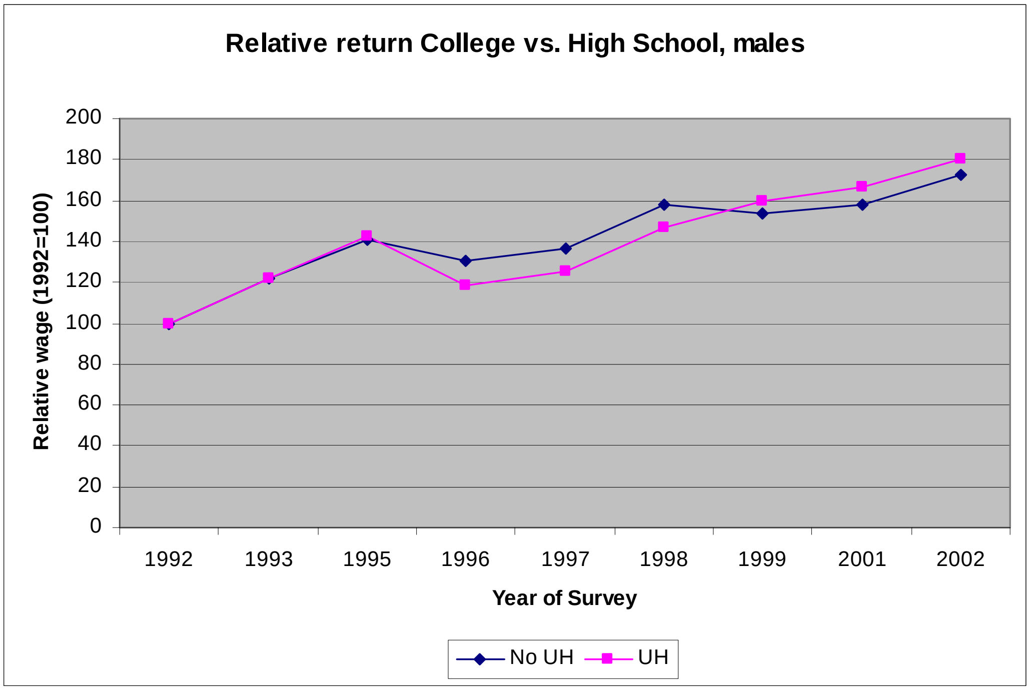

Evolution compositional changes
Secondary level, females
Probability
0.4500
0.4000
0.3500
0.3000
0.2500
0.2000
0.1500
0.1000
0.0500
0.0000
1940 1942 1944 1946 1948 1950 1952 1954 1956 1958 1960 1962 1964 1966 1968 1970 1972 1974 1976 1978
Year of birth
Prob(Sk=2|X)
Prob(h=high|Sk=2,X)
Prob(Sk=2|h=high,X)

