IZA/CEPR 10th European Summer Symposium in Labour Economics

Supported and hosted by the Institute for the Study of Labor (IZA)

Ammersee

9-11 October 2008

Does Speed Signal Ability? The Impact of Grade Repetitions on Employment and Wages

Thomas Brodaty, Robert J. Gary-Bobo and Ana Prieto

We are grateful to the following institutions for their financial and organizational support: the Institute for the Study of Labor (IZA)

Does Speed Signal Ability? The Impact of Grade Repetitions on Employment and Wages*

Thomas O. Brodaty*, Robert J. Gary-Bobo† and Ana Prieto‡

14 May 2008

Abstract

We propose a new test for the presence of job-market signaling in the sense of Spence (1973), based on an equation in which log-wages are explained by two endogenous variables: the student's degree and the student's time to degree, not simply by years of education. Log-wages are regressed on a measure of *education*, which is a position on a scale of certificates and degrees, and a measure of the student *delay*, defined as the difference between the individual's school-leaving age and the average school-leaving age of students holding the same certificate or degree. We use past school-opening instruments, and distance-to-the-nearest-college, also measured in the past, when students were entering grade 6, to identify the parameters. We find a robust, significant and negative impact of the delay variable on wages, averaged over the first five years of career. A year of delay causes a 9% decrease of the student's wage. The only reasonable explanation for this effect is the fact that longer delays signal unobserved characteristics with a negative productivity value. We finally estimate a nonlinear model of education choices and cannot reject the assumption that the data is generated by a job-market signaling equilibrium.

*University of Cergy-Pontoise, THEMA, France. email: thomas.brodaty@eco.u-cergy.fr

†CREST-INSEE and Paris School of Economics, Université Paris 1, 106-112 boulevard de l'Hôpital, 75647 Paris Cedex 13, France. email: garybobo@univ-paris1.fr

‡CNRS, THEMA, France. email: ana.prieto@eco.u-cergy.fr

*We thank the Institut Geographique National for the geographical data file used in this article, Denis Fougere, Thierry Kamionka, Francis Kramarz, Guy Laroque, Edwin Leuven, Thierry Magnac, David Margolis, Eric Maurin and Jean-Marc Robin for their help and remarks at various stages of this research. We also thank Kjell Salvanes, Steve Machin, organizers of CEPR conferences held in Bergen, Norway, and Madrid, Spain, in 2007; Christian Belzil, organizer of a conference at IZA, Bonn, spring 2007; the audiences of these meetings; the members of the Microdata network, as well as seminar audiences in Paris, Toulouse and Leuven for a number of very useful remarks.

1 Introduction

We propose a new test for the presence of job-market signaling in the sense of Spence (1973), based on an extension of the Mincer equation. In our model, log-wages are explained by two variables: the student's degree and the student's time to degree, not simply by years of education. There is a substantial amount of individual variability in time to degree-completion, conditional on the highest certificate or diploma earned by the student. We show that an appropriately defined function of time-to-degree, called *delay*, has a negative impact on wages at the beginning of a worker's career. To be more precise, we regress log-wages on a measure of *education*, which is a position on a scale of certificates and degrees and on a measure of the student *delay*, defined as the difference between the individual's school-leaving age and the average school-leaving age of students holding the same certificate or degree. These two variables happen to be orthogonal by construction. They are also potentially endogenous in the wage equation. It follows that two instruments at least are needed to identify and consistently estimate their coefficients by means of IV methods. To this end, we use past school-opening instruments, and distance-to-the-nearest-college, also measured in the past, when students were entering grade 6. Using 3SLS, we find a robust, significant and negative impact of the delay variable on wages, averaged over the first five years of career: a year of delay causes a 9% decrease of the student's wage. This figure is far from negligible and stems from the fact that delay conveys information about the unobserved ability of young workers.

To obtain these results, we first constructed a linear simultaneous-equations model that can be used to test for the presence of statistical discrimination based on delay. The return-to-education coefficients that we find are the sum of two effects: the productivity-enhancing effect and the signaling effect, the latter being due to the fact that education is correlated with unobserved ability characteristics, but only the sum of the two effects is identifiable. For the same reasons, the coefficient on delay in the wage equation is also potentially the sum of two effects, but a significant and negative impact of delay cannot be obtained in a human-capital model with fully informed employers, unless we are ready to assume that workers's experience and maturity, or partial completion of degrees, have a negative productivity, contrary to well-established empirical findings. Thus, signaling in the sense of Spence is the only likely reason for which delay has a significant and negative impact in the wage equation. A number of checks show that the result is fairly robust.

We then specify and estimate a nonlinear model of individual educational investment in which students choose an education level, based on private information, while bearing two sorts of risk: the risk affecting future wages and the risk affecting education costs, which is mainly due to time-to-degree uncertainty. The nonlinear model is derived from the assumption of expected utility maximization by imperfectly informed students, under rational expectations, and has very clear sources of identification: these are essentially the same as in the linear model. Students are assumed able to predict delay conditional on private information, and to take its impact on future wages into account. The nonlinear results fully confirm the linear results. In other words, we cannot reject the theory: the negative impact of delay is compatible with correct anticipation of this impact by students and therefore with a full-fledged signaling equilibrium. Students choose education but delay acts as an additional, not fully controled signal of their ability.

To estimate the models, we use a very rich sample of 12,310 male students, based on a survey[1] of young workers conducted in France between 1992 and 1997. The survey provides a wealth of details on family background, educational achievement, type of certificates and degrees, and a month-by-month reconstruction of labor market experience during the first five years of career. The data permit one to distinguish the duration of schooling from effective certificates. This is relevant because a substantial fraction of students failed exams and (or) repeated grades. The total accumulated delay of a student can be computed, since we observe his school-leaving age and can compare it with the average school-leaving age of the students who passed the same exams.

The relevance of degrees (as opposed to years of education) has been discussed in the literature. A number of authors have identified "sheepskin effects"; see e.g. Hungerford and Solon (1987), Jaeger and Page (1996). Degree holders tend to obtain higher wages than the workers with the same number of years of education, but who failed to pass the final exams.

[1]The survey is called *Génération 92*, and produced by a French state-sponsored institution called CEREQ. See section 2 below for details.

At the same time, Kane and Rouse (1995) showed that, among those who failed to earn the degree, the number of credits (i.e., partial completion of a Two-Year College's degree, for instance) does matter. Time-to-degree and other forms of schooling delays are in fact important from the empirical point of view. For instance, Brunello and Winter-Ebner (2003) have analyzed the expected completion time of college students in 10 European countries; they show that the percentage of students completing their degree at least one year later than the required time ranges from 30% in Sweden and Italy to zero in the UK. The problem seems to be important in the US, at the undergraduate as well as graduate levels[2]. The recent work of Garibaldi, Giavazzi, Ichino and Rettore (2006) identifies the impact of tuition fees on the time-to-degree of students at the Bocconi University in Milan[3].

Grade repetitions in primary and secondary school are frequent in some countries (like for instance France, Portugal and Spain), and absent in some others (the UK, Scandinavian countries, and in general in countries enforcing automatic promotion policies; see Paul (1997)). In the US, only a small percentage of students are repeating in any given year, but 20% repeated at least one grade by age 15 (see, e.g. Brophy (2006)). Grade repetitions are also frequent in developing countries (see, e.g. Gomes-Neto and Hanushek (1994)). The data used below are generated by an educational system that has several sizeable sources of delay, because grade repetitions in primary, secondary and higher education are very common in France[4]. Our delay variable is the result of an addition of these sources of age variation.

The impact of delay can be identified only with the help of instruments. Indeed, OLS estimations of log-wage regressions yield a small positive coefficient on delay. Instrumentation is therefore crucial. Distance to the nearest college (at the time of junior high-school entry) is one of our instruments (although we can dispense with it and still obtain the results), and has been generated with the help of detailed geographical data of the French

[2]There is also a literature on the time taken to complete PhDs in various countries (see Booth and Satchell (1995), Ehrenberg and Mavros (1995), Van Ours and Ridder (2003)). See Garibaldi *et al.* (2006), and their references.

[3]They show that a 1000 euros increase in tuition, in the last year of the programs, would reduce the probability of late graduation by 6 percentage points (with respect to an average probability of 80%).

[4]In this country, 45% of grade 9 male students had already accumulated at least a year of delay in 2002.

National Geographic Institute[5]. College proximity instruments, measuring a form of exogenous variation of education costs, have been used by a number of contributions, including the pioneering work of Card (1995) (see also, e.g., Duflo (2001), Carneiro *et al.* (2003)). This instrument has been criticized because mobility in geographical space is likely to be endogenous. In the present research, the line of defense is that distance-to-college is measured at the time of entry in the French equivalent of grade 6, that is, years before the age at which students effectively decide to go to college, and thus predetermined.

Our core instruments are based on school openings, and we took inspiration from Currie and Moretti (2003). We used the complete listing of the addresses and dates of opening of all secondary high schools, vocational colleges, colleges and universities in France since the early fifties, using a file from the Ministry of Education[6]. We tried various possibilities, but the instruments exploiting vocational high-school supply variation happened to work better than others. First, we computed the stock of vocational high schools in the county of residence at the age of entry into grade 6 of the student, and we divided this stock by the population aged 15 to 19 in the same county, at the same moment. Second, we took differences of this county-level per-capita stock of schools between two exogenously (but appropriately) chosen years.

The consequences of incomplete information on labor markets have been explored in various ways (on the empirical assessment of adverse selection problems, see, for instance, Gibbons and Katz (1991), Foster and Rosenzweig (1993)). Empirical tests of job-market signaling in the sense of Spence (1973) versus human-capital theories, inspired by Becker (1964) and Mincer (1974) are difficult to construct[7]. The basic reason for this difficulty is that the two theories predict the same increasing relationship between education and wages. A handful of contributions only have been devoted to this topic since the mid-seventies. Early attempts are, for instance, due to Wolpin (1977), Riley (1979) and Albrecht (1981). A decomposition of the productivity-enhancing and signaling effects of education can be obtained with the help of a structural model but only at the cost of strong restrictions (see

[5] *Institut Géographique National.*

[6] The *Base Centrale des Etablissements.*

[7] On the debate about job-market signaling, see for instance Weiss (1995), Riley (2001), Lange and Topel (2006).

Fang (2006)). Some identification strategies rely on shocks affecting the supply or demand of education (e.g. Lang and Kropp (1986), Bedard (2001)). Under the assumption of employer full-information, the opening of a new college in some remote county would reduce the costs of higher education and therefore only induce some of the students to study longer. In contrast, if we observe a signaling equilibrium, the same college opening would change the average productivity of high-school graduates, because the best students would choose to go to college. The wages of high-school graduates would then decrease to reflect the lower quality of the group of students choosing to stop after high school, and the incentives to finish high school would be reduced for the weakest students. As a consequence, the number of high-school dropouts would also increase. In essence, Bedard (2001) shows that her data support this prediction.

Closer to our approach are strategies testing for the presence of statistical discrimination in wages, based on the presence or absence of a certificate. For instance, Tyler, Murnane and Willett (2000) test if the GED (General Educational Development) credential has a signaling value on the US labour markets[8]. Tyler *et al.* (2000) assume that the total test score measures human capital. But the passing standards differ from one US state to another. This constitutes a natural experiment, because some high-school dropouts with equal GED test scores differ in GED status. Tyler *et al.* (2000) then show that the young workers who passed the GED earn significantly higher wages than workers who failed because of stricter standards but have the same underlying test score. This is a sure sign that signaling is taking place.

Our approach borrows elements from the recent literature on employer learning (e.g., Farber and Gibbons (1996), Altonji and Pierret (2001), Lange (2007)). According to employer-learning theory, the impact of job-market signaling effects is limited to the beginning of a worker's career, because employers learn the unobserved ability characteristics of employees after a few years only[9]. Thus, we have a good chance to find a signaling effect before it fades out, precisely because our data cover the first years of career of young workers.

[8]The GED is a battery of tests that a high-school dropout can take as a "second chance". On this question see also Cameron and Heckman (1993).

[9]See Lange (2007).

Following Farber and Gibbons (1996), we distinguish 4 kinds of variables: (i) variables observed by both the employer and the econometrician (education, age, various controls); (ii) variables observed neither by the econometrician nor by the employers (unobservable individual "ability" characteristics); (iii) variables observed by the employer but not by the econometrician (potentially some individual characteristics of the employee), and crucially, (iv) variables observed by the econometrician, but not by the employer (past environmental factors that matter for education and delay but are not reported in the employee's CV). In the work of Farber and Gibbons (1996), Altonji and Pierret (2001) and Lange (2007), the role of category-(iv) variables is played by the Armed Forces Qualification Test (AFQT): it is assumed that the AFQT measures unobserved ability and is not observed by employers. In the present research, we assume that employers do not observe our instruments, that is, variations in the local ratio of schools to school-age population that have an impact on delay.

The following story gives the intuition for our identification strategy. Assume that Adam and Bob are two equally able students with the same type of parents and the same degree, but that they differ in their school-leaving age. We observe that Adam earns more than Bob at the beginning of his career, in spite of the fact that they are equally gifted and productive. Adam was raised in a favorable environment and finished college sooner than Bob. Because he was born in a remote spot, Bob has faced unobserved difficulties during his school years that caused him to repeat a grade. In spite of these conditions, he fully caught up with Adam in terms of competence, but he unfortunately sends a less positive signal. In our model, if delay did not convey information and was not used by employers to statistically discriminate, then consistent estimates of the impact of delay on wages should be zero.

We now add a few comments on our structural econometric approach, and its relationship with the literature on human capital. The literature on returns to education has been entirely renewed in the 90s by the quest for instrumental variables, aimed at solving the problem posed by the endogeneity of education; see the surveys of Card (1999), Heckman *et al.* (2003). At the same time, a small number of structural econometric approaches have tested theories of individual schooling investments in models in which schooling decisions are derived from expected utility maximization, using dynamic programming; see, e.g., Keane and Wolpin (1997), Eckstein and Wolpin (1999), Taber (2001), Belzil and Hansen (2002), Magnac and Thesmar (2002), Attanasio *et al.* (2005). The increasing need to take individual heterogeneity of returns and information into account has led to contributions proposing a decomposition method for the cross-section variance of earnings. This variance is broken into a component that is predictable at the time students decide to go to college, and an unforecastable component, using a method of separation of individual heterogeneity from pure earnings-risk: see Carneiro *et al.* (2003b), and Cunha *et al.* (2005). The latter approaches are based on the identification of underlying, unobservable factor structures[10].

For our nonlinear econometric specification, we took inspiration from these contributions. Our model has a simple structure as compared to dynamic programming models à la Keane and Wolpin; it is essentially derived from static expected-utility maximization: as if the student would decide at the age of say, 13, his highest targeted degree, bearing the risk of random completion time (with correlative random costs) and random earnings. This simplification has two sorts of advantages. First, it allows for the use of simple estimation methods: straightforward Maximum Likelihood, without any additional difficulty due to nested algorithms. Second, the model has a closely comparable "reduced-form" counterpart, which is an Ordered-Probit, latent-index model à la Heckman, and its sources of identification are very clear. In spite of its relative simplicity, our model captures a dynamic element, which is the fact that individual schooling decisions are made on the basis of wage expectations, conditional on some background characteristics observed by the student, but unobserved by the econometrician.

In the following, Section 2 describes the data. Section 3 is devoted to the linear model, estimation results and various robustness checks. Section 4 presents the nonlinear

[10]On this methodology, see also Bonhomme and Robin (2006). These authors use two different measures of education to identify a factor structure in the residuals of a wage equation. The first measure is school-leaving age — call it 'age', for short. The second is a coding of the highest diploma obtained by the individual in 16 categories: this latter variable taking the median value of school-leaving age in the sample, in each diploma category — call it 'diploma', for short. These two variables are close to our (education level, delay) pair, although different in principle. Bonhomme and Robin (2006) identify two factors, explaining 'age', 'diploma' and wages simultaneously. They conclude that the 'true education' variable would be a certain combination of the two factors, and that 'diploma' and 'age' measure 'true education' with error.

model and its estimation results.

2 Data

To perform the estimations presented below we used "Génération 92", a large scale survey conducted in France. The survey and associated data base have been produced by CEREQ (*Centre d'Etudes et de Recherches sur les Qualifications*), a public research agency, working under the aegis of the Ministry of Education[11]. Génération 92 is a sample of 26, 359 young workers of both sexes, whose education levels range from the lowest (i.e., high-school dropouts) to graduate studies, and who graduated in a wide array of sectors and disciplines. Observed individuals have left the educational system between January 1rst and December[12] 31rst, 1992. They have left the educational system for the first time, and for at least one year in 1992[13]. The labor market experience of these individuals has been observed during 5 years, until 1997. The survey provides detailed observations of individual employment and unemployment spells, of wages and occupation types, as well as geographical locations of the students at the age of entry into junior high-school (roughly 11), and in 1992, when they left school. The personal labor-market history of each survey respondent has been literally reconstructed, month after month, during the period 1993-1997, by means of an interview. Before 1992, the individual's educational achievement is also observed.

For the purpose of estimation, we have created several variables with the help of the data. More precisely, we computed four endogenous variables: (i) an *earnings* statistic, (ii) a probability (or rate) of *employment*, (iii) *education*, and (iv) *delay*. We also studied variants of the earnings and employment variables. We first describe the education variable, then the wages and employment variables, and finally delay.

Education levels, representing degrees, are indicator variables. But to explore the impact of degrees in a linear model, we constructed a synthetic schooling variable, dubbed

[11] Articles and descriptive statistics, concerning various aspects of the survey, are available at www.cereq.fr.

[12] To fix ideas, the number of inhabitants of France who left school for the first time in 1992 is estimated to be of the order of 640,000.

[13] They did not return to school for more than one year after 1992, and they had not left school before 1992 except for compulsory military service, illness, or pregnancy.

education. By definition, it is the individual's "normal age" after a number of years of successfully concluded education. The "normal" number of years needed to reach the individual's grade, sit the exam and earn the degree, is a conventional age, associated with each individual's school-leaving degree. For each degree or certificate, the normal age is thus the age of those who earned this degree or certificate, without any grade repetition or delay of any kind — not the average completion age. Our education variable is thus a particular construction that, albeit natural, is different from the traditional years-of-schooling used in the literature. A number of conventions have been used: (i) the high-school dropouts have a normal age of 13 years; (ii) the vocational high-school degree holders have a normal age of 16 or 18 years, depending on the category of their certificate[14]; (iii) those who passed the national high-school diploma, i.e., the *baccalauréat*[15], have a normal age of 18 years; (iv) two years of college[16] correspond to a normal age of 20 years, and so on. In the linear model studied in Section 3 below, this education variable is used instead of the years-of-schooling measure of human capital. In the nonlinear model, the education variable is not used anymore, because education levels are indicated by dummy variables.

Each individual's curriculum on the job market is an array of data including a number of jobs, with their corresponding wages and durations in months, and unemployment spells, again with a length in months. To estimate the returns to education, we rely on a single, scalar index of earnings for each worker. We constructed four different wage variables with the help of the data. The first statistic is simply the arithmetic average of the full-time wages earned during full-time employment spells, weighted by their respective spell durations. In the following, this index is called the *mean wage*. We also consider a second and a third statistic: the wage earned by the individual in his first full-time job, called the *first full-time wage*, and the *last wage*, earned in the last full-time job observed.

[14] i.e., the so-called Certificats d'Aptitude and Brevet d'Etudes Professionnelles.

[15] Grade 12 students in the US correspond (roughly) to the French *classe terminale*, and the students of this grade sit an examination called *baccalauréat*. There exist vocational versions of the diploma.

[16] The corresponding exam is called DEUG (*Diplôme d'Etudes Universitaires Générales*), which is the equivalent of an Associate's degree, or DUT (*Diplôme Universitaire de Technologie*). There are exams at the end of each of the college years in French universities, and the DEUG or DUT correspond to the end of grade 14.

We can also compute a rate of employment. Each young worker is observed during 5 years, but depending on the exact month during which he or she left school, the number of observed months can vary a little. It varies from 60 to 72 months, to be precise. Our *employment* variable is the logarithm of the ratio of the number of months spent in employment over the total number of months in the observation period. For descriptive statistics and further details, see Appendix A.

A substantial part of the variance of school-leaving age, conditional on education level or degrees (again, see Appendix A), happens to be due to *repeated grades*. Grade repeaters are quite common, even in college[17]. Delays are thus generated by grade repetitions in primary, secondary and higher education. They are also computed with the help of some conventions. An individual i's *delay* δ_i is defined as this individual's school-leaving age d_i, minus the average school-leaving age τ of those for which this degree is the highest (and who thus left school with that degree)[18]. This particular definition has a nice property, proved in the next section: education and delay in our sense are by construction orthogonal. The efficiency of grade repetition in primary and secondary education is of course a hotly debated issue, but until today, the institution has survived. For instance, an individual who finished high school and passed the national examinations (i.e., the *baccalauréat*) at the age of 19.33 is below par and would get a delay of approximately $-1.45 = 19.33 - 20.78$ years, because the average age of those who left school at this level is 20.78[19]. The national high-school diploma is required for admission to colleges (i.e., *Universités*) in France. Thus, a person who passed the *baccalauréat* at the age of 18.5 and spent two years in college but failed to pass an Associate's or any equivalent degree has an education level of only 18 (which corresponds to that person's highest degree) and would have a delay of -0.28 years (since the average

[17]Freshmen repeating the first and second years of college are quite common.

[18]We also studied a variant, in which delay was defined as school-leaving age minus normal age (i.e.,"education"). The differences between the two approaches are small, but the chosen definition seemed to yield better results.

[19]School-leaving age can in fact be measured in months, and then converted back into years. We observe that the average age at which those who went to college passed the national high-school exam is of course lower than 20.78, but the national high school exam is not their highest degree. In a number of regressions, we used delay measured in years, and the results are very similar.

age of those who left school with the *baccalauréat* is 20.78). Employers do observe the school-leaving age and compare it to the average school-leaving age of similar students: our central hypothesis is of course that delay conveys information about ability. Figures 1a-1c provide various representations of the distribution of delay for males. Fig. 1a is the plain non-parametric estimate of the density. Fig. 1b plots the densities, conditional on father education; Fig. 1c shows the density of delay, conditional on the student's education. Some stochastic dominance is visible on Fig 1c, but the overall impression is that the distribution of delay is quite stable and doesn't depend on education.

On top of this, the survey provides information on family background: the father's and the mother's occupation in 92, the father's and the mother's education are the most important of these variables. We also know if the parents are unemployed, inactive, retired or deceased (in 92). Are also observed, for each individual: the number of sisters, the number of brothers, the rank among siblings (i.e., birth order)[20]. We know the geographical location of the student's family at the age of junior high-school entry (roughly at 11) and the student's location at school-leaving age (i.e. in 1992). Location is rather precise since we know the *code* of each *commune*, and there are more than $35,000$ *communes* in France.

Part of our instruments and some additional controls are based on data with a geographical structure. Using a file from the National Geographical Institute (i.e. *Institut Geographique National*), which permits one to link fine territorial-division codes with geographical coordinates on the map of France, we have computed a *distance-to-college* instrument, which is the Euclidean distance between the jurisdiction (i.e., the *commune*) of residence at the age of entry into grade 6 and the *nearest* college[21] (i.e., the nearest *Université*). Note that distance to college is not computed with the student's location in 92, but with the coordinates of his residence at a much younger age. This data source also yields a measure of local population density, which we used as an additional control.

A number of other instruments that we constructed are based on inter-county variation, where by county we mean the French *département*[22]. With the help of Census and

[20] But we don't know the order of sexes in an individual's sibship.

[21] This distance is the Euclidean distance between the two points on the map of France, in kilometers.

[22] There are 95 *départements* in France. Communes are a much finer territorial division. So the distance-to-college variable is close to being individual-specific.

regional macro data from the National Statistical Institute (i.e., *INSEE*), we constructed a county-level share of population aged 15-19 in 1982, that is used as a control. We constructed an average county-level unemployment rate in in the years 1982-1987, in the county where the student was residing at junior high-school entry. This variable can possibly be used as an instrument if we control for the rates of unemployment experienced by the individual after 1992. So, we use the average rate of unemployment in the years 1992-1997, in the 1992 county of residence of the student, as a control in the wage and employment equations. This is tantamount to exploiting inter-county variability of past unemployment rates.

Finally, we constructed a battery of school-opening instruments, using a file from the Ministry of Education (the *Base Centrale des Etablissements*) which lists all high-school and two-year college openings in the country since 1950. The file enables one to distinguish between vocational and general high-schools. The instruments based on vocational high-school openings in each county happened to be the strongest. The 1980s in France witnessed a rapid growth in the number of vocational high-schools (i.e., to be precise, the *lycées professionnels* and *lycées techniques*). Figure 2 shows the historical development of the national stock of such schools, and displays a *per capita* version of this measure, namely, the stock divided by the number of 15-year-olds. Both curves are strongly increasing and correlated in the 70s and 80s. Interestingly, the data exhibit a substantial degree of inter-county variability of the stock of vocational high-schools, *per capita* of 15-to-19-year-olds. The density of this county-level *per capita* measure in the year 1982 is depicted on Figure 3. We use this variable as an intrument. In a recent paper, Currie and Moretti (2003), have used the same kind of school-opening per capita, measured in the years when the individual was at a crucial age, say 17 or 18. Here, given the structure of our data, we must avoid a potential problem of negative correlation of the individual's education with the high-school stock. This correlation would simply reflect the fact that educated students are older at the end of their studies and therefore experienced an environment with less high-schools during their teens. To avoid this problem, we have chosen to fix the year at which the stock is evaluated. The choice of 1982 as a fixed point in time, ten years before the school-leaving year of students, characterizes the school-supply environments, roughly around the age of junior high-school entry.

Now, one might argue that it is not the stock of high schools itself that plays a role, but its growth rate or first difference (we will discuss this point again below). We then also computed the variation of our county-level stock of vocational schools between two fixed points in time, namely between 1989 and 1982, and use the variation as an instrument. These years cover the relevant time span during which most of our students were teenagers. Again, with this definition, the years at which temporal variations are evaluated do not depend on the individual's age, but only on the individual's county of residence at the age of junior high-school entry. Now, is there enough inter-county variation in these stock variations in the sample? The answer is yes, as suggested by Figure 4, which gives a plot of the empirical density of stock variations. The distribution is skewed (there are more increases than decreases in the stock of high-schools per capita), but there is a non-negligible number of counties in which the stock has been reduced by closings. We postpone any further discussion of the validity of instruments to the next section.

3 The Linear Model and Estimation Results

We first studied the data with the help of standard econometric techniques. To this end, we used 3SLS on a system of four linear equations, involving several instruments. Our main empirical result is that delay and education, as defined above, have identifiable and opposite effects on labour market outcomes. We show here how this estimation strategy can be interpreted as a test of the presence of signaling in the sense of Spence (1973).

3.1 Derivation of the Model

Let s_i be individual i's education level and let d_i be i's school-leaving age. Let in addition $\tau(s_i)$ denote the average school-leaving age of the subset of individuals with the same education level s as i, that is, with $s = s_i$. Function $\tau(s)$ is the empirical counterpart of $E(d \mid s)$. Individual i's delay is defined as $\delta_i = d_i - \tau(s_i)$. Delay is thus the part of school-leaving age that is not predicted by education s. Alternative definitions are possible, but would lead to very similar econometric formulations (see our remarks below and the appendix). We now drop the index i to simplify notation.

We assume that an individual's productivity, denoted q, is given by the relation

$$\ln(q) = a_0 s + b_0 \delta + X c_0 + \theta_1 + \theta_2, \tag{1}$$

where X is a vector of covariates observed by both the employers and the econometrician, θ_1 is an ability factor observed by the employer, but not by the econometrician, and θ_2 another ability factor, observed neither by the employer nor by the econometrician. Both are assumed to have a zero mean, finite variances and a non-negative covariance. The productivity effect of education is denoted a_0; it is in general nonnegative. The direct productivity effect of delay is denoted b_0. It is not clear that this second effect is nonzero, but it is presumably positive, because it is a measure of experience, a measure of the student's maturity. We will return to this point later. Parameters c_0 are the coefficients on control variables. In addition to this, employers are assumed to observe both education s and delay δ, and these variables are determined by the auxiliary equations:

$$\begin{aligned} s &= X c_1 + Z g_1 + f_1 \theta_0 + \xi_1, \\ \delta &= X c_2 + Z g_2 + f_2 \theta_0 + \xi_2. \end{aligned} \tag{2}$$

In these expressions, we drop index i for simplicity; Z is a vector of at least two variables excluded from the productivity equation; θ_0 is a "talent" factor that is not observed, neither by employers nor by econometricians, and potentially correlated with ability factors θ_1 and θ_2. In addition, ξ_1 and ξ_2 are random error terms, assumed independent of Z and the θs, and c_1 , c_2, g_1 , g_2, f_1 , f_2 are parameters. Technically, variables Z are exogenous sources of variation of s and δ, assumed uncorrelated with of the θs. To be precise, using vector notation, we assume

$$p\lim(N^{-1}Z'\theta_1) = p\lim(N^{-1}Z'\theta_2) = 0, \tag{3}$$

i.e., these variables are instruments for education and delay. We assume that vector Z is not included in the employer's information set. The central argument here is that Z is a list of environmental factors that had an impact during the individual's childhood but are typically not reported in a CV — something that is not written on the candidate's face. Even if variables Z are in fact observed by the employer, there are some relatively mild assumptions under which they can still be excluded from the wage equation, as shown below.

Standard economic theory suggests that wages w are equal to the expected productivity of employees conditional on employers' information, that is,

$$w = E[q \mid X, s, \delta, \theta_1].$$

Now, assuming that our random factors and random error terms are normally distributed, $\ln(q)$ conditional on (X, s, δ, θ_1) is also normal. Using the well-known formulae for the expectation of a log-normal random variable, we thus have,

$$w = \exp\{E(\ln(q) \mid X, s, \delta, \theta_1) + (1/2)Var(\ln(q) \mid X, s, \delta, \theta_1)\}.$$

A special property of normal vectors is that the conditional variance $Var(\ln(q) \mid X, s, \delta, \theta_1)$ doesn't depend on the value of (X, s, δ, θ_1); it can therefore be treated as a constant, included in X. Thus, we get the equation,

$$\ln(w) = a_0 s + b_0 \delta + X c_0 + \theta_1 + E(\theta_2 \mid X, s, \delta, \theta_1). \tag{4}$$

Under the assumed normality of the variables, conditional expectations are linear, and we get the convenient formula,

$$E(\theta_2 \mid X, s, \delta, \theta_1) = a_3 s + b_3 \delta + X c_3 + f_3 \theta_1, \tag{5}$$

where a_3, b_3, c_3 and f_3 are theoretical regression coefficients. Substituting this result in the log-wage equation, we finally get the model,

$$\ln(w) = (a_0 + a_3)s + (b_0 + b_3)\delta + X(c_0 + c_3) + (1 + f_3)\theta_1. \tag{6}$$

We obtain a two-dimensional signaling model à la Spence. Delay now appears as a variable in the log-wage equation also because it is a signal which conveys information about the hidden talent factor θ_2. The same is true with the education variable's coefficient, which is the sum of two effects: the direct productivity-increasing effect of education a_0 plus Spence's signaling effect a_3. It is not possible to identify a_0 and a_3 (or b_0 and b_3) separately without making strong additional assumptions, but we can easily test if b_3 is significantly negative, if we can find a consistent estimator of $b = b_0 + b_3$, and if we are ready to assume that the pure productivity effect of delay b_0 is nonnegative.

Assume now that Spence's hypothesis, as reflected by the model stated above is wrong, and assume on the contrary that a full-information version of the theory holds. For convenience, we call "Becker's hypothesis" the model in which ability factors θ_1 and θ_2 are fully observable by the employers (but not by the econometrician). Under Becker's hypothesis, we have

$$\ln(w) = a_0 s + b_0 \delta + X c_0 + \theta_1 + \theta_2, \tag{7}$$

and s and δ are given by the same linear equations as above. If we estimate the regression $\ln(w) = as + b\delta + Xc + \nu$, where ν is a random error term, by ordinary least squares, the estimations of a and b are potentially biased under Becker's as well as under Spence's hypotheses. Education s and delay δ are clearly endogenous under either assumption. If we can find at least two valid instruments, under Spence's hypothesis, the IV estimates $(\widehat{a}_{IV}, \widehat{b}_{IV})$ are consistent and we find,

$$p\lim(\widehat{a}_{IV}) = a_0 + a_3, \qquad p\lim(\widehat{b}_{IV}) = b_0 + b_3.$$

In contrast, under "Becker's hypothesis", the same IV estimates are consistent but satisfy $p\lim(\widehat{a}_{IV}, \widehat{b}_{IV}) = (a_0, b_0)$. It follows that if Becker's hypothesis holds, we strongly expect to find $\widehat{b}_{IV} \geq 0$ and if in contrast, this coefficient happens to be significant and negative, then, Spence's hypothesis holds. This is because the result can only be due to the fact that $b_3 < 0$ under the prior assumption that $b_0 \geq 0$: delay signals negative unobservable productivity characteristics to the employer. This is indeed what we will find. We thus have a test of the presence of signaling in the sense of Spence, conditional on the assumption that b_0 is nonnegative. The most plausible interpretation of our results is therefore that a form of Spence's signaling hypothesis is present in the data. Our estimates would be compatible with a full-information view only if delay destroyed productivity (because then b_0 would be significant and negative). This could for instance be due to the fact that higher delay in fact measures a form of lower "quality" of the diplomas and certificates. But this interpretation contradicts the full-information assumption according to which employers observe the productivity-relevant aspects of quality directly: under "Becker's view," employers do not need to use delay as a proxy for something that they are supposed to observe directly. The only consistent interpretation of the presence of the delay variable in the log-wage equation, under employer full-information, is therefore to view it as a measure of experience, maturity or a measure of partial degree completion. Students that are slightly older than the average of the group holding the same degree could only be (slightly) more productive *ceteris paribus*, given that negative characteristics have been taken into account (i.e., controled for) by the employer.

It is possible to dispense with the assumption that Z is not observed by employers. We can admit instead that employers observe Z, provided that (i), Z has no direct impact on productivity q, and (ii), if Z doesn't contribute to the explanation of θ_2, given X, s, δ, θ_1; in other words, if the regression function $E(\theta_2 \mid X, s, \delta, \theta_1, Z)$ doesn't depend on Z. Finally, if delay itself is not observed by the employers, then, it is easy to check that we must have $p\lim(\widehat{b}_{IV}) = 0$.

Our data set contains detailed information on the length of employment and unemployment spells during the first five years of career of a young worker, enabling us to construct a second outcome variable: the probability of employment, denoted π. This variable is modelled in a way which is analogous to wages. We have in mind that the probability of employment depends on factors X, s, δ, and θ, for the same reasons as above. This is because students exhibiting higher delay and smaller abilities are more likely to be turned down by employers during the job search process. Such students are also more likely to experience difficulties to find a stable job. A log-linear approximation is reasonable for this model and we can thus estimate a second regression equation to explain individual rates of employment. We thus get the following system of four simultaneous equations,

$$\begin{aligned} \ln(w_i) &= as_i + b\delta_i + X_i c + \nu_i, \\ \ln(\pi_i) &= \alpha s_i + \beta\delta_i + X_i\gamma + \zeta_i, \\ s_i &= X_i c_1 + Z_i g_1 + \varepsilon_i, \\ \delta_i &= X_i c_2 + Z_i g_2 + \eta_i. \end{aligned} \tag{8}$$

Formally, (a, b, c), (α, β, γ), (c_1, c_2, g_1, g_2) are parameters to be estimated, X is a vector of controls, Z is a vector of instruments for s and δ, and $(\nu_i, \zeta_i, \eta_i, \epsilon_i)$ is a vector of random disturbances, with covariance matrix Ω. We keep in mind that the random terms $(\nu_i, \zeta_i, \eta_i, \epsilon_i)$ are correlated because they are functions of the same unobservable ability terms θ_i. A first crucial test is to check whether the estimated coefficients $(\widehat{a}_{IV}, \widehat{b}_{IV})$ and $(\widehat{\alpha}_{IV}, \widehat{\beta}_{IV})$ are significantly different from zero. A second crucial test is to check whether $\widehat{b}_{IV}$ and $\widehat{\beta}_{IV}$ are negative, which, according to our theory, indicates that signaling is taking place. The model has a particular structure, because $(s, \ln(w), \ln(\pi))$ do not contribute to the explanation of δ and $(\delta, \ln(w), \ln(\pi))$ do not contribute to the explanation of s, while finally $\ln(\pi)$ is excluded from the $\ln(w)$ equation, and vice-versa. It follows that to identify (a, b) and (α, β), we only need two exclusions from the wage and employment equations, in other words we need two instruments at least. We can at least use distance-to-college and school opening variables to do this job. In fact we will show that it is legitimate to exclude more than two variables from the wage and employment equations, and use an over-identified model. The model and its variants have been estimated by means of 3SLS (Three-stages least squares), which also yields an estimate of the covariance structure $\widehat{\Omega}$.