Panel 4

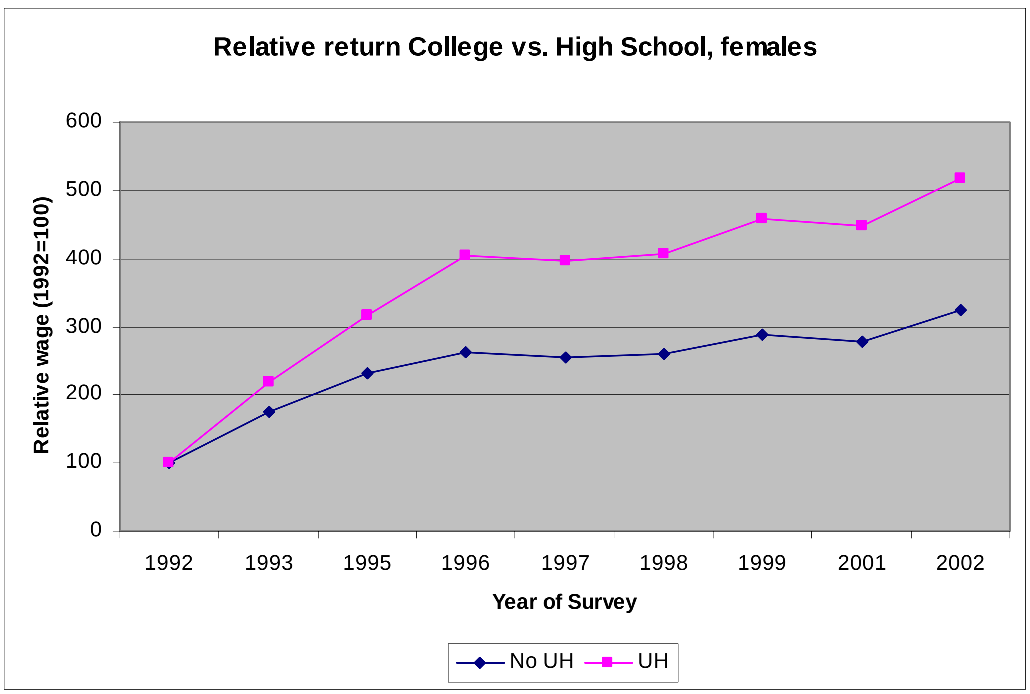

Evolution compositional changes
Primary level, males
Probability
0.8000
0.7000
0.6000
0.5000
0.4000
0.3000
0.2000
0.1000
0.0000
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
Year of birth
Prob(Sk=1|X)
Prob(h=high|Sk=1,X)
Prob(Sk=1|h=high,X)


Evolution compositional changes
Primary level, females
Probability
0.9000
0.8000
0.7000
0.6000
0.5000
0.4000
0.3000
0.2000
0.1000
0.0000
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
Year of birth
Prob(Sk=1|X)
Prob(h=high|Sk=1,X)
Prob(Sk=1|h=high,X)