The education variable s and the delay variable δ have a remarkable property: their empirical covariance is zero. Let $\{1, ..., N\}$ be the set of observed individuals. For the ease of exposition, define the subsets of the agents with education level equal to s as $B(s) = \{i \mid s_i = s\}$. These subsets constitute a partition of the set of observed individuals. Let $N(s) = |B(s)|$ be the number of observations in $B(s)$. Then, by definition,

$$\tau(s_i) = \frac{1}{N(s_i)} \sum_{j \in B(s_i)} d_j, \tag{9}$$

and of course, $\tau(s_i) = \tau(s_k)$ for all k in $B(s_i)$. For a proof that the overall average delay is zero and that the empirical covariance of s and δ is zero by construction, see appendix B. Remark that, according to our definition, delay δ is uncorrelated with any deterministic function f of s, that is: $\widehat{cov}(f(s), \delta) = 0$ for any mapping f. In particular, $\widehat{cov}(\tau, d - \tau) = 0$. This is because the $\tau(s)$ are the coefficients of a regression of d on a set of indicators of s. Delay is the residual of this regression, and is therefore orthogonal to τ. Remark that $p \lim \tau(s) = E(d \mid s)$. It is also easy to check that $cov(\delta, s) = E[(d - E(d \mid s))s] = 0$.

A more general definition of delay also comes to mind. Delay could have been defined as the residual of the theoretical regression of d on all the variables observed by the employer, i.e., $\delta^* = d - E(d \mid X, s, \theta_1)$. If δ^* is the real delay variable, given that the econometrician doesn't observe θ_1, it can be replaced with the above defined δ. It then follows that we estimate a model with a built-in error in the delay variable $\delta^* = \delta + U$, where $U = E(d \mid s) - E(d \mid X, s, \theta_1)$. We show in Appendix C that this formulation leads to a very similar model and an equivalent econometric specification.

Another possibility is to define delay as school-leaving age minus "normal age", that is, $\delta^{**} = d - s$. This model is also explored in Appendix C and yields very similar results.

3.2 First Stage; Validity of Instruments

A glance at the first stage, that is, the education and delay equations estimated by OLS, shows the strength and impacts of the chosen instruments. Table 1 gives excerpts of the complete results for a benchmark linear specification with four equations, given in Appendix I. Education and delay have a parallel specification. The controls are: father and mother occupation dummies, falling in 7 categories; mother and father education dummies, in 5 categories; the population aged 15 to 19 in the county of residence at grade 6 entry (but measured in 1982), the population density in 1982 in the town of residence at grade 6 entry. The instruments are: distance to college at grade 6 entry; distance to college squared; the 1982 per capita stock of vocational high schools in the county of residence at grade 6 entry; the variation (denoted $\Delta Stock$)) of this variable between 1982 and 1989 in the county of residence at grade 6 entry; residence in the Paris region at grade 6 entry; and finally the average unemployment rate 1982-1987 in the county of residence at grade 6 entry. Table 1 shows that, except for distance to college squared in the education equation (first two columns of the table), all instruments are highly significant (columns 2 and 4 of Table 1 give the t-statistics of the corresponding coefficients for the education and delay equations respectively). The F-test for the joint significance of the instruments is far above 10 in both equations. The instruments are not weak. There is a lot of noise in the delay variable but the instruments do explain some of its variance.

As expected, distance-to-college has a negative impact on education; it also happens to have a negative impact on delay. It is understandable that those who study less because of the costs captured by distance-to-college also tend to accumulate less delay, probably because studying less means making less efforts to pass an exam (and therefore less repetitions as well as shorter studies). For the same reasons, the *Stock* and $\Delta Stock$ of vocational schools

TABLE 1: First Stage of the Benchmark Linear Model

	Education		Delay	
	Coeff	t	Coeff	t
Constant	14.95	(74.44)	-1.114	(-10.20)
Distance to college	-0.0067	(-2.17)	-0.0065	(-4.09)
(Distance to college)^2	0.0001	(1.39)	0.0001	(4.19)
Stock of vocational high schools 1982	0.0171	(14.19)	0.0048	(7.30)
ΔStock of vocational schools 1982-89	0.0317	(9.77)	0.0074	(4.38)
Local unemployment rate before 87	-0.0525	(-3.66)	0.0545	(6.89)
Paris area at grade 6 entry	0.5368	(5.47)	0.2898	(5.45)
F-test of instruments	31.2		18.06	
R^2	0.21		0.021	
Number of observations	12,310			

Both columns were estimated by ordinary least squares. t-statistics are in parentheses. Coefficients on controls are not reported.

clearly increase both education and delay. A higher stock means more opportunities to study and smaller costs of education, so the signs of the corresponding coefficients are easily interpretable. Residence in the Paris area at grade 6 entry has a large impact: again, longer studies go along with longer delays for Paris residents. Finally, it is interesting to note that local unemployment rates reduce education and increase delay. The impact on delay can be understood as the result of smaller opportunity costs of education in regions of high unemployment: there are less interesting job market opportunities and the incentives to earn a degree quickly are reduced. But the high unemployment regions are also regions in which young men tend to study less.

Appendix D gives further detail son the properties of the instruments; in particular, it shows the i:pact of instruments in sub-samples. To sum up, the *Stock* and $\Delta Stock$ instruments have a relatively balanced effect in sub-samples, while the distance-to-college instrument seems to affect mostly students from highly educated families.

Are these instruments truly exogenous? We will of course show that our instruments pass the tests of over-identifying restrictions very well, but some more fundamental arguments can also be put forward. The distance-to-college instruments introduced by Card (1995) have been criticized on the grounds that location in geographical space is endogenous, and that distance is correlated with regional variations of labor market conditions. The validity of an instrument must be assessed with respect to the quality of the controls introduced in the equations of interest (the wage and employment equations here). We are able to control for many family background variables, for instance, we know if the students are the sons of farmers, likely to live far away from the nearest university, or the sons of executives, likely to reside very close to a university, just because universities, as well as executives, are located in or near major cities. We also control for residence in the Paris region, both at the age of grade 6 entry, and the end of studies. We have added more geographical dummies as controls to purge wages and employment rates from regional variations, but this has not been very useful. The most important and significant geographical indicator is that of the Paris region, versus the rest of France. We also control for county population aged 15-19, local population density in the county of residence at grade 6 entry, and local unemployment rates post 1992, which can all contribute to the explanation of regional variations of wages and employment rates. In the present study, given the relatively rich set of controls, our line of defense is mainly the fact that this distance is measured at the age of grade 6 entry (not when students finish high school or during higher education of course). It follows that this distance is in essence pre-determined: the related instrument could be weak because student mobility would imply that the past distance-to-college is not relevant as a measure of cost, but it should not be endogenous. France is also characterized by a much smaller mobility in geographical space than, for instance, the US. The students who moved (from one county to another) between the age of grade 6 entry and 1992 are a minority: 14% of the sample to be precise. It is reasonable to assume that the location of an individual at age 10 or 11 is mainly determined by parental occupation and parental job opportunities. Note that in our data set, distance to the nearest college is not affected by variations of family location within the same city (i.e. the same *commune*). It follows that if location within a given city can be partially determined by (elementary and secondary) school choice, these variations are not reflected in the distance-to-college variable. To sum up, conditional on family background, our distance-to-college instrument reflects the existence of pre-determined variations of distance, measuring variations in the costs of education.

The arguments that one can put forward in favor of the per capita stock of vocational high schools in the county are similar. This stock is pre-determined as well, because it is measured in the county in which the student was residing at the age of grade six entry, but in a fixed year (i.e., 1982). So, if the fixed year is not well chosen in the past, or if the measure of education-supply is not appropriate, the instrument will be weak, but it will not be more correlated with unobserved determinants of wages for that reason, given the many controls. For instance, two blue-collar families that are otherwise similar may have faced different environments in terms of local school availability when their son was aged 10 or 11; this had an influence on the education and delay of their son, but we assume that, at the time, their location was mainly determined by the availability of jobs for the parents. We also assume that regional variations of the stock of vocational high schools in 1982 are not correlated with unobserved regional conditions of post-1992 labor markets, given that we control for some important sources of local variations of wages and employment rates.

The time variation of the per capita stock of vocational high schools between two adequately chosen points in time is likely to suffer from less problems, if any, than the per capita stock itself. This is because this instrument reflects inter-county differences in the local speed of development of the vocational school system which are likely to be due to variations in the policies and lobbying activities of local governments. The variability of $\Delta Stocks$ could of course be due to differences in regional growth rates but only to a limited extent, given the importance of central government funding in the education sector and the redistributive nature of the grant-in-aid system in France. This type of school-opening instrument has given good results in the recent work of Currie and Moretti (2003). One can of course think of unobserved determinants that would make both the per capita stock variable and its variation endogenous, and it is impossible to prove that an instrument is absolutely valid, so the quality of the statistical tests and results obtained will be our final line of defense.

TABLE 2: Benchmark Linear Model. Equations of Interest

	Mean Wage			
	OLS		3SLS	
	A	B	A	B
Delay	1.03%	0.94%	-9.27%	**-9.26%**
	(6.10)	(5.58)	(-3.50)	**(-3.29)**
Education	6.92%	6.51%	8.73%	**9.12%**
	(80.99)	(69.77)	(13.08)	**(10.49)**
R^2	0.3778	0.3927	0.0651	0.1654
p-value of overid. F-test	....	....	**0.0001**	**0.3984**
	Employment			
	OLS		3SLS	
	A	B	A	B
Delay	-1.78%	-1.61%	-15.79%	-19.29%
	(-3.31)	(-3.00)	(-2.00)	(-2.28)
Education	5.60%	5.55%	7.89%	12.98%
	(20.68)	(18.60)	(4.01)	(5.03)
R^2	0.0405	0.0470	0.0073	0.0197
p-value of overid. F-test	....	....	0.1500	0.2812
Number of observations	12,310			

Note: t-statistics are in parentheses. Model A is without any controls. Model B is the benchmark. Coefficients on controls are not reported. The controls are : Father's and mother's occupation; parental education; residence in Paris area in 1992; local unemployment rate 92-97; local population density in 1982; local share of population aged 15-19 in 1982. "Local" means "in the county of residence". Instruments (exclusions) are: Local stock of vocational high school per capita in 1982; variation of local stock of vocational high schools per capita 1989-1982; distance to nearest college at grade 6 entry; distance to nearest college squared; local unemployment rate 1982-87; residence in Paris area at grade 6 entry.

3.3 Benchmark Linear Model

We have estimated the linear four-equations specified above under a benchmark specification and a number of variants mainly defined by changes in the choice of instruments. We first present the benchmark specification results and later study the variants as robustness checks. The full benchmark results are given in Appendix I. In the benchmark specification, the controls are: father and mother occupation dummies, father and mother education dummies, population aged 15-19 (in the county of residence at grade 6 entry), local population density (in the town of residence at grade 6 entry), average county unemployment rates 1992-1997 (in the 1992 county of residence) and an indicator of residence in the Paris region in 1992. The instruments are: the distance to college at grade 6 entry, the distance to college squared, the per capita stock of vocational high schools (in the county of residence at grade 6 entry), the variation of the same per capita stock of schools from 1982 to 1989; the indicator of residence in the Paris region at grade 6 entry, the average over years 1982-1987 of the county unemployment rate. We will discuss below the possible role of each of these instruments, but we first focus on the impact of IV estimation with this benchmark model.

Table 2 gives the results, i.e., the estimated values of a, b, α, β, by means of OLS and 3SLS, in two different specifications. The t-statistics are given in parentheses, and the coefficients are expressed in percentage (i.e., multiplied by 100). The top of Table 2 gives the results for the wage equation; the bottom displays the equivalent results for the employment equation. Column A in each group of two reports the results of a very crude model, estimated without any family-background controls, while column B reports the results of the benchmark itself. If we consider the impact of delay on wages first, it is striking that the OLS estimates are small, positive and significant, while in stark contrast, 3SLS estimates are strongly negative and significant: a year of delay causes a -9% decrease of the mean wage, during the first five years of career. At the same time, the effect of education on wages is standard, OLS returns to a year of education are around 7%, while IV estimates of the returns are around 9%. Given that delay and education are orthogonal regressors, these results cannot be due to some form of multicollinearity. OLS estimates of the returns to education are slightly downward biased, while OLS estimates of the impact of delay are strongly upward biased. We will show later that these facts can be reconciled. Results show that a year of delay will approximately wipe out the benefits of an additional step on our scale of degrees. The differences between column A and column B are small, giving the impression that family-background controls do not play a major role, but in fact they do, because instruments are rejected when controls are not included. This is clear if we look at the p-value of the over-identification restrictions F-test. In column 3SLS-A, we reject the fact that the instruments are valid exclusions from the wage equation, while in contrast, in column 3SLS-B, the p-value being 39%, we really cannot reject the exclusions (provided that we accept at least two of them).

The employment equation has similar features. Table 2 shows that OLS estimates of the coefficient on delay are negative, but the bottom-right part of the table shows that these OLS estimates are also upward biased. A year of delay causes a 20% reduction in the probability of employment, during the first five years of career: this effect is very striking, it represents an additional year of search during the first five. Appendix E gives further results related to the linear benchmark. In particular it shows whatthe results would be if the traditional years of schooling variable had been used and delay was ignored. Appendix E discusses the OLS biases and provides an explanation for the sign of these biases.

3.4 Robustness Check I: Variants

We will continue our analysis with various forms of robustness checks. We first change the list of instruments, then change the outcomes, consider alternative definitions of delay and check for the impact of experience and mobility.

Table 3a shows the results of a number of variants. The top lines of Table 3a first give the results of the wage equation: coefficients a, b with t-statistics in parentheses, then the results of the employment equation: coefficients α, β and the related ts, the list of additional controls (added in both the wage and the employment equations) and the list of instruments, included in both the education and delay equations and excluded from the equations of interest. The first column restates the results of the benchmark. In variant 1, many controls have been added to the basic family background variables: family structure, which includes birth order, number of sisters, number of brothers, and the person's age at grade 6 entry. The inclusion of these additional controls is not changing the results much. Variant 2 is not using the Paris region indicator and past local unemployment rates as instruments (contemporary unemployment rates and the Paris region indicator are also removed from the list of controls). The wage equation results resist quite well, although it is clear that the removed instruments help improving the significance of the coefficient on delay. Yet, we still get the same negative sign and the same order of magnitude for the crucial coefficients. Variant 3 is just Variant 2 with the number of brothers and number of sisters used as additional instruments: this variant is a very good compromise, if one is ready to accept the validity of the number of siblings as an instrument for wages. In any case, these variants pass the test of over-identifying restrictions very well. Variant 4 is not making use of the per capita stock and $\Delta Stock$ of vocational high schools, Paris region and local unemployment variables as instruments: distance to college, number of siblings and the mother-at-home dummy are used as a source of exogenous variation. The results of Variant 4 are similar to those of Variant 3 and confirm the findings of the benchmark, while using a very different set of instruments (they have only the distance-to-college variables in common). Variant 5 is the benchmark with parental education as an additional set of exclusions: this variant excludes to many variables and very clearly fails to pass the test of overidentifying restrictions (but still provides a significant impact of delay on wages of -7%).

The study of variants is pursued in Table 3b. Variant 6 is the benchmark without the Paris region instruments and controls: the typical benchmark results are still obtained. Variant 7 is the benchmark without the local unemployment variables: removing these variables yields a very strong (but significant) impact of delay. Variant 8 is the benchmark without the distance-to-college instruments: results are again very similar to the benchmark. Variant 9 is the benchmark without the per capita stock instruments: the negative impact of delay is still present, with the same order of magnitude, but less significant. It seems that only the Stock and $\Delta Stock$ variables play an essential role in the significance of our results. If these variables are removed, as in Variants 4 and 9, they must be replaced with some other source of variation, like the number of siblings, to restore significance. Finally, Variant 10 is just identified, using $\Delta Stock$ and local unemployment only: the effect of delay is still estimated with precision in the wage equation, but not in the employment equation (although we still get the sign and order of magnitude). Remark that all the variants, with the exception of Variant 9, pass the overidentifying restrictions test very well.

We conclude from the inspection of variants that our results are quite robust and do not seem to depend on a particular instrument.

3.5 Robustness Checks II: Other outcomes; Impact of Experience

Another robustness check will be to test if the effect of delay is still significant and negative when (i), we change the definition of the dependent variable and (ii) if we control for the experience accumulated before the recorded school-leaving time. Instead of using the mean wage statistic, we can use the first full-time wage or the last wage observed in the 5-year observation period. Instead of using the employment variable, we can measure the duration of search, in months, to the first full-time job and regress it on education, delay and controls. Table 4 shows the results of these regressions and does the additional job of controlling for the effect of experience. The A columns of Table 4 report the 3SLS estimates of the benchmark model with just a change of dependent variable, as indicated. Experience accumulated during studies is measured by the total sum of months spent by the student in summer jobs and internships before labor market entry. In the B columns of Table 4, we just reestimate the benchmark with a control for experience added, treating the summer-jobs-and-internships variable as exogenous. Given that the benchmark is overidentified (we have more than 3 instruments), it is technically possible to treat the jobs-and-internships variable as endogenous in a linear model with 5 equations. Doing this yields the C columns of Table 4, in which experience is instrumented. We observe a remarkable stability of the crucial coefficients (a, b). We always find a significant and negative impact of delay on the first wage, the mean wages and the last wage, the effect being around -7%. Experience is significant and positive in the B columns, but the coefficients lose their significance once experience is instrumented (i.e., in C columns). Thus, the introduction of the summer-jobs variable in the model does not change the main results. Yet, there is a more subtle point: that the p-value of the overidentification F-test is low for the last wage, but high for the first wage. This indicates that some of the instruments work less well with the last wage statistic:

they are slightly correlated with the residual of the last-wage equation. This could be due to employer learning[23]: as time passes, more is learned about individuals and after only five years of career, some of the instruments Z start to explain wages. It is reassuring to see that the opposite phenomenon is true for the first full-time wage: in spite of being more noisy, it still has the negative impact of delay and at the same time, the highest p-values for the overidentifying restrictions test.

If we now look at the employment equation, it resists very well the introduction of experience. In constrast, the attempt at modelling duration of search along the same lines is a failure (the only thing that seems to come out of these last regressions is that summer jobs and internships reduce search duration).

We conclude from these tests that internship and summer-job experience, which has a positive value for employers, does not invalidate the negative impact of delay: experience can increase wages while delay, at the same time, still reduces wages.

3.6 Robustness Check III: Alternative Definitions of Delay

Table 5 presents another series of tests, based on alternative models obtained when we change the delay variable in the benchmark specification. Again, the 3SLS estimates of (a, b) and (α, β) are given with the t-statistics in parentheses, and each column is a variant of the benchmark. The benchmark itself is recalled as column 1. Our standard delay variable is measured in years, but given that the data give employment and unemployent spells in months, as well as the exact school-leaving month in the year, we can in fact compute a delay in months (and divide it by 12). Column 2 in Table 5 gives the results when delay in months is used instead of delay in years, and we see that the results ar essentially the same.