Panel 5

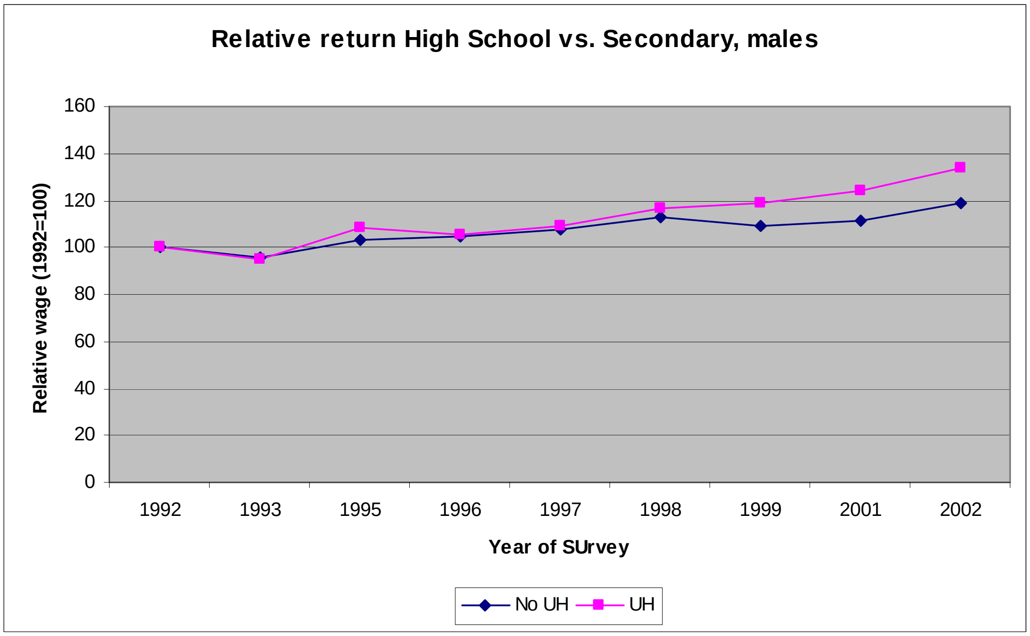


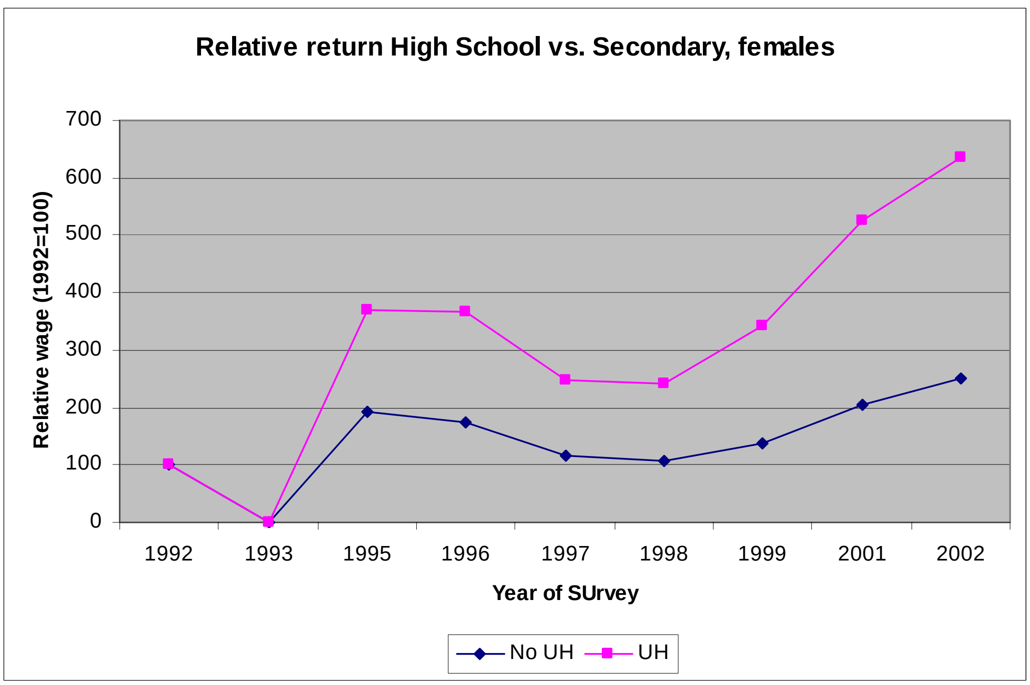
Relative return College vs. High School, males
Relative wage (1992=100)
200
180
160
140
120
100
80
60
40
20
0
1992
1993
1995
1996
1997
1998
1999
2001
2002
Year of Survey
No UH
UH



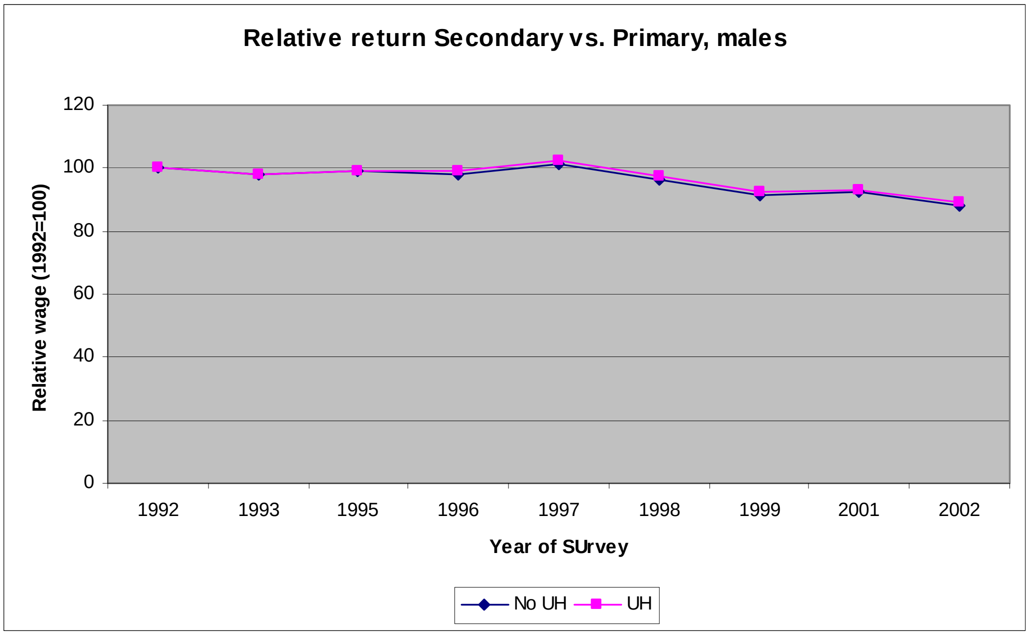
Relative return College vs. High School, females
Relative wage (1992=100)
600
500
400
300
200
100
0
1992
1993
1995
1996
1997
1998
1999
2001
2002
Year of Survey
No UH
UH