One could suspect that delay is in part search in disguise, so we change the definition of delay in a way that will test for the fact that delay is in fact partly a form of job search. The standard definition of delay is based on the following convention: for instance, a student who has spent just one year in college without passing any exam has an education level equal to the high-school diploma, and an additional year of delay. In other words, the standard

[23] On this question, see again Farber and Gibbons (1996), Altonji and Pierret (2001) and Lange (2007).

definition of delay is computed with the help of school-leaving age, while education is based on the highest earned degree. This is consistent with our view of the education variable as measuring the degrees, not the effective but potentially misspent years in high school or college. The "alternate 1" and "alternate 2" columns of Table 5 report the results of two variants based on the same idea. In alternate 1, both the delay and employment variables have been changed so that the months spent in school without producing any new diploma are added to job-search time, and therefore reduce the employment ratio. The alternate 2 specification is based on the same change of the delay variable, without changing the employment variable. These changes are not innocent, and could destroy the main result, but we see that the significant and negative impact of delay on wages is still present, although it vanishes in the employment equation with alternate 1.

We can produce yet another robustness test with the age at grade 6 entry. We know that a substantial fraction of the students has already accumulated delay while entering grade 6. This variable is a good predictor of both final delay and educational achievement. The last three columns of Table 5 show the results of three variants: the first uses age at grade 6 entry in years; the second uses delay at grade 6 entry (i.e., age minus average age at grade 6 entry in the group of students who reached the same *final* level of education); the third uses delay accumulated after grade 6 entry (standard delay is of course the sum of delay accumulated before grade 6 and after grade 6). If the impact of delay is less significant in the first two age-at-grade 6 columns, it is still negative. The last column of Table 5 shows that post-grade 6 delay yields approximately the same results as delay itself, with the benchmark specification.

3.7 Robustness Check IV: Impact of Mobility

The fact that mobility in geographical space during education years is likely to be endogenous could possibly induce some bias in our results. We use an indicator of residence in the Paris region at two different points in time: roughly at the beginning of high school and at the end of studies, and our local unemployment instrument is based on location at the beginning of grade 6, while we control for local unemployment at the end of studies. It is legitimate to control wage and employment equations for residence in the Paris region and local unemployment rates at the age of labor-market entry, but endogenous mobility could perturb the estimates. We can however hope that they are only slightly perturbed. Note that we estimated variants of the model in which these additional instruments were not used, and in which the impact of delay on wages was negative and significant. But we will submit our model to more difficult tests.

We first re-estimated the benchmark model with the subsample of immobile students: to be precise, the students who reside in the same county at the age of grade 6 entry and at the end of studies. The mobile students in this sense amount to 14% of the total number of observations only. The first column of Table 6 gives the result of this subsample estimation. The impact of delay is still present and negative, of the same order of magnitude (yet smaller), but less significant. It is likely that the loss of significance is due to the reduction in sample size. Then we constructed a dummy variable called *Move*, equal to 1 if the student's county of residence at the end of studies is not the same as his county of residence at grade 6 entry. This variable doesn't change if the student moved within the same county. We don't know if the move was motivated by education choices or by other causes, like changes in parental job location. In the third column of Table 6, *Move* is added as a control in the four equations, and we see that it has a positive and significant coefficient: those who moved earn 4.5% more on average. This is of course not very surprising. Yet, given the simple fact that the movers are mainly those who went to Universities, and given that we control for education, *Move* seems to indicate that movers are somewhat self-selected: they tend to be better than the average. The important point is that we observe that the coefficients on delay are stable and remain significant.

To measure mobility more finely, we computed a *Distance* variable, defined as the Euclidean distance, on the map of France, between towns (i.e., *communes*) of residence at grade 6 entry and at the end of studies. This is the distance covered by the individual during his education years, in kilometers, between two residences. Controling for *Distance* (Column 4 of Table 6) is not perturbing the results, and *Distance* is significant in the wage equation.

Finally, we estimated a 5-equations linear model, which is the benchmark, with an additional mobility equation. In the mobility equation, either *Move* or *Distance* are regressed on the benchmark controls and instruments, plus the education variable. Education is added in the mobility equation because covered distance is strongly correlated with higher education. Identification of this extended model is possible because the benchmark is overidentified. There is no specific instrument for mobility, but it is nevertheless interesting to look at possible changes in the estimated coefficients. The results of the 5-equations model are reported in the last 2 columns of Table 6: it is easy to check that the crucial coefficients do not change significantly, and do not lose significance (at least in the wage equation).

4 The Nonlinear Model

To sum up, it seems that our benchmark model is fairly robust: the main conclusions about the impact of delay on wages can be obtained with various subsets of intruments, with variants of the endogenous variables, and do not seem to depend on the introduction of additional controls for experience and mobility. Yet, there are some remaining questions. First, are results robust to changes of our conventional education scale? We propose to answer this question radically by using a set of endogenous dummy variables, indicating education levels, and an Ordered Probit structure, to model education choices. Thus, we get rid of the implicit constraints embodied in the education measure used until know. Second, to really conclude that our negative impact of delay is a proof of the presence of signaling, we should in principle try to show that the data is compatible with a full-fledged signaling equilibrium, that is, a situation in which the distribution of education is the result of students making rational risky education choices, based on private information and knowledge of the wage equation, including a correct appreciation of the negative signaling impact of delay. If not, we might have detected the presence of statistical discrimination on the part of the employers, but not necessarily signaling à la Spence *stricto sensu*. The nonlinear model studied here serves both purposes, and confirms both the results and the intuition: we cannot reject the fact that a signaling equilibrium generates the observed data. The nonlinear model thus proposes a theoretical rationalization for the linear model studied so far, and more precisely, it provides a positive answer to the following question: there exists a human-capital investment model, based on individual rationality (expected utility maximization) which underpins the observed joint distribution of wages, employment, degrees and delays.

4.1 Basic Assumptions

Individuals are indexed by $i = 1, ..., N$. Let $s = 0, 1, ..., S$, denote the certified schooling level, and let w denote the wage, as before. Let $x_i = \ln(w_i)$. Let s_i be the education level chosen by individual i. Let π_i denote the probability of employment, that is, the number of months in employment divided by the number of observed months, and let $y_i = \ln(\pi_i)$. Utility is logarithmic, and defined as follows,

$$u(w, \pi) = \ln(w\pi), \tag{10}$$

Let d_{is} denote the individual's age while leaving school at level s, and let $\tau(s_i)$ denote the average age of students finishing with a degree of level $s = s_i$. We assume that individuals form expectations about their future wages as a function of education level s, of delay $\delta_i = d_i - \tau(s_i)$, and of exogenous variables X_i, by means of an extended Mincer equation,

$$x_i = \ln(w_i) = \sum_{s=1}^{S} \chi_{is} a_s + b\delta_i + X_{i1}\gamma_0 + \nu_i, \tag{11}$$

where, $\chi_{is} = 1$ if $s = s_i$ and 0 otherwise, and ν_i is a Gaussian error term. To simplify notation, we drop index i and define,

$$x_s = \ln(w_s) = a_s + b(d_s - \tau(s)) + X_1\gamma_0 + \nu, \tag{12}$$

where d_s is the age at which the individual would pass degree s (and the real school-leaving age if s is the effective education level).

We assume that individual i predicts her (his) employment probability as the conditional expected value of π_i, using the model,

$$y_i = \ln(\pi_i) = \sum_{s=1}^{S} \chi_{is}\alpha_s + \beta\delta_i + X_{i1}\gamma_1 + \zeta_i, \tag{13}$$

where ζ_i is a Gaussian error term. To simplify notation, we again drop index i and denote,

$$y_s = \ln(\pi_s) = \alpha_s + X_1\gamma_1 + \beta(d_s - \tau(s)) + \zeta. \tag{14}$$

Then, define the individual's instantaneous utility as

$$u_s = \ln(\pi_s w_s) = x_s + y_s. \tag{15}$$

We use 6 different education levels.

Delay is assumed to be given (predicted) by the equation,

$$\delta_i = X_{i2}\gamma_2 + \eta_i. \tag{16}$$

where η_i is a Gaussian error term, and where X_{i2} is a vector of exogenous variables including instruments Z_i that are excluded from the wage and employment equations and γ_2 is a vector of parameters. Dropping index i, this equation can be rewritten,

$$d_s = \tau(s) + X_2\gamma_2 + \eta, \tag{17}$$

and the $\tau(s)$ are known parameters.

Finally, we specify the education costs of a year spent preparing for the exams of level s, as a fraction $1 - h_s$ of the expected wage $\pi_{s-1}w_{s-1}$, where $0 \leq h_s \leq 1$. The costs are thus a fraction of the wage that could have been earned if the individual did go to work with education level $s - 1$, instead of studying to reach level s. We thus assume that the opportunity and direct costs of education, incurred by an individual per period, are of the form $(1 - h_s)\pi_{s-1}w_{s-1}$. We adopt the following specification for h_s,

$$h_{si} = \exp(-X_{i3}\gamma_3 - c_s + \epsilon_i), \tag{18}$$

where X_{i3} is a vector of exogenous variables, related to environment and family background, including instruments Z_i that are excluded from the wage and employment equations; γ_3 and c_s are parameters, and ϵ is an error term with a normal distribution, interpreted as unobserved resources, or "help" from the family. In contrast, the error term $-\eta$ can be viewed as unobserved "talent" at school.

We have introduced 4 error terms, $(\nu, \zeta, \eta, \epsilon)$, respectively: "ability" at work, "ability" in job search, "handicap at school" (the opposite of talent at school), and unobserved "family help". This vector is assumed multivariate-normal with a zero mean and covariance matrix

$$\Omega = \begin{pmatrix} \sigma_\nu^2 & \sigma_{\nu\zeta} & \sigma_{\nu\eta} & \sigma_{\epsilon\nu} \\ \sigma_{\nu\zeta} & \sigma_\zeta^2 & \sigma_{\zeta\eta} & \sigma_{\epsilon\zeta} \\ \sigma_{\nu\eta} & \sigma_{\zeta\eta} & \sigma_\eta^2 & \sigma_{\epsilon\eta} \\ \sigma_{\epsilon\nu} & \sigma_{\epsilon\zeta} & \sigma_{\epsilon\eta} & 1 \end{pmatrix} \tag{19}$$

Note that $E(\epsilon^2) = 1$, for the sake of identification.

4.2 Expected Utility Maximization

Assuming that each individual lives for T periods (i.e., years) and has a zero rate of time preference (or a discount rate equal to one), we can express the individual's expected utility, conditional on ϵ, as follows,

$$\begin{aligned} V(s \mid \epsilon) &= E\left[\sum_{t=1+d_s}^{T} u_s + \sum_{z=1}^{s} \sum_{t=1+d_{z-1}}^{t=d_z} \ln(h_z \pi_{z-1} w_{z-1}) \mid \epsilon\right] \\ &= E\left[(T - d_s)u_s + \sum_{z=1}^{s}(\Delta d_z)(\ln(h_z) + u_{z-1}) \mid \epsilon\right], \end{aligned} \quad (20)$$

where $\Delta d_z = d_z - d_{z-1}$. Each individual is then assumed to choose level s so as to maximize V, the expected utility over the life-cycle, knowing the unobserved family factors ϵ, but bearing several kinds of risk, affecting employment, wages, and the costs of education (the duration of studies being random). Remark that $(T-d_s)+\sum_{z=1}^{s}(\Delta d_z) = T-d_0$. We assume that d_0 is exogenously given, for instance, $d_0 = 0$. Define then

$$\Delta V(s \mid \epsilon) = V(s \mid \epsilon) - V(s-1 \mid \epsilon). \quad (21)$$

We easily get,

$$\Delta V(s \mid \epsilon) = \Delta u_s[T - E(d_s)] - \Delta d_s[X_3\gamma_3 + c_s] + \epsilon[\Delta d_s - \sigma_{\epsilon\eta}\Delta u_s], \quad (22)$$

(see Appendix F for a derivation of this result).

4.3 Necessary Conditions for Optimal Choice of Education

We can now state the necessary conditions for an individually optimal choice of s as: $\Delta V(s \mid \epsilon) \geq 0$ *and* $\Delta V(s+1 \mid \epsilon) \leq 0$. Assume that $\Delta d_s - \sigma_{\epsilon\eta}\Delta u_s \geq 0$. This property holds if $\Delta u_s \geq 0$, and $\sigma_{\epsilon\eta} \leq 0$, which, given that ϵ represents help, and $-\eta$ represents unobservable academic talent, is a reasonable assumption. It is then easy to see that $\Delta V(s \mid \epsilon) \geq 0$ is equivalent to,

$$\epsilon \geq -\frac{\Delta u_s(T - E(d_s))}{(\Delta d_s - \sigma_{\epsilon\eta}\Delta u_s)} + \frac{\Delta d_s}{(\Delta d_s - \sigma_{\epsilon\eta}\Delta u_s)}(X_3\gamma_3 + c_s) \equiv k_s, \quad (23)$$

where $\Delta u_s = \Delta a_s + \Delta \alpha_s$, $\Delta d_s = \Delta \tau(s)$, and we define: $\Delta a_s = a_s - a_{s-1}$, $\Delta \alpha_s = \alpha_s - \alpha_{s-1}$. In addition, $E(d_s) = \tau(s) + X_2\gamma_2$. Family help ϵ must be greater than a threshold denoted k_s (the right-hand side of the above inequality).

It follows that s is an individually optimal level of education, knowing ϵ, only if

$$k_s \leq \epsilon \leq k_{s+1}. \tag{24}$$

Thus, education is determined by an ordered discrete choice (Ordered Probit) model with cuts k_s. Remark that if $\sigma_{\epsilon\eta} = 0$, the above necessary condition boils down to the following easily interpretable expression,

$$\frac{\Delta u_s(T - E(d_s))}{\Delta d_s} \geq X_3\gamma_3 + c_s - \epsilon, \tag{25}$$

i.e., marginal utility $\Delta u_s / \Delta d_s$, multiplied by the expected number of years at work, $T - E(d_s)$, must be greater or equal than marginal costs $X_3\gamma_3 + c_s$ minus family help ϵ.

The model has a meaning only if it is true that $k_s < k_{s+1}$, for all s, which is equivalent to say that second-order conditions hold. See Appendix G for a study of these second-order conditions: they are likely to hold. The parameters to be estimated are a_s, α_s, b, β, $\gamma_0, ..., \gamma_3$, c_s and Ω. The complete nonlinear model, called Model A, can be estimated by straightforward ML. Model A is fully identified. The likelihood is derived in Appendix H.

4.4 Variants of the Nonlinear Model: Model A and Model B

In the course of estimating the full model, we can also estimate a simplified version, hereafter called Model B, in which the particular functional form of the thresholds $k_1, ..., k_s$, as defined by (30), is not imposed. In this simplified version, the Ordered Probit part can be simply specified as

$$\kappa_s + X_3\gamma_3' \leq \epsilon \leq \kappa_{s+1} + X_3\gamma_3', \tag{26}$$

meaning that the individual chooses level s if and only if his (her) realization of ϵ falls in the above interval, with constant cuts denoted κ_s. The values of γ_3' and the cuts κ_s are not the same as γ_3 and c_s in Model A, and they do not play exactly the same role. Model B is a standard system of equations with endogenous dummy variables à la Heckman (see

Heckman (1978)). Model B can also be viewed as an extension of Cameron and Heckman's (1998) model, because it relies on the Ordered Probit to describe educational choices.

The advantage of our specification is now that we have two models, Model A and Model B, a "structural" and a "reduced-form" model, respectively, that are closely comparable. An immediate term-by-term comparison of Models A and B is possible for almost all parameters, except the cuts and γ_3. We will of course compare the likelihoods of the two models.

Model A embodies more structure, because it imposes a particular functional form of the cuts k_s. This form is not simply a nonlinear function of other structural parameters, because it involves individual variations of expected durations, through the terms $E(d_s) = \tau(s) + X_2\gamma_2$, and of education costs, through $X_3\gamma_3$, which all depend on observations. But, as suggested by the preliminary explorations of this model by means of standard linear methods, identification does not hinge upon functional forms: two instruments would in principle be enough. Intuitively, we need only one instrument for education, and one for delay, to identify the crucial parameters a and b. We use the same instruments and controls as in the linear benchmark model studied above. The instruments are excluded from X_1 but included in X_2 and X_3. Model B and the benchmark linear model have the same sources of identification: exclusions from the wage and employment equations. Model A uses the same sources of variability than Model B plus the variability of individual $k_{s,i}$ terms, which is itself generated by the variability of predicted durations.

4.5 Results

The estimation of Model A and Model B produces remarkably similar results, confirming the results obtained with the linear model. Table 8a shows the complete results for Model B with the sample of male students. Table 8a shows significant coefficients on delay in both the wage and employment equations; the order of magnitude and signs of these coefficients are roughly the same as the corresponding results for the benchmark linear model. In addition, we now get estimates of the returns to education in terms of wages and employment rates, for each education level s. More precisely, Table 8a lists the estimated Δa_s and $\Delta\alpha_s$. Given that each education level takes around two years, we find that an additional year of education yields approximately a 7% increase of the wage (and a much bigger increase of the probability of employment in the first years of career). Note in passing that the father's education increases the son's education *and* delay, while the mother's education very distinctly increases the son's education and *reduces* the son's delay.

The impact of excluded variables on education and delay are similar to the corresponding effects in the linear model. Distance-to-college now loses significance, but distance-to-college squared is significant. The *Stock* and $\Delta Stock$ variables still have a strong effect on education and delay: this is reassuring.

The estimated correlation matrix shows that the correlation coefficient of ϵ and η is small, but its sign implies that $\sigma_{\epsilon\eta}$ is negative. This is the expected sign, since ϵ represents an unobservable positive push from family background, while η is the negative of personal talent at school (or an unobservable personal handicap). The sign pattern of the correlation matrix Ω is the same as in the linear model.

Table 8b reports the results of a variant of Model B in which delay is replaced with interactions of delay with education levels. This is to check if the impact of delay varies with the degree. The first columns of the table give the coefficients of the interacted variables in the log-wage equation: they are all significant, negative and around −10%. The interaction of education and delay are also around −20% in the employment equation, less precisely estimated for the highest education levels. It seems that the impact of delay is stable and does not depend much on the degree.

To sum up, in essence, the results of Model B confirm those obtained with the linear model. In particular, the negative signaling effect of delay is not altered if we use dummies indicating education levels instead of the conventional years-of-education scale used in the linear model.

We now turn to estimations of Model A. The complete results are displayed on Table 9. Model B and Model A are very close, and exhibit the same qualitative and quantitative properties. The estimated parameters are numerically close, with the exception of some of the γ_3 coefficients and of the Ordered Probit cuts, as expected. Detailed comments of Table 9 are therefore not necessary.