Panel 6

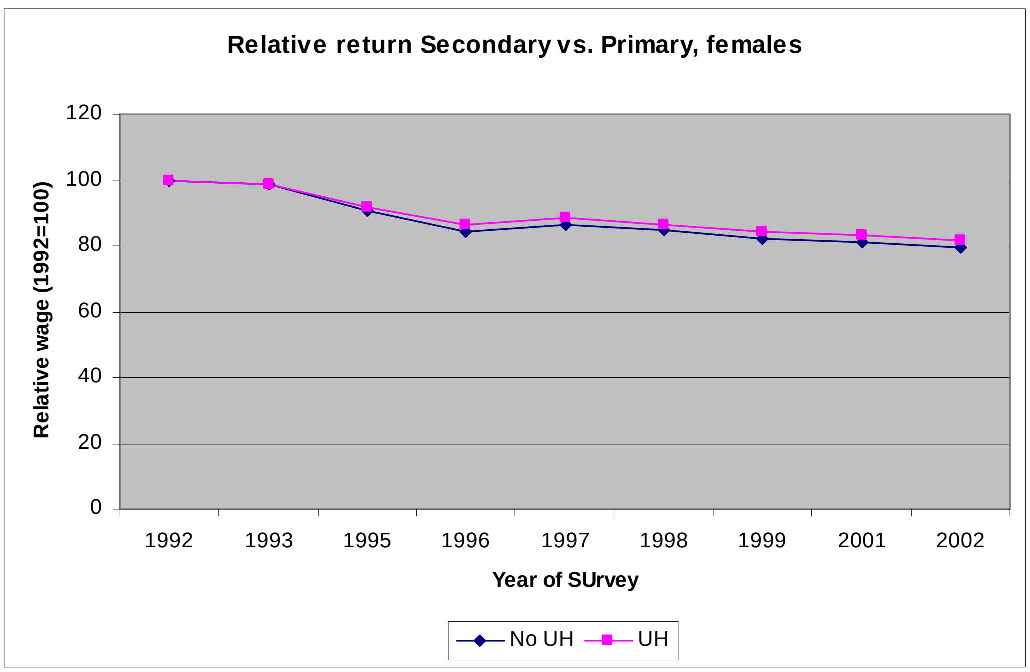

Relative return High School vs. Secondary, males
Relative wage (1992=100)
0
20
40
60
80
100
120
140
160
1992
1993
1995
1996
1997
1998
1999
2001
2002
Year of SUrvey
No UH
UH


Relative return High School vs. Secondary, females
Relative wage (1992=100)
0
100
200
300
400
500
600
700
1992
1993
1995
1996
1997
1998
1999
2001
2002
Year of SUrvey
No UH
UH

Panel 7

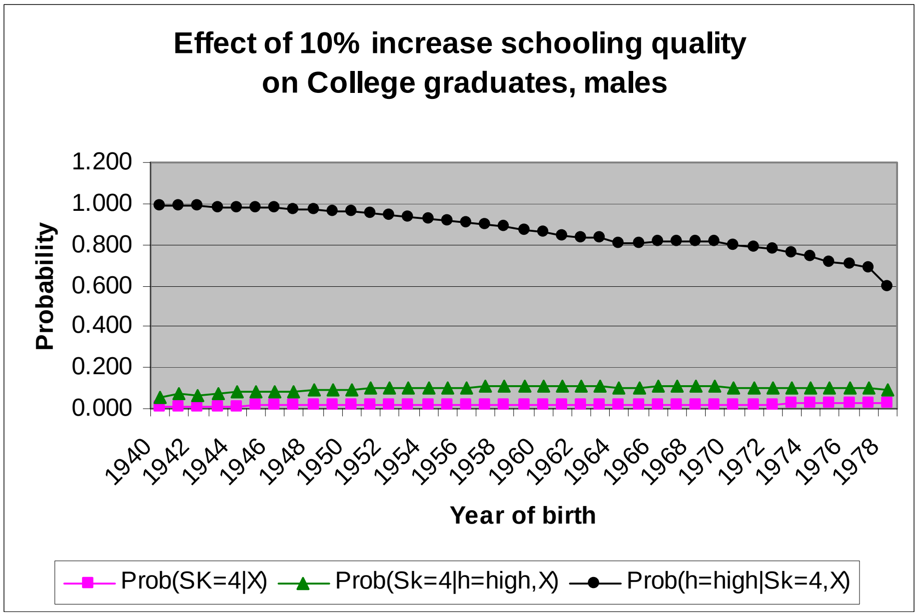


Relative return Secondary vs. Primary, males
120
100
80
60
40
20
0
Relative wage (1992=100)
1992
1993
1995
1996
1997
1998
1999
2001
2002
Year of SUrvey
No UH
UH