Yet, can we say that Model A dominates Model B as a description of the data? We use Vuong's test of non-nested hypotheses to compare Model A and Model B rigorously. The test is based on the difference of the log-likelihoods, and its value is +1.11. The positive sign means that Model A has a slightly higher likelihood, but is not significantly better than Model B (this would have required a value of the test higher than 2, say; see Vuong (1989)). But at the same time, the test does not reject Model A, when compared with Model B (rejection would have required a negative value of Vuong's test, a value smaller than −2). This might seem surprising, but the data seem compatible with the description proposed by Model A, which provides a richer and more "structural" description of individual education decisions. So, we can conclude that our theory of signaling by means of delay is not rejected by the data. The structural analysis shows that the negative effects of delay do exist even if rationally anticipated by individuals, and that they are compatible with educational decisions made under risk by individuals, when the risk affects education costs as well as wages and employment.

5 Conclusion

Using an extension of the Mincer log-wage equation, we found a way of testing for the presence of signaling in the sense of Spence (1973). Log-wages have been regressed on two orthogonal variables: *education*, which is a level on a scale of degrees, and *delay*, computed as school-leaving age minus the average school-leaving age of the group with the same degree. Using instruments based on past school-openings and distance to colleges, we found that delay has a significant, robust, and negative impact on the wages of young workers. A year of delay causes a 9% decrease of wages, averaged over the first five years of career. At the same time, we found standard values of the returns to education with our degree-based education variable. IV estimation is crucial to obtain these results, because OLS estimates of the coefficient on delay are small and positive. We provided explanations for the likely source of this bias. A number of checks, based on (i) variants of the model, (ii) changes in the intruments, (iii) taking care of experience and mobility, (iv) making use of alternative definitions of delay, education and wages, and alternative outcomes, like the employment rate, showed that the estimated effect is robust. The negative effect of delay that we find is a job-market signaling phenomenon. Employers being incompletely informed, delay, as defined above, conveys information about the young workers' productivity-relevant characteristics. Human capital theory under employer full-information does not predict a negative impact of delay. It follows that if the full information assumption was (nearly) true, the IV estimates of the effect of delay on wages would be approximately zero. To check that our model is compatible with a complete signaling equilibrium, we constructed a nonlinear model, based on expected utility maximization, in which students choose education. In this model, students form rational predictions of future wages as a function of education, delay, and of information that the econometrician does not observe. Students also predict delay, which is a source of risk, conditional on private information. The nonlinear model is estimated by maximum likelihood. The results are compatible with the assumption that the data is generated by a signaling equilibrium in which students take the risk affecting time-to-degree and its impact on wages into account.

6 References

Albrecht, James (1981), "A Procedure for Testing the Signaling Hypothesis", *Journal of Public Economics*, **15**, 123-132.

Altonji, Joseph, G. and Charles Pierret (2001), "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 116, 313-350.

Attanasio, Orazio, Meghir, Costas, and Ana Santiago (2005), "Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA", manuscript, University College, London, *http://www.homepages.ucl.ac.uk/*

Becker, Gary S. (1964), *Human Capital: A Theoretical And Empirical Analysis with Special Reference to Education*, Third edition 1993, NBER and the University of Chicago Press, Chicago, Illinois.

Bedard, Kelly (2001), "Human Capital vs Signaling Models: University Access and High-School Drop-outs", *Journal of Political Economy*, **190**, 749-775.

Belzil, Christian, and Jürgen Hansen, (2002), "Unobserved Ability and the Returns to Schooling", *Econometrica*, **70**, 2075-2092.

Bonhomme, Stéphane, and Jean-Marc Robin (2006), "Using Higher-Order Moments to Estimate Linear Independent Factor Models", manuscript, Université Paris 1, Panthéon-Sorbonne, *http://eurequa.univ-paris1.fr/membres/robin.*

Booth, Alison L., and S. E. Satchell (1995), "The Hazards of Doing a PhD: An Analysis of Completion and Withdrawal Rates of British PhD Students in the 1980s", *Journal of the Royal Statistical Society, Series A*, **158**, 297-318.

Bound, John, and David A. Jaeger, (2000), "Do Compulsory School Attendance Laws Alone Explain the Association Between Quarter of Birth and Learning?", *Research in Labor Economics*, **19**, 83-108.

Brophy, Jere, (2006), *Grade Repetition*, Education Policy Series, International Institute for Educational Planning, International Academy of Education, UNESCO, Paris.

Brunello, Giorgio, and R. Winter-Ebmer (2003), "Why Do Students Expect to Stay Longer in College? Evidence from Europe", *Economics Letters*, **80**, 247-253.

Cameron, Stephen V., and James J. Heckman (1993), "The Non-equivalence of High-School Equivalents", *Journal of Labor Economics*, **11**, 1-47.

Cameron, Stephen V., and James J. Heckman (1998), "Life-Cycle Schooling and Dynamic Selection of Bias: Models and Evidence for Five Cohorts of American Males", *Journal of Political Economy*, **106**, 262-333.

Card, David (1999),"The Causal Effect of Education on Earnings", chapter 30 in: Ashenfelter, O. and D.Card, eds., *Handbook of Labor Economics, Volume 3*, Elsevier Science, Amsterdam.

Card, David (1995), "Using Geographical Variation in College Proximity to Estimate the Return to Schooling", in L. N. Christofides, E. K. Grant and R. Swidinsky, eds., *Aspects of Labor Market Behavior: Essays in Honor of John Vanderkamp*, University of Toronto Press, Toronto.

Carneiro, Pedro, Heckman, James J., and Edward Vytlacil (2003a), "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education", University of Chicago, manuscript.

Carneiro, Pedro, Hansen, Karsten,T. and James J. Heckman (2003b), "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice", *International Economic Review*, **44**, 361-422.

Cunha, Flavio, Heckman, James J., and Salvador Navarro (2005), "Separating Uncertainty from Heterogeneity in Life-Cycle Earnings", *Oxford Economic Papers*, **57**, 191-261.

Currie, Janet and Enrico Moretti (2003), "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings," *Quarterly Journal of Economics*, **118**,1495-1532.

Duflo, Esther (2001), "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment", *American Economic Review*, **91**, 795-813.

Eckstein, Zvi, and Kenneth I. Wolpin (1999), "Why Youths Drop out of High School? The Impact of Preferences, Opportunities, and Abilities", *Econometrica*, **67**, 1295-1340.

Ehrenberg, Ronald G., and Panagiotis G. Mavros (1995), "Do Doctoral Students' Financial Support Patterns Affect Their Times-to-Degree and Completion Probabilities?", *Journal of Human Resources*, **30**, 581-609.

Fang, Hanming (2006), "Disentangling the College Wage Premium: Estimating a Model with Endogenous Education Choices", *International Economic Review*, 47, 1151-1185.

Farber, Henry, S. and Robert Gibbons (1996), "Learning and Wage Dynamics", *Quarterly Journal of Economics*, 111, 1007-1047.

Foster, Andrew D. and Mark R. Rosenzweig (1993), "Information, Learning, and Wage Rates in Low-Income Rural Areas", *Journal of Human Resources*, **28**, 759-790.

Garibaldi, P., Giavazzi, F., Ichino, Andrea, and Enrico Rettore (2006), "College Cost and Time to Obtain a Degree: Evidence from Tuition Discontinuities", manuscript, *http://www.iue.it /Personal/Ichino/*.

Gibbons, Robert and Lawrence F. Katz (1991), "Layoffs and Lemons", *Journal of Labor Economics*, **9**, 351-380.

Gomes-Neto, Joao B., and Eric Hanushek (1994), "Causes and Consequences of Grade Repetitions: Evidence from Brazil", *Economic Development and Cultural Change*, 43,117-148.

Heckman, James J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica*, **46**, 931-959.

Heckman, James J., Lochner, Lance J., and Petra E. Todd (2003), "Fifty Years of Mincer Earnings Regressions", IZA DP n°775, Institute for the Study of Labor, *IZA*, Bonn.

Hungerford, T., and Gary Solon (1987), "Sheepskin Effects in the Returns to Education", *Review of Economics and Statistics*, **69**, 175-177.

Jaeger, David A., and Marianne E. Page (1996), "Degrees Matter: New Evidence on Sheepskin Effects in the Returns to Education", *Review of Economics and Statistics*, **78**, 733-740.

Kane, Thomas J., and Cecilia E. Rouse (1995), "Labor-Market Returns to Two- and Four-Year College", *American Economic Review*, **85**, 600-614.

Keane, Michael, and Kenneth I. Wolpin (1996), "Career Decisions of Young Men", *Journal of Political Economy*, **105**, 473-522.

Lang, Kevin, and David Kropp (1986), "Human Capital vs Sorting: The Effects of Compulsory Attendance Laws", *Quarterly Journal of Economics*, **101**, 609-624.

Lange, Fabian, and Robert Topel (2006), "The Social Value of Education and Human Capital", chapter 8 in: Eric Hanushek and Finnis Welch, eds, *Handbook of Economics of Education, Vol.1*, Elsevier, Amsterdam.

Lange, Fabian (2007), "The Speed of Employer Learning," *Journal of Labor Economics*, **25**, 1-35.

Magnac, Thierry, and David Thesmar (2002), "Identifying Dynamic Discrete Decision Processes", *Econometrica*, **70**, 801-816.

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

Paul, Jean-Jacques (1997), "Le redoublement à l'école: une maladie universelle?", *International Review of Education*, 43, 611-627.

Riley, John G. (1979), "Testing the Educational Screening Hypothesis", *Journal of Political Economy*, **87**, S227-252.

Riley, John G. (2001), "Silver Signals: Twenty-Five Years of Screening and Signaling", *Journal of Economic Literature*, **39**, 432-478.

Spence, Michael (1973), "Job Market Signalling", *Quarterly Journal of Economics*, **87**, 355-374.

Taber, Christopher R. (2001), "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?", *Review of Economic Studies*, **68**, 665-691.

Tyler, John H., Murnane, Richard, J., and John B. Willett (2000), "Estimating the Labor Market Signaling Value of the GED", *Quarterly Journal of Economics*, **115**, 431-468.

Van Ours, Jan C., and Gert Ridder (2003), "Fast Track or Failure: A Study of the Graduation and Dropout Rates of PhD Students in Economics", *Economics of Education Review*, **22**, 157-166.

Vella, Francis, and R.G. Gregory (1996),“Selection Bias and Human Capital Investment: Estimating the Rates of Return to Education for Young Males”, *Labour Economics*, **3**, 197-219.

Vuong, Quong H., (1989), “Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses”, *Econometrica*, **57**, 307-334.

Weiss, Andrew (1995), “Human Capital vs Signalling Explanations of Wages”, *Journal of Economic Perspectives*, **9**, 133-154.

Wolpin, Kenneth, I. (1977), “Education and Screening”, *American Economic Review*, **67**, 949-958.

7 Appendix

Appendix A is devoted to details on the sample and additional descriptive statistics; Appendix B proves that delay and education are orthogonal; Appendix C explores the models obtained with alternative definitions of delay; Appendix D gives additional details on the first-stage regressions; Appendix E discusses the OLS bias; Appendices W, X, Y are devoted to the nonlinear model; Appendix I gives the complete estimation results of the benchmark linear model, of Model A and Model B.

7.1 Appendix A. Data and Descriptive Statistics

Table A1 shows the empirical distribution of school-leaving age, conditional on the education level reached by male students (the displayed figures are frequencies). As can be seen, school leaving-age is substantially dispersed, even conditional on final education level[24]. Figure A1 gives the distribution of the education variable itself, for males and females[25]. Table A2 gives

[24]For instance, the first line of Table A1 says that 33 percent of the high-school dropouts left at the age of 18.

[25]The probabilities of 16 and 19 are zero because, due to our conventions, nobody leaves school with an education equal to 16 or 19. There is some bunching of post-graduation diplomas such as Master’s degree at level 23.

further indications on the distribution of the education variable, conditional on parental education levels. Table A2 shows some well-known facts; for instance, that a student's probability of reaching the highest degrees is much higher than for any other category when his (her) father went to college.

A difficulty with wages is that we do not observe the hours worked (but we know if the individual worked full-time or part-time). To solve this problem, we decided to select the individuals who experienced at least a full-time employment spell during the five-years observation period. More precisely, we first removed 717 individuals who had never worked (no employment spell recorded during 5 years). The remaining 25,642 individuals are the addition of 14,213 men and 11,429 women who worked at least once during the observation period. We then selected the individuals who experienced at least one full-time employment spell during the five years. As a consequence, we lost 11.7% of the male sub-sample, but still had 12,538 men. The final stage was to match the sample with geographical data from the National Geographical Institute, in order to compute the *distance-to-college* instruments, and other geography-related variables. Some observations of the individual's location at the age of entry into junior high-school (the jurisdiction of residence's code) were missing. This left us with only 12,310 males. The possible bias introduced by this selection procedure is limited in the case of men[26]. In the present article, we focus on the male subsample.

Yet, a clear advantage of our selection procedure is that it permits us to compare earnings more precisely, given that full-time employment means a 39 hours working week for most wage-earning employees (and given the heavily regulated French labor market of the 90s). More importantly, it tends to select a relatively homogeneous population of youths willing to work full-time (which has some advantages).

The mean wage variable ignores the length of unemployment spells, and the difficulties faced by the individual to find a stable (and well-paid) job. To capture the effect of job instability on average earnings, we defined a second average, simply called *earnings*. To compute this average, wages and unemployment benefits are weighted by the corresponding employment or unemployment spell duration[27]. Figure A2 presents a plot of the density of

[26]The same mode of selection would have left us with a sample of 8630 women, all willing to work full-time. It is therefore likely that there is a sizeable selection bias in our female sub-sample.

[27]A worker is eligible for unemployment benefits if he or she has worked in the recent past. Students thus

wages and earnings (in the men's subsample[28]).

See Figure A3 for a plot of the density of employment and search-duration indices. An alternative endogenous employment measure is the time spent searching for the first job, in months: we call this variable *search duration* for short. We also consider the duration of search before the first full-time job is found, called search duration to full-time job.

7.2 Appendix B. Orthogonality of Delay and Education Variables

The overall average delay is zero by construction:

$$\begin{aligned}\overline{\delta} &= \frac{1}{N}\sum_{s}\sum_{i\in B(s)}(d_i-\tau(s)) \\ &= \frac{1}{N}\sum_{s}\left\{N(s)\left[\sum_{i\in B(s)}\frac{d_i}{N(s)}-\tau(s)\right]\right\} \\ &= 0.\end{aligned} \tag{27}$$

Let $\overline{s}=(1/N)\sum_{i=1}^{N}s_i$. The empirical covariance of s and δ can be computed as follows,

$$\begin{aligned}\widehat{cov}(s,\delta) &= \frac{1}{N}\sum_{i}s_i\delta_i-\overline{s}\overline{\delta}=\frac{1}{N}\sum_{i}s_i(d_i-\tau(s_i)) \\ &= \frac{1}{N}\sum_{s}\left[N(s)\left(\sum_{i\in B(s)}s\frac{d_i}{N(s)}\right)-sN(s)\tau(s)\right] \\ &= 0.\end{aligned} \tag{28}$$

7.3 Appendix C. Other Possible Definitions of Delay

7.3.1 Delay as a residual

Instead of defining delay as above, we could have chosen $\delta^*=d-E(d\mid s,X,\theta_1)$ if Spence's hypothesis holds, or even $\delta^*=d-E(d\mid s,X,\theta_1,\theta_2)$ if Becker's hypothesis holds. Suppose

get zero before their first job. The unemployment benefits are roughly a half of the lost job's wage.

[28]The first wage density depicted on Fig. A2 is not necessarily a first full-time wage.

that Becker's hypothesis is true, then, the real model is

$$\ln(w) = a_0 s + b_0 \delta^* + X c_0 + \theta_1 + \theta_2, \quad (29)$$

$$s = X c_1 + Z g_1 + f_1 \theta_0 + \xi_1, \quad (30)$$

$$\delta^* = Z g_2 + f_2 \theta_0 + \xi_2. \quad (31)$$

We define $\delta^* = \delta + U$, where $U = E(d \mid s) - E(d \mid s, X, \theta_1, \theta_2)$. Under the normality assumption, we get $U = a_4 s + X c_4 + f_4 \theta_1 + f_5 \theta_2$, for some coefficients (a_4, c_4, f_4, f_5). Substituting $\delta +$ U in the log-wage equation and then s by its expression in the delay equation, we obtain a model expressed in terms of s and δ instead of s and δ^*,

$$\ln(w) = (a_0 + b_0 a_4) s + b_0 \delta + X(c_0 + b_0 c_4) + [(1 + b_0 f_4)\theta_1 + (1 + b_0 f_5)\theta_2], \quad (32)$$

$$s = X c_1 + Z g_1 + [f_1 \theta_0 + \xi_1], \quad (33)$$

$$\delta = X(-c_4 - c_1 a_4) + Z(g_2 - g_1 a_4) + [(f_2 - f_1 a_4)\theta_0 - f_4 \theta_1 - f_5 \theta_2 + \xi_2 - a_4 \xi_1]. \quad (34)$$

The terms between square brackets being linear combinations of the original random perturbations, given the assumptions about Z, the θs and the ξs, we still get a standard simultaneous-equations model, with the same structure as above, that can be estimated by 3SLS. The coefficients cannot all be identified, of course: we only get $a = a_0 + b_0 a_4$ and $b = b_0$ and $c = c_0 + b_0 c_4$.

Under Spence's hypothesis, θ_2 is not observed by employers. We thus now have $\delta^* = d - E(d \mid s, X, \theta_1)$. This is tantamount to setting $f_5 = 0$. Now, $U = E(d \mid s) - E(d \mid s, X, \theta_1)$. Given that (s, X, θ_1, δ) and $(s, X, \theta_1, \delta^*)$ provide the same information, we can still write a prediction of θ_2 of the form,

$$E(\theta_2 \mid s, \delta^*, X, \theta_1) = a_3 s + b_3 \delta + X c_3 + f_3 \theta_1. \quad (35)$$

Using the same derivation as with the benchmark model above, we get the model,

$$\ln(w) = (a_0 + b_0 a_4 + a_3) s + (b_0 + b_3)\delta + X(c_0 + b_0 c_4 + c_3) + (1 + b_0 f_4 + f_3)\theta_1, \quad (36)$$

$$s = X c_1 + Z g_1 + f_1 \theta_0 + \xi_1, \quad (37)$$

$$\delta = X(-c_4 - c_1 a_4) + Z(g_2 - g_1 a_4) + (f_2 - f_1 a_4)\theta_0 - f_4 \theta_1 + \xi_2 - a_4 \xi_1. \quad (38)$$

Again, the structure of this model is the same as that of the benchmark under Spence's hypothesis. The coefficients cannot all be identified, of course: we only get $a = a_0 + b_0 a_4 + a_3$ and $b = b_0 + b_3$ and $c = c_0 + b_0 c_4 + c_3$.The coefficient on delay is exactly the same and the model can be estimated by 3SLS with the same intruments Z. The error U is completely inoccuous if, as we believe, $b_0 = 0$.