Relative return Secondary vs. Primary, females
120
100
80
60
40
20
0
Relative wage (1992=100)
1992
1993
1995
1996
1997
1998
1999
2001
2002
Year of SUrvey
No UH
UH

Figure 5


Effect of 10% increase schooling quality
on College graduates, males
Probability
1.200
1.000
0.800
0.600
0.400
0.200
0.000
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
Year of birth
Prob(SK=4|X)
Prob(Sk=4|h=high,X)
Prob(h=high|Sk=4,X)

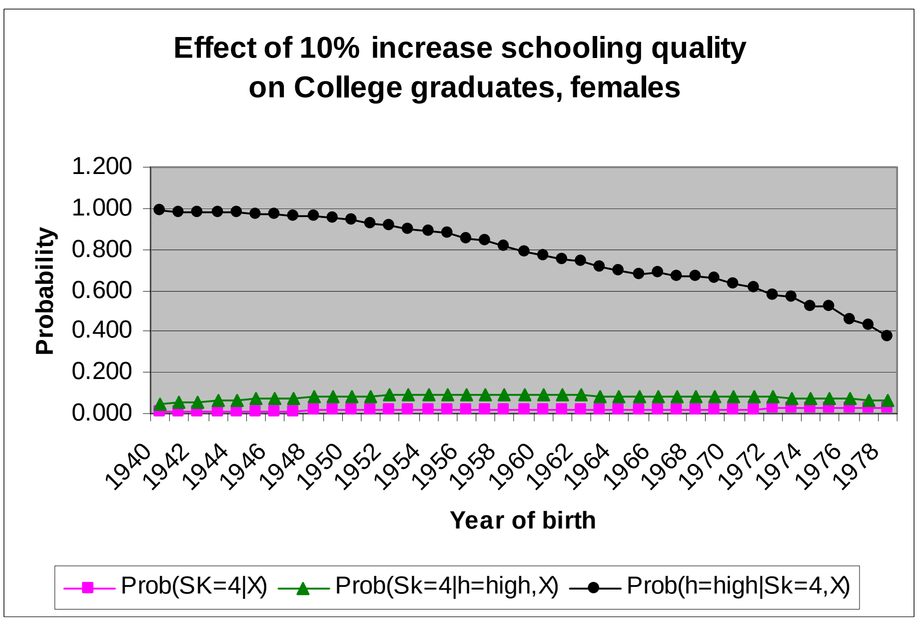
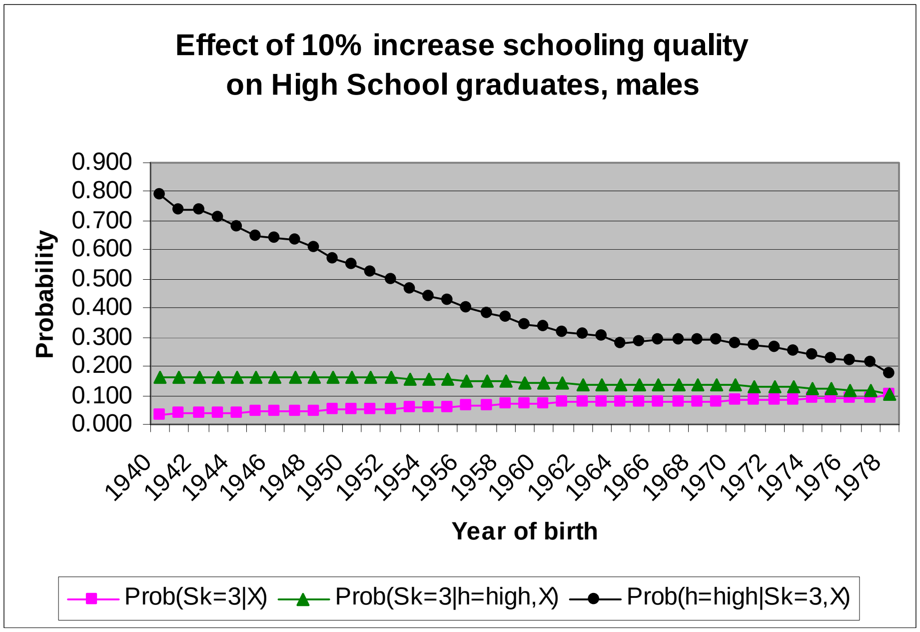