7.3.2 Delay based on normal age

If we now assume that delay is school-leaving age minus normal age, we get $\delta = d - s$. Assume that the log-productivity equation is the same as (1) above, the education equation is the same as (2) and

$$\delta = Xc_2 + Zg_2 + f_2\theta_0 + \mu s + \xi_2, \tag{39}$$

where $\mu > 0$. The latter parameter is nonnegative because pushing studies further can only increase delay, *ceteris paribus*. In this model, the log-wage equation is the same as before, with a residual equal to $\theta_1 + \theta_2$ under Becker's hypothesis, and to $(1+ f_3)\theta_1$ under Spence's hypothesis. In (44), replace s with its expression in (2). We obtain,

$$\delta = X(c_2 + \mu c_1) + Z(g_2 + \mu g_1) + (f_2 + \mu f_1)\theta_0 + \mu\xi_1 + \xi_2.$$

We again derived a simultaneous equation model with the same overall structure as the benchmark. In this model, note that

$$Cov(\delta, \theta_1) = (f_2 + \mu f_1)Cov(\theta_0, \theta_1). \tag{40}$$

Then, if $\mu = 0$, the covariance given by (45) should be negative if θ_0 and θ_1 are positively correlated talents and $f_2 < 0$. But if $\mu > 0$, since we have in mind that $f_1 > 0$, it is possible to find that $Cov(\delta, \theta_1) > 0$. Under these assumptions, b would be overestimated by OLS. The results obtained with this latter model are close to the benchmark results discussed above.

7.4 Appendix D. Further Properties of the First Stage

Table D1 gives the correlation coefficients of the main instruments with the endogenous variables and some of the controls. These results are of course not a proof of the validity of instruments, but they give some interesting indications. If distance-to-college has a significant but moderate (i.e. 10%) correlation with wages, it is not at all correlated with the employment rate variable. The stock of high schools and $\Delta Stock$ variables are not significantly correlated with wages and employment. Distance-to-college has an 11% negative correlation with the father-went-to-college dummy, but the *Stock* and and $\Delta Stock$ variables are either weakly or not significantly correlated with father-went-to-college. Low density places tend to have more vocational high schools and to be more distant from Universities.

Table D2 sheds some light on the impact of instruments in sub-samples. We have partitioned the sample according to the father's occupation, and the table reports the results of first-stage regressions in the various sub-samples (t-statistics are in parentheses). Distance-to-college works well for the sons of executives only (this category including the sons of doctors, lawyers, engineers and teachers — to sum up, highly educated fathers). In contrast, the *Stock* and $\Delta Stock$ variables are significant, with the same order of magnitude, in the education equation for all sub-samples. These stock variables have a significant impact, of the same order of magnitude in all delay equations, except for the executives' sons. Removing the students residing in the Paris region at grade 6 entry does not affect the significance of the results. Those living in the Paris region (which includes several counties or *départements*) at the age of grade 6 entry are more likely to study less in the suburbs where the stock of vocational high schools is high: this is due to the clear social stratification of these counties. To sum up, the *Stock* and $\Delta Stock$ instruments have a relatively balanced effect in sub-samples, while the distance-to-college instrument seems to affect mostly students from highly educated families.

7.5 Appendix E. Additional Results and Comments on the Linear Benchmark Model

Table E1, column C, shows the results obtained when school-leaving age is used instead of our education variable (but ignoring delay); column D shows that using our education variable instead of school-leaving age would not yield very different results, if we still ignore delay. Finally, column E in table E1 shows the results if we use only delay and ignore education: in this case, delay has a positive return. In each case, instruments are rejected by the over-identification F-test: this is because instruments then capture the impact of the missing endogenous variable, as they would in a reduced-form specification.

Table E2 gives the matrix of correlation coefficients of the benchmark's residuals $(\nu, \zeta, \varepsilon, \eta)$. The correlation of the education and wage equations' residuals is negative, $corr(\nu, \varepsilon) = -19\%$, reflecting a negative ability bias, while the correlation of the delay and wage equation's residuals is substantial and positive, i.e., $corr(\nu, \eta) = 45\%$.

The OLS bias on the coefficients given by column A in Table 2 is easy to compute. Using the fact that delay and education are orthogonal, while delay has a zero mean, and the assumed properties of instruments, we get the simple formulas,

$$p \lim(\widehat{a}_{OLS}) = a + \frac{Cov(\varepsilon, \nu)}{Var(s)}, \quad p \lim(\widehat{b}_{OLS}) = b + \frac{Cov(\eta, \nu))}{Var(\delta)}.$$

Using the notation introduced above, under Spence's Hypothesis, we have $Cov(\varepsilon, \nu) = f_1(1+f_3)Cov(\theta_0, \theta_1)$ and $Cov(\varepsilon, \nu) = f_2(1+f_3)Cov(\theta_0, \theta_1)$. So, to explain the facts, there must be an element of negative covariance in the factor structure related to ability. First, remark that $1 + f_3 > 0$ is a reasonable assumption, because $f_3 \simeq Cov(\theta_1, \theta_2)/Var(\theta_1)$ (if s and δ are not strongly correlated with θ_1). Then, if f_1 and f_2 have opposite signs, and θ_0 is interpreted as a talent, i.e., $f_2 < 0$, we need to assume $Cov(\theta_0, \theta_1) < 0$ to get the right sign pattern. If, on the contrary, we assume $Cov(\theta_0, \theta_1) \geq 0$, then θ_0 must have a negative impact on education and a positive impact on delay, i.e., θ_0 is a handicap. To sum up, a possible way to explain the estimated sign pattern of the correlation matrix is to assume the existence of two ability factors with a negative covariance. For instance, there would exist an "academic ability" and a "practical ability" factor, both productive, but with the likely property that the individuals who possess more of one talent have less of the other.

There are other explanations for this negative ability bias in the literature. The negative covariance of θ_0 and θ_1 could also be explained by job opportunities: the occurrence of an attractive opportunity would at the same time increase wages and shorten education and delay (see e.g., Vella and Gregory (1996)).

Table E1: School-Leaving Age, Education and Delay

Mean Wage						
	C		D		E	
	OLS	3SLS	OLS	3SLS	OLS	3SLS
School-Leaving Age	6.24%	6.91%	—	—	—	—
	(57.57)	(9.56)	—	—	—	—
Education	—	—	6.50%	7.47%	—	—
	—	—	(69.57)	(11.12)	—	—
Delay	—	—	—	—	0.66%	5.49%
	—	—	—	—	(3.33)	(2.05)
Fisher Overid. test (*p*-value)	—	0.0001	—	0.0062	—	0.0001
R^2	0.3316	0.1890	0.3911	0.2041	0.1520	0.1441

Employment						
	C		D		E	
	OLS	3SLS	OLS	3SLS	OLS	3SLS
School-Leaving Age	6.06%	7.99%	—	—	—	—
	(18.35)	(3.63)	—	—	—	—
Education	—	—	5.57%	9.55%	—	—
	—	—	(18.67)	(4.45)	—	—
Delay	—	—	—	—	-1.85%	2.87%
	—	—	—	—	(-3.39)	(0.40)
Fisher Overid. test (*p*-value)	—	0.0157	—	0.1068	—	0.0001
R^2	0.0454	0.0207	0.046	0.0209	0.0201	0.019 1

Table E2: Correlation Matrix of Residuals (Benchmark Model)

	Mean Wage	Employment	Education	Delay
Mean Wage	1	0.3134	-0.1922	0.4531
Unemployment	0.3134	1	-0.1756	0.2346
Education	-0.1922	-0.1756	1	-0.0279
Delay	0.4531	0.2346	-0.0279	1

7.6 Appendix F. Derivation of ΔV

We easily get,

$$\begin{aligned} \Delta V(s \,|\, \epsilon) &= E\left[(T-d_s)u_s - (T-d_{s-1})u_{s-1} + \Delta d_s(\ln(h_s)+u_{s-1}) \,|\, \epsilon\right] \\ &= E\left[(T-d_s)\Delta u_s + \Delta d_s \ln(h_s) \,|\, \epsilon\right]. \end{aligned} \tag{41}$$

Denoting $\Delta a_s = a_s - a_{s-1}$, and $\Delta\alpha_s = \alpha_s - \alpha_{s-1}$, we find

$$\Delta u_s = \Delta a_s + \Delta \alpha_s \quad \text{and} \quad \Delta d_s = \Delta\tau(s). \tag{42}$$

It then follows that,

$$\Delta V(s \,|\, \epsilon) = \Delta u_s[T - E(d_s \,|\, \epsilon)] + \Delta d_s[-X_3\gamma_3 - c_s + \epsilon]. \tag{43}$$

But

$$E(d_s \,|\, \epsilon) = \tau(s) + X_2\gamma_2 + E(\eta \,|\, \epsilon) = E(d_s) + E(\eta \,|\, \epsilon). \tag{44}$$

Because of normality, $E(\eta \,|\, \epsilon) = (\sigma_{\epsilon\eta}/\sigma_\epsilon^2)\epsilon = \epsilon\sigma_{\epsilon\eta}$. Hence,

$$\begin{aligned} \Delta V(s \,|\, \epsilon) &= \Delta u_s[T - E(d_s) - \epsilon\sigma_{\epsilon\eta}] + \Delta d_s[-X_3\gamma_3 - c_s + \epsilon] \\ &= \Delta u_s[T - E(d_s)] - \Delta d_s[X_3\gamma_3 + c_s] + \epsilon[\Delta d_s - \sigma_{\epsilon\eta}\Delta u_s]. \end{aligned} \tag{45}$$

7.7 Appendix G. Second-Order Conditions

The discrete concavity condition $k_s < k_{s+1}$, for all s, can be written,

$$\frac{\Delta d_{s+1}}{A_{s+1}}(X_3\gamma_3 + c_{s+1}) - \frac{\Delta d_s}{A_s}(X_3\gamma_3 + c_s) > \frac{\Delta u_{s+1}(T - E(d_{s+1}))}{A_{s+1}} - \frac{\Delta u_s(T - E(d_s))}{A_s},$$

where by definition, $A_s = \Delta d_s - \sigma_{\epsilon\eta}\Delta u_s$. Assume that $\sigma_{\epsilon\eta} \simeq 0$, so that $A_s \simeq \Delta d_s$; then, the above inequality is approximately equivalent to

$$\Delta c_{s+1} > \left(\frac{\Delta u_{s+1}}{\Delta d_{s+1}} - \frac{\Delta u_s}{\Delta d_s}\right)(T - E(d_s)) - \frac{\Delta u_{s+1}}{\Delta d_{s+1}}(E(d_{s+1} - d_s)).$$

Given that $E(d_{s+1}) - E(d_s) = E(\Delta d_{s+1}) = \Delta d_{s+1}$, we get the equivalent condition,

$$\Delta c_{s+1} + \Delta u_{s+1} > \left(\frac{\Delta u_{s+1}}{\Delta d_{s+1}} - \frac{\Delta u_s}{\Delta d_s}\right)(T - E(d_s)).$$

This latter condition is sufficient for $k_s < k_{s+1}$, provided that $\sigma_{\epsilon\eta}$ is sufficiently small. It is easy to see that the condition holds under the stronger conditions of "increasing cost", i.e., $\Delta c_{s+1} \geq 0$, increasing utility $\Delta u_{s+1} \geq 0$, and concave utility, i.e., if $\Delta u_s/\Delta d_s$ is decreasing with s. But the model can easily accomodate moderately increasing returns, i.e., $\Delta u_s/\Delta d_s$ increasing with s, provided that $\Delta c_s + \Delta u_s$ is high enough, and T is not too large.

7.8 Appendix H. Estimation of the Nonlinear Model and Likelihood

Given the normality assumptions, we can write,

$$\epsilon = E(\epsilon|\nu, \zeta, \eta) + \xi, \tag{46}$$

where ξ is an independent error term, orthogonal to (ν, ζ, η), and we have

$$E(\epsilon|\nu, \zeta, \eta) = r_0\nu + r_1\zeta + r_2\eta, \tag{47}$$

where $r = (r_0, r_1, r_2)$ are theoretical regression coefficients, given by the formula,

$$r = (r_0, r_1, r_2) = (\sigma_{\epsilon\nu}, \sigma_{\epsilon\zeta}, \sigma_{\epsilon\eta})\Omega_0^{-1}. \tag{48}$$

where Ω_0 is the submatrix,

$$\Omega_0 = \begin{pmatrix} \sigma_\nu^2 & \sigma_{\nu\zeta} & \sigma_{\nu\eta} \\ \sigma_{\nu\zeta} & \sigma_\zeta^2 & \sigma_{\zeta\eta} \\ \sigma_{\nu\eta} & \sigma_{\zeta\eta} & \sigma_\eta^2 \end{pmatrix}, \tag{49}$$

Using the constraint $Var(\epsilon) = 1$, we see that the variance of ξ satisfies,

$$\begin{aligned} Var(\xi) &= 1 - Var(r_0\nu + r_1\zeta + r_2\eta) \\ &= 1 - \rho^2 \end{aligned} \tag{50}$$

where by definition,

$$\rho^2 = r'\Omega_0 r = \begin{pmatrix} \sigma_{\epsilon\nu} & \sigma_{\epsilon\zeta} & \sigma_{\epsilon\eta} \end{pmatrix} \Omega_0^{-1} \begin{pmatrix} \sigma_{\epsilon\nu} \\ \sigma_{\epsilon\zeta} \\ \sigma_{\epsilon\eta} \end{pmatrix}. \tag{51}$$

We can now derive individual contributions to likelihood, denoted, L_i. First, given the ordered Probit structure, we have,

$$\Pr(s_i = s, x_i, y_i, \delta_i) = \int_{k_s}^{k_{s+1}} pdf(\epsilon \mid x_i, y_i, \delta_i) pdf(x_i, y_i, \delta_i) d\epsilon, \tag{52}$$

using the decomposition $pdf(x_i, y_i, \delta_i, \epsilon) = pdf(\epsilon \mid x_i, y_i, \delta_i) pdf(x_i, y_i, \delta_i)$, and the densities involved are normal. Now define,

$$\widehat{\nu}_{is} = x_i - a_s - X_i\gamma_0 - b\delta_i, \tag{53}$$

$$\widehat{\zeta}_{is} = y_i - \alpha_s - X_i\gamma_1 - \beta\delta_i, \tag{54}$$

$$\widehat{\eta}_{is} = \delta_i - X_{i2}\gamma_2. \tag{55}$$

The transformation $(x, y, \delta) \mapsto (\nu, \zeta, \eta)$ is linear, one-to-one and upper triangular, that is,

$$\begin{pmatrix} \nu \\ \zeta \\ \eta \end{pmatrix} = \begin{pmatrix} 1 & 0 & -b \\ 0 & 1 & -\beta \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ \delta \end{pmatrix} + C, \tag{56}$$

where C is a vector of functions of X, s, and model parameters. It follows that the Jacobian determinant J of this transformation is equal to the product of its diagonal terms, i.e., $J = 1$. Thus, on the integration domain, the vector (x_i, y_i, δ_i) is normal and has the following p.d.f., denoted ψ,

$$\psi(x_i, y_i, \delta_i) = \frac{1}{(\sqrt{2\pi})^3 \sqrt{\det \Omega_0}} \exp\{-\frac{1}{2} \begin{pmatrix} \widehat{\nu}_{is} & \widehat{\zeta}_{is} & \widehat{\eta}_{is} \end{pmatrix} \Omega_0^{-1} \begin{pmatrix} \widehat{\nu}_{is} \\ \widehat{\zeta}_{is} \\ \widehat{\eta}_{is} \end{pmatrix} \}. \tag{57}$$

We can therefore factor out $\psi(x_i, y_i, \delta_i)$ in the expression of $\Pr(s_i = s, x_i, y_i, \delta_i)$. This yields,

$$\Pr(s_i = s, x_i, y_i, \delta_i) = \psi(x_i, y_i, \delta_i) \int_{k_s}^{k_{s+1}} pdf(\epsilon \mid x_i, y_i, \delta_i) d\epsilon.$$

Let $\Phi(x) = \int_{-\infty}^{x} \phi(v)dv$, be the standard normal c.d.f., and $\phi(x) = (\sqrt{2\pi})^{-1}\exp(-x^2/2)$ be the standard normal p.d.f. The distribution of ϵ_i conditional on (x_i, y_i, δ_i), X_i and the parameters, is the same as the distribution of ϵ_i conditional on (ν_i, ζ_i, η_i), X_i and the parameters, given the one-to-one mapping between the two vectors (the conditioning with respect to X is kept implicit everywhere). This conditional distribution is normal, with mean $(r_0\widehat{\nu}_{is} + r_1\widehat{\zeta}_{is} + r_2\widehat{\eta}_{is})$ and variance $1-\rho^2$. Therefore, we have,

$$\Pr(s_i = s, x_i, y_i, \delta_i) = \psi(x_i, y_i, \delta_i) \int_{k_s}^{k_{s+1}} \phi\left(\frac{\epsilon - r_0\widehat{\nu}_{is} - r_1\widehat{\zeta}_{is} - r_2\widehat{\eta}_{is}}{\sqrt{1-\rho^2}}\right) \frac{d\epsilon}{\sqrt{1-\rho^2}}. \tag{58}$$

Integration finally yields,

$$\Pr(s_i = s, x_i, y_i, \delta_i) = \psi(x_i, y_i, \delta_i)(\Phi^s_{s+1,i} - \Phi^s_{s,i}), \tag{59}$$

where by definition,

$$\Phi^s_{s,i} = \Phi\left[\frac{k_{i,s} - r_0\widehat{\nu}_{is} - r_1\widehat{\zeta}_{is} - r_2\widehat{\eta}_{is}}{\sqrt{1-\rho^2}}\right] \tag{60}$$

and

$$\Phi^s_{s+1,i} = \Phi\left[\frac{k_{i,s+1} - r_0\widehat{\nu}_{is} - r_1\widehat{\zeta}_{is} - r_2\widehat{\eta}_{is}}{\sqrt{1-\rho^2}}\right].$$

The contribution to likelihood of an individual i is precisely $L_i = \psi(x_i, y_i, \delta_i)(\Phi^s_{s+1,i} - \Phi^s_{s,i})$ with $s = s_i$.

7.9 Appendix I. Complete results

Tables 7, 8a, 8b and 9 give the complete estimation results, for the Benchmark Linear Model, Model B, Model B with interactions of education and delay and Model A, respectively.