Effect of 10% increase schooling quality
on College graduates, females
Probability
1.200
1.000
0.800
0.600
0.400
0.200
0.000
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
Year of birth
Prob(SK=4|X)
Prob(Sk=4|h=high,X)
Prob(h=high|Sk=4,X)

Figure 6


Effect of 10% increase schooling quality
on High School graduates, males
Probability
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
Year of birth
Prob(Sk=3|X)
Prob(Sk=3|h=high,X)
Prob(h=high|Sk=3,X)

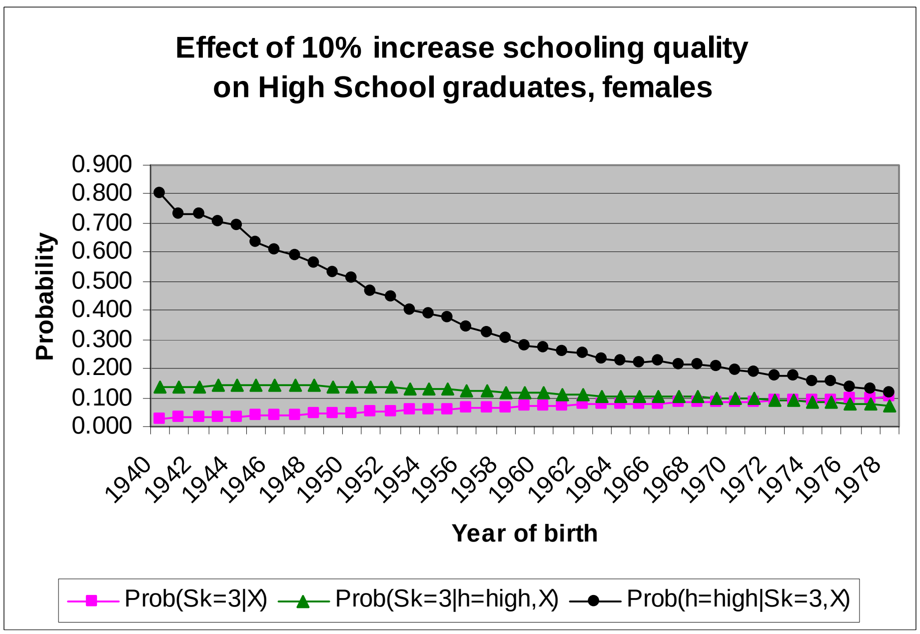

Effect of 10% increase schooling quality
on High School graduates, females
Probability
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
1940
1942
1944
1946
1948
1950
1952
1954
1956
1958
1960
1962
1964
1966
1968
1970
1972
1974
1976
1978
Year of birth
Prob(Sk=3|X)
Prob(Sk=3|h=high,X)
Prob(h=high|Sk=3,X)

Table 6

MALES

Prob.	Prob(Sk=3|X)	Prob(Sk=3|h=high,X)	Prob(h=high|Sk=3,X)
Before Policy	27.93%	22.69%	13.81%
After Policy	10.20%	10.50%	17.50%

Prob.	Prob(Sk=4|X)	Prob(Sk=4|h=high,X)	Prob(h=high|Sk=4,X)
Before Policy	15.09%	75.75%	85.35%
After Policy	2.60%	9.10%	59.40%

FEMALES

Prob.	Prob(Sk=3|X)	Prob(Sk=3|h=high,X)	Prob(h=high|Sk=3,X)
Before Policy	33.90%	15.70%	7.80%
After Policy	10.20%	6.90%	11.60%

Prob.	Prob(Sk=4|X)	Prob(Sk=4|h=high,X)	Prob(h=high|Sk=4,X)
Before Policy	18.76%	83.53%	75.68%
After Policy	2.80%	6.30%	37.90%