Table A1 : Empirical Distribution of Male School-Leaving Age, Conditional on Education Level

Age while leaving school	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
High school dropouts	*0.01*	***0.17***	***0.24***	***0.33***	***0.16***	***0.07***	*0.01*	*0.01*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Vocational degree	*0*	*0*	*0.03*	***0.30***	***0.37***	***0.21***	***0.06***	*0.02*	*0.01*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
High school graduates (grade 12)	*0*	*0*	*0*	*0.02*	***0.12***	***0.31***	***0.32***	***0.15***	***0.06***	*0.02*	*0.01*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
Two years of college (grade 14)	*0*	*0*	*0*	*0*	*0*	***0.13***	***0.27***	***0.28***	***0.19***	***0.08***	*0.02*	*0.01*	*0.01*	*0*	*0*	*0*	*0*	*0*
Four years of college (grade 16)	*0*	*0*	*0*	*0*	*0*	*0*	*0.04*	***0.17***	***0.20***	***0.27***	***0.13***	***0.09***	*0.04*	*0.03*	*0.01*	*0.01*	*0.01*	*0.01*
Graduate studies	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0.04*	***0.24***	***0.29***	***0.15***	***0.11***	***0.07***	*0.04*	*0.02*	*0.02*	*0.01*	*0*

Table A2 : Distribution of Education Variable, Conditional on Parental Education

	13	14	15	17	18	20	21	22	23	Total
Father's education										
Without Qualification	4.7	18.4	16.2	24.1	12.7	13.3	2.2	2.8	5.6	16.9
Elementary Certificate	1.6	9.3	13.0	24.7	17.7	19.8	2.3	3.1	8.6	33.7
Vocational Degree	1.9	11.7	17.7	22.8	15.7	18.0	1.6	2.8	7.8	22.6
High School Degree	0.8	5.0	7.6	15.7	18.7	22.4	4.5	7.0	18.3	7.0
College	0.5	3.3	4.9	11.2	12.8	16.7	3.0	8.1	39.4	10.3
Observation Missing	6.5	22.5	24.5	24.2	10.7	8.3	0.9	0.9	1.5	9.5
Mother's education										
Without Qualification	4.5	17.7	16.9	23.0	14.0	13.8	1.7	2.3	6.0	22.1
Elementary Certificate	1.4	9.4	13.4	24.3	17.0	19.4	2.4	3.3	9.4	38.6
Vocational Degree	2.1	9.7	15.7	22.4	15.8	19.2	2.5	4.0	8.7	14.1
High School Degree	0.9	5.7	8.8	16.7	15.8	19.9	3.9	6.6	21.6	9.2
College	0.2	3.7	4.6	10.6	14.4	17.8	2.1	8.0	38.6	6.9
Observation Missing	6.3	22.2	24.4	24.5	10.2	8.5	0.5	0.8	2.7	9.1
Total	2.5	11.7	14.5	22.1	15.3	17.1	2.2	3.6	11.1	100.0

Figure A1: Duration of Schooling

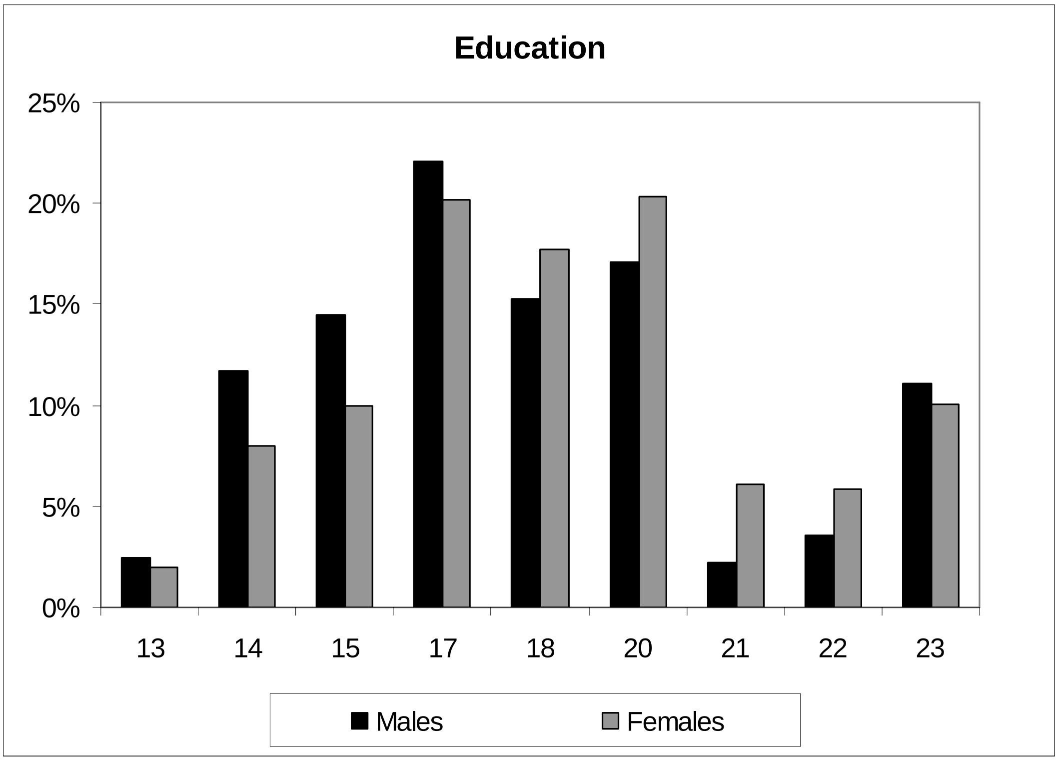

Figure A2 : Male Wage Distributions (in Euros)

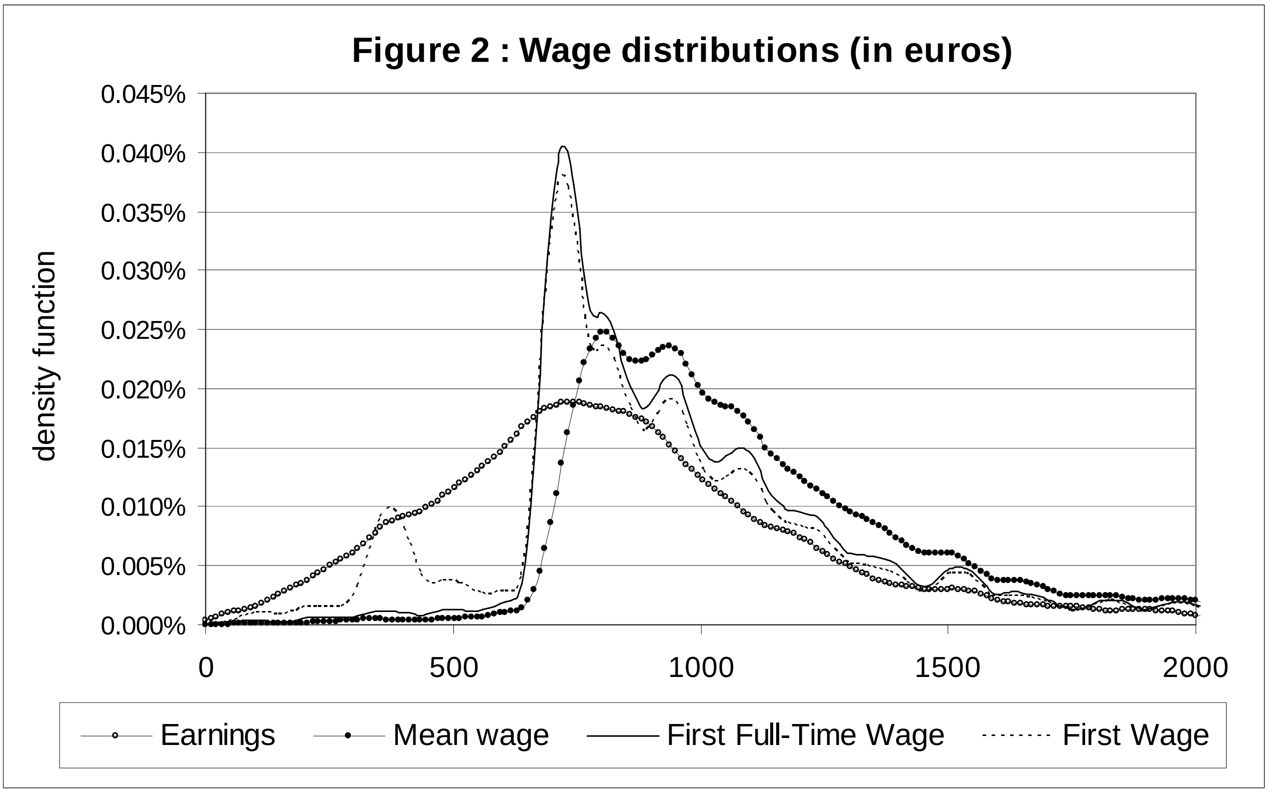

Figure A3: Male Unemployment Rate Distributions

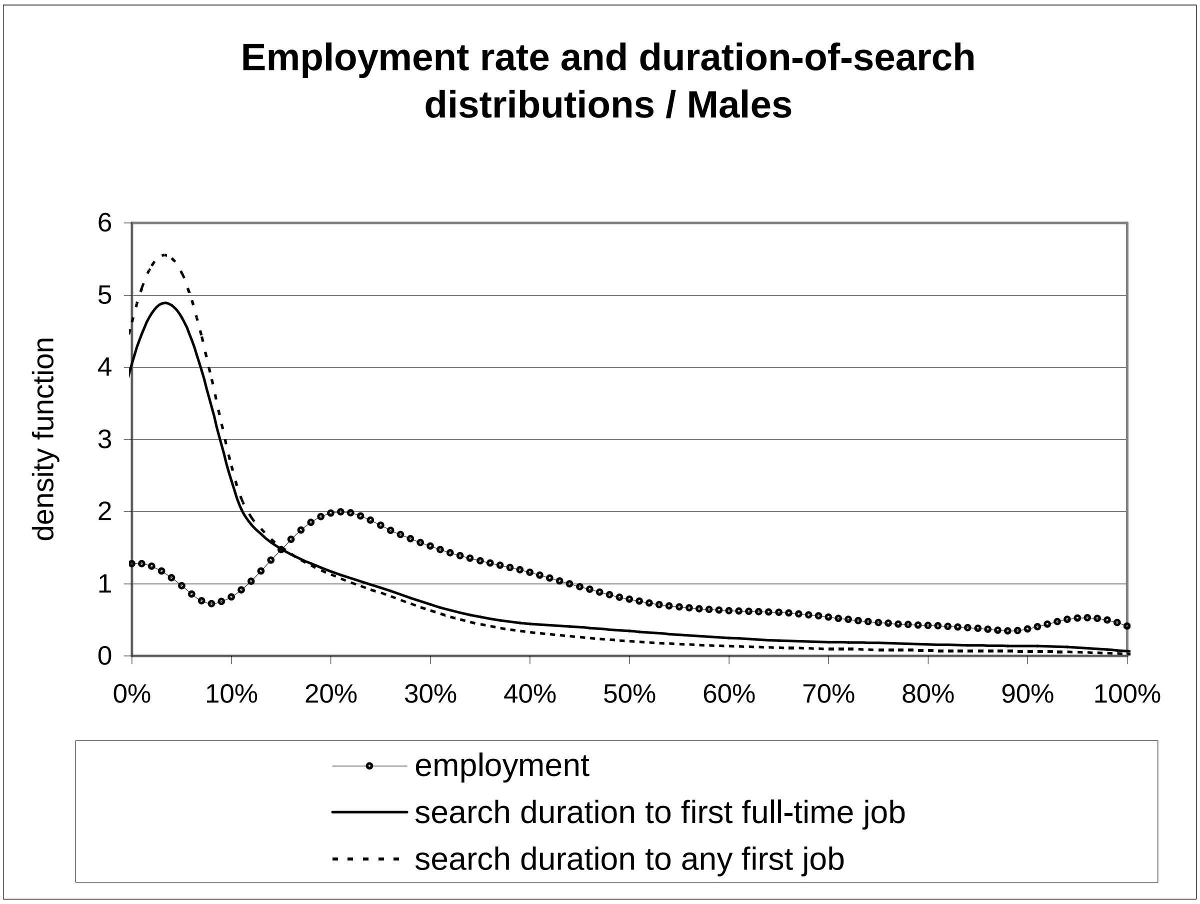

Figure 1a : Distribution of Delay

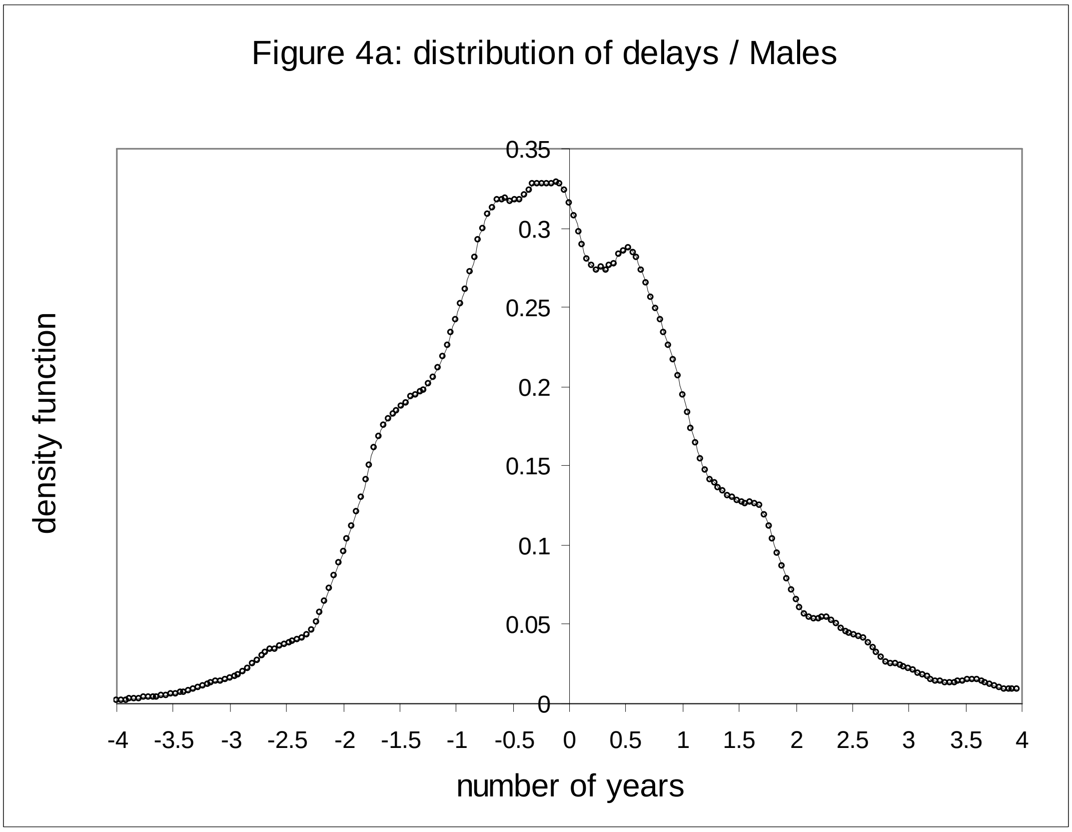

Figure 1b : Conditional Distribution of Delay (Impact of Father's Education)

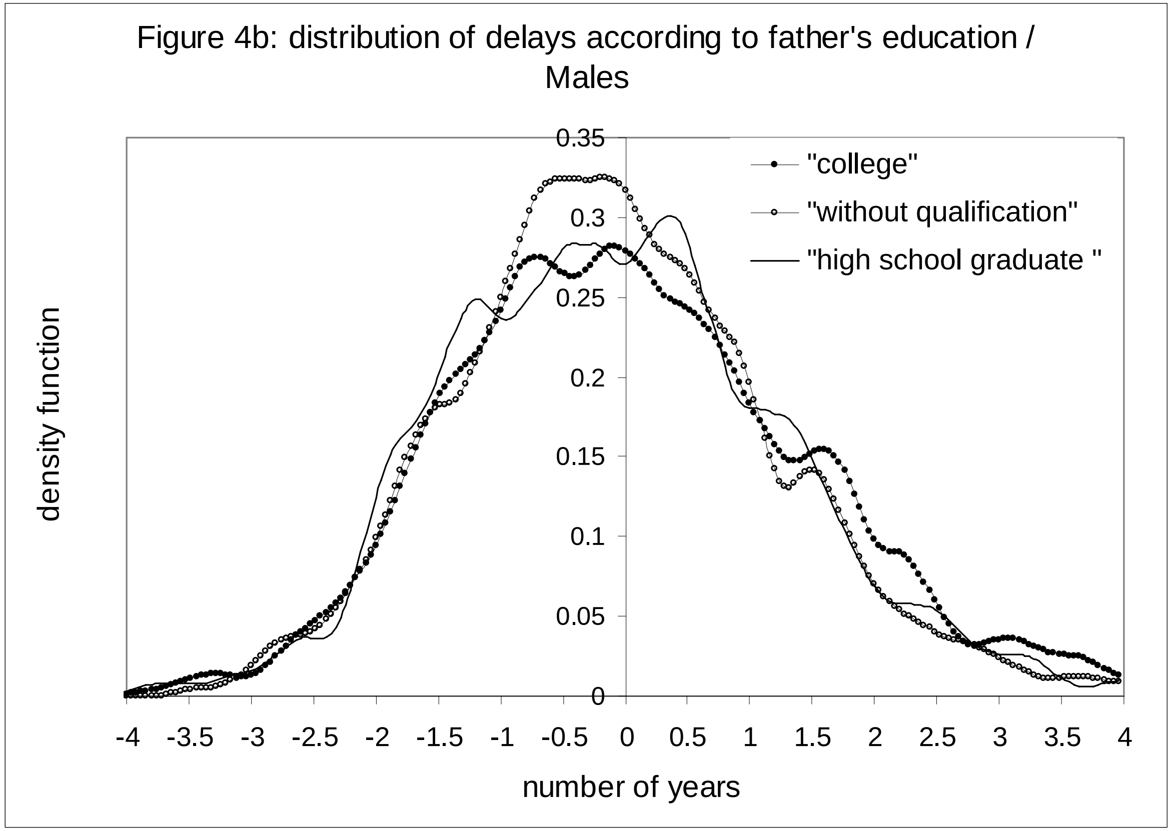

Figure 1c : Conditional Distribution of Delay (Impact of Son's Education)

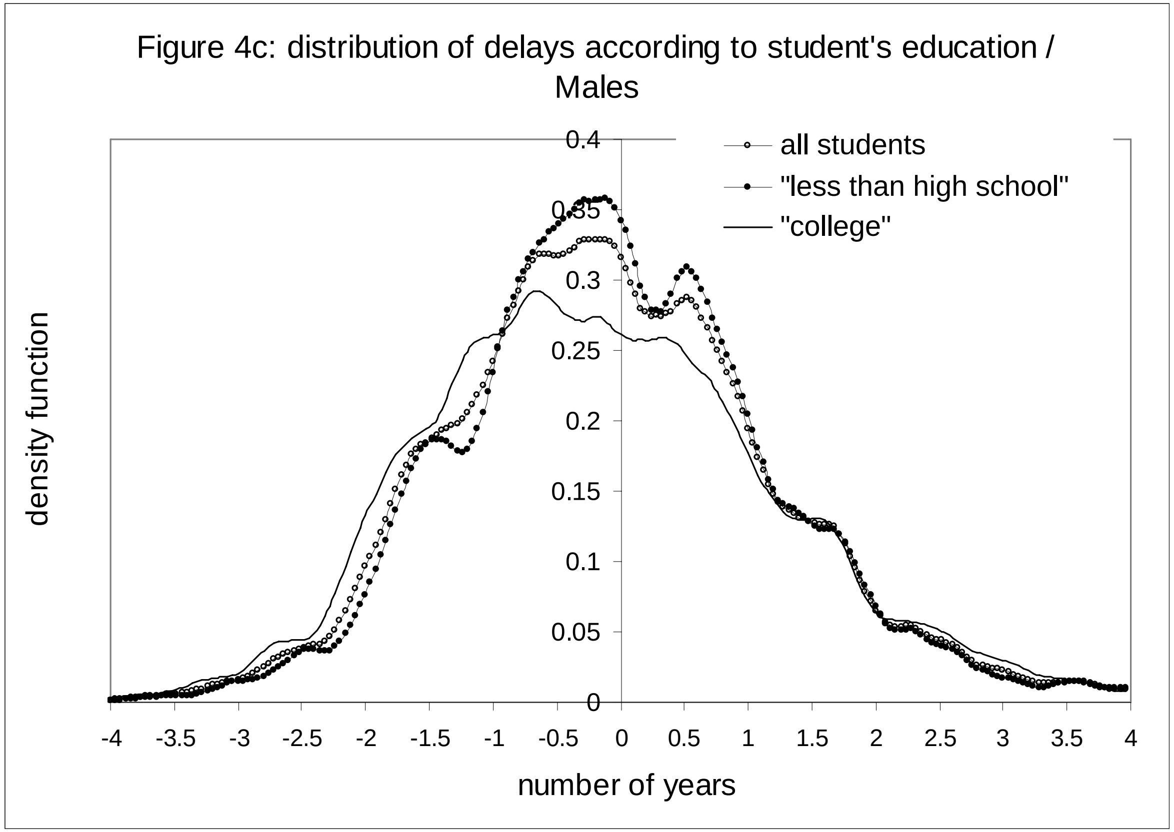

Figure 2 : Historical Growth of Vocational Secondary Education

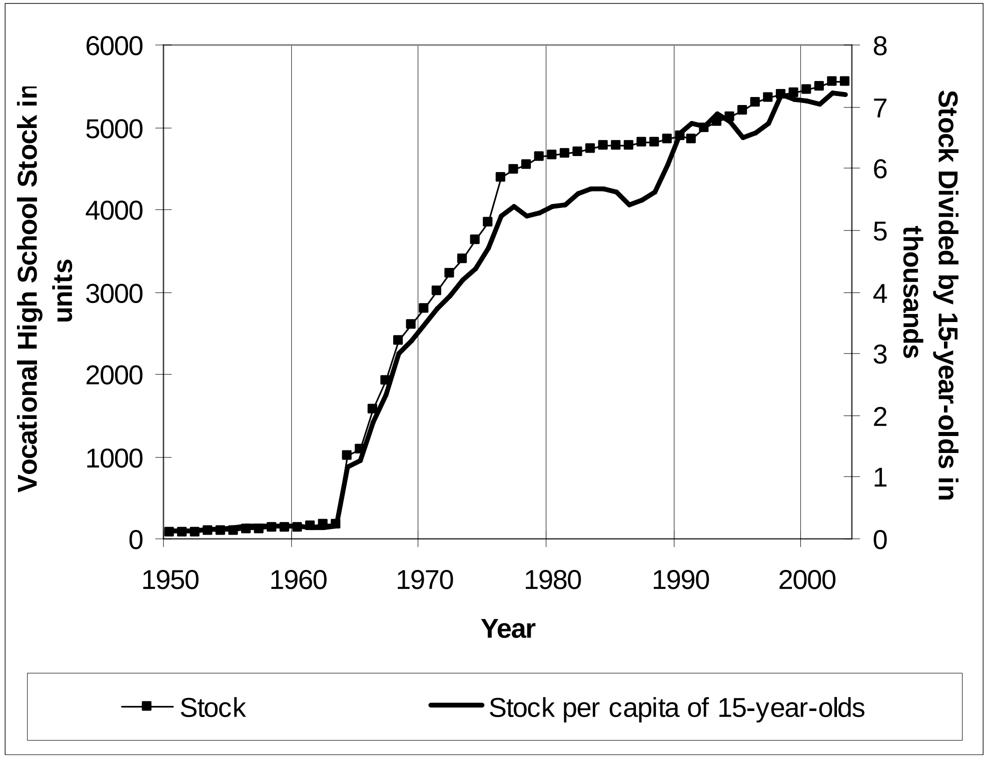

Figure 3 : Distribution of Stock of Vocational High Schools 1982

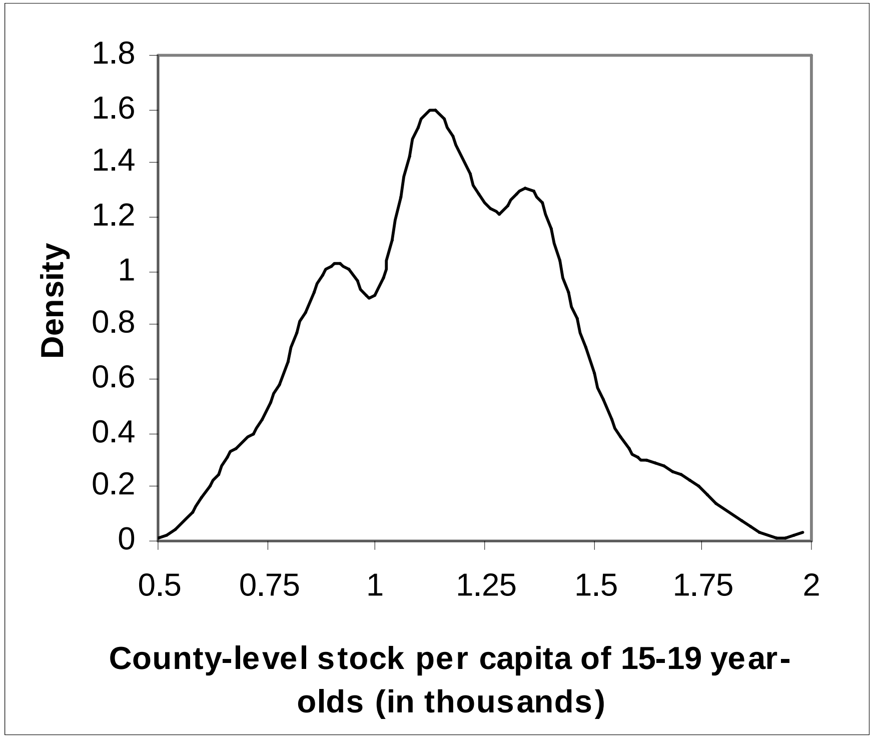

Figure 4 : Distribution of Δ Stock of Vocational High Schools 1989-82

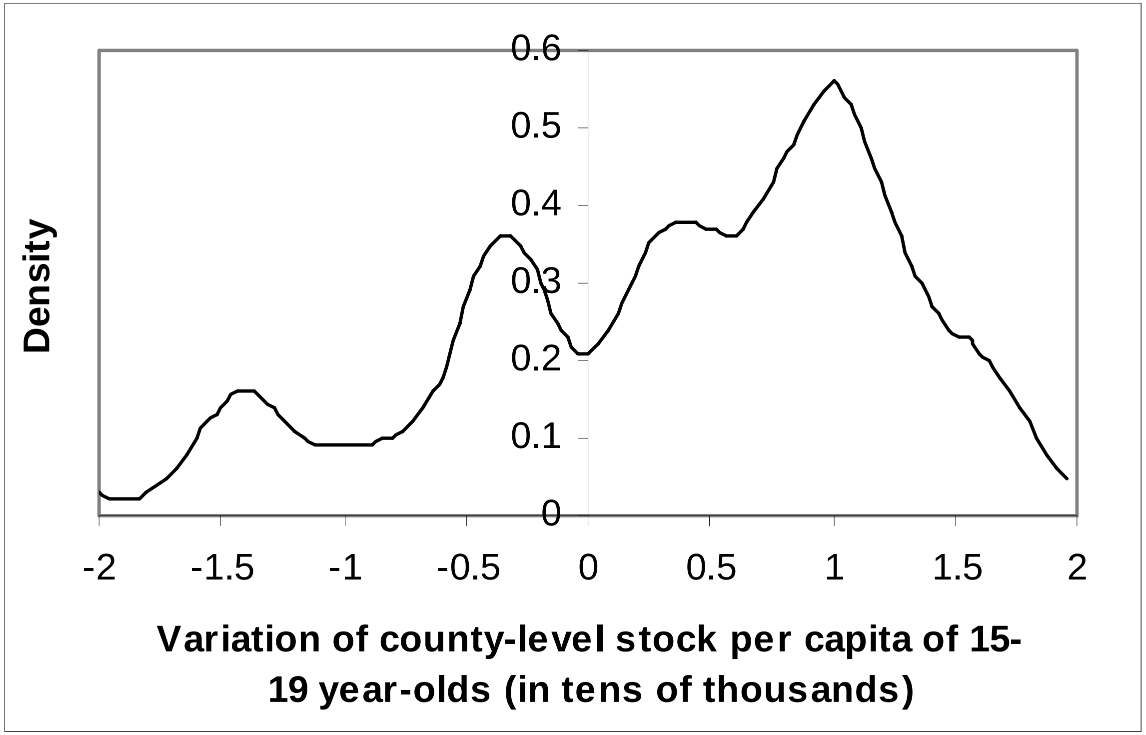

Table D1: Correlation of Main Instruments with Various Variables

	Distance to college	Stock of vocational high schools 1982	Δ stock of vocational schools 1989-82	Mean Wage	Employment
Distance to college	1	0.1930	0.0477	-0.0972	0.0047
		<.0001	*<.0001*	*<.0001*	*0.6035*
Stock of vocational high schools 1982	0.1930	1	-0.4621	-0.0093	-0.0001
	<.0001		*<.0001*	*0.3003*	*0.9908*
Δ stock of vocational schools 1989-82	0.0477	-0.4621	1	-0.0066	0.0105
	<.0001	*<.0001*		*0.4642*	*0.2462*

	Education	Delay	Father went to college	Local population density
Distance to college	-0.0325	-0.0830	-0.1140	-0.3821
	0.0003	*<.0001*	*<.0001*	*<.0001*
Stock of vocational high schools 1982	0.0566	0.0466	-0.0539	-0.1295
	<.0001	*<.0001*	*<.0001*	*<.0001*
Δ stock of vocational schools 1989-82	-0.0185	0.0196	-0.0016	-0.1239
	0.0404	*0.0295*	*0.8618*	*<.0001*

p-value of significance test is given in italics below estimated correlation

Table D2: Impact of Instruments in Various Sub-Samples

	Occupation of the Father							
	Farmer		Craftsman		Executive		Middle Manager	
	Education	Delay	Education	Delay	Education	Delay	Education	Delay
Distance to college	-0.0012	0.0023	-0.0081	-0.0069	-0.0295	-0.0099	-0.0140	-0.0088
	(-0.08)	(0.28)	(-0.90)	(-1.70)	(-3.34)	(-1.94)	(-1.38)	(-1.54)
Stock of vocational high schools 1982	0.0139	0.0116	0.0157	0.0036	0.0147	0.0000	0.0178	0.0088
	(3.23)	(4.86)	(4.43)	(1.90)	(4.53)	(0.01)	(4.70)	(4.08)
Δ stock of vocational schools 1989-82	0.0390	0.0190	0.0362	0.0041	0.0402	-0.0037	0.0256	0.0230
	(3.14)	(2.74)	(3.82)	(0.89)	(4.41)	(-0.70)	(2.41)	(3.82)
Number of Observations	679		1359		1956		1271	

	Occupation of the Father				Location at Grade 6 Entry			
	White Collar		Blue Collar		Not in Paris Area		Paris Area	
	Education	Delay	Education	Delay	Education	Delay	Education	Delay
Distance to college	-0.0080	-0.0018	0.0000	-0.0070	-0.0038	-0.0050	-0.0392	-0.0179
	(-1.20)	(-0.49)	(0.01)	(-2.52)	(-1.08)	(-2.65)	(-2.23)	(-1.65)
Stock of vocational high schools 1982	0.0167	0.0055	0.0204	0.0056	0.0210	0.0059	-0.0384	-0.0189
	(6.04)	(3.81)	(9.36)	(4.75)	(16.85)	(8.58)	(-6.09)	(-4.92)
Δ stock of vocational schools 1989-82	0.0436	0.0072	0.0218	0.0068	0.0354	0.0068	0.0660	0.0484
	(5.75)	(1.81)	(4.08)	(2.58)	(10.61)	(3.79)	(3.62)	(4.33)
Number of Observations	2377		3717		10842		1468	

Table 3a : Robustness Check Ia: Variants

		Benchmark	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5
Mean Wage	Delay	-9.26%	-9.83%	-11.92%	-7.89%	-7.55%	-7.15%
		(-3.29)	(-3.25)	(-1.85)	(-3.44)	(-2.27)	(-3.36)
	Education	9.12%	9.67%	7.16%	5.81%	6.55%	8.49%
		(10.49)	(9.41)	(2.79)	(6.76)	(5.40)	(23.02)
	p-value of overidentification F-test	**0.3984**	**0.3409**	**0.2468**	**0.5404**	**0.5210**	**0.0004**
Employment	Delay	-19.29%	-21.25%	-21.77%	-16.35%	-21.91%	-7.04%
		(-2.28)	(-2.35)	(-1.22)	(-2.40)	(-2.15)	(-1.12)
	Education	12.98%	14.87%	8.73%	7.15%	6.10%	7.11%
		(5.03)	(4.90)	(1.22)	(2.81)	(1.64)	(6.62)
	p-value of overidentification F-test	**0.2812**	**0.3165**	**0.2407**	**0.2317**	**0.1942**	**0.0178**
Additional controls of equation of interest	Paris Area 1992	Yes	Yes				Yes
	Average local unemployment rate 92-97	Yes	Yes				Yes
	Family structure		Yes				
	Age at grade 6 entry		Yes				
Instruments	Distance to college	Yes	Yes	Yes	Yes	Yes	Yes
	Stock of vocational high schools 1982	Yes	Yes	Yes	Yes		Yes
	Δ stock of vocational schools 1989-82	Yes	Yes	Yes	Yes		Yes
	Average local unemployment rate 82-87	Yes	Yes				Yes
	Paris Area at grade 6 entry	Yes	Yes				Yes
	Number of siblings				Yes	Yes	
	Mother at home					Yes	
	Parental education						Yes

Table 3b : Robustness Check Ib: Other Variants

		Variant 6	Variant 7	Variant 8	Variant 9	Variant 10
Mean Wage	Delay	-11.77%	-14.80%	-8.13%	-6.53%	-13.30%
		(-3.46)	(-2.33)	(-2.71)	(-1.61)	(-2.10)
	Education	8.78%	10.21%	8.91%	11.07%	7.93%
		(8.77)	(4.84)	(10.27)	(8.17)	(6.40)
	p-value of overidentification F-test	**0.5726**	**0.3052**	**0.2014**	**0.1815**	**---**
Employment	Delay	-14.16%	-34.61%	-15.88%	-16.76%	-9.69%
		(-1.49)	(-1.88)	(-1.74)	(-1.38)	(-0.60)
	Education	10.44%	15.42%	13.35%	11.77%	10.74%
		(3.85)	(2.51)	(5.13)	(2.87)	(3.68)
	p-value of overidentification F-test	**0.2216**	**0.3790**	**0.5802**	**0.0886**	**---**
Additional controls of equation of interest	Paris Area 1992		Yes	Yes	Yes	
	Average local unemployment rate 92-97	Yes		Yes	Yes	Yes
	Family structure					
	Age at grade 6 entry					
Instruments	Distance to college	Yes	Yes		Yes	
	Stock of vocational high schools 1982	Yes	Yes	Yes		
	Δ stock of vocational schools 1989-82	Yes	Yes	Yes		Yes
	Average local unemployment rate 82-87	Yes		Yes	Yes	Yes
	Paris Area at grade 6 entry		Yes	Yes	Yes	
	Number of siblings					
	Parental education					

Table 4 : Robustness Check II: Other Outcomes, Impact of Experience

	Last Wage			Mean Wage			First Full-Time Wage		
	A	B	C	A	B	C	A	B	C
Delay	-7.95% (-2.71)	-7.14% (-2.66)	-6.99% (-2.05)	-9.26% (-3.29)	-8.57% (-3.33)	-7.91% (-2.42)	-7.36% (-2.21)	-6.64% (-2.16)	-6.13% (-1.59)
Education	8.62% (9.59)	8.35% (9.73)	8.34% (8.03)	9.12% (10.49)	8.86% (10.74)	8.72% (8.70)	8.30% (8.16)	8.05% (8.20)	7.90% (6.81)
Jobs & Internships before 92	--- ---	4.35% (4.71)	2.01% (0.21)	--- ---	4.26% (4.83)	4.68% (0.51)	--- ---	4.45% (4.20)	5.05% (0.47)
R^2	0.1705	0.1938	0.1640	0.1653	0.1904	0.1598	0.1046	0.1196	0.1033
p-val. Fisher Overid. Test	**0.1151**	**0.0675**	**0.0471**	**0.3984**	**0.2629**	**0.1853**	**0.5140**	**0.4175**	**0.3020**

	Employment			Search to Full-Time Job		
	A	B	C	A	B	C
Delay	-19.29% (-2.28)	-17.55% (-2.23)	-15.49% (-1.58)	8.69% (0.68)	7.95% (0.66)	1.01% (0.07)
Education	12.98% (5.03)	12.36% (4.94)	11.88% (4.05)	-1.42% (-0.37)	-0.91% (-0.24)	0.71% (0.16)
Jobs & Internships before 92	--- ---	10.96% (4.04)	15.04% (0.56)	--- ---	-11.89% (2.86)	-63.17% (-1.52)
R^2	0.0196	0.0268	0.0194	0.0127	0.0175	0.0124
p-val. Fisher Overid. Test	**0.2812**	**0.2289**	**0.1316**	**0.0002**	**0.0003**	**0.0004**

Model A is the benchmark specification. Model B uses experience as a control. Model C treats experience as an endogenous variable (benchmark instruments).

Table 5 : Robustness Check III: Alternative Definitions of Delay

Definition of Delay		**in years**	**in months (divided by 12)**	**in months (divided by 12)**	**in months (divided by 12)**	**in years**	**in years**	**in years**
		benchmark	**benchmark**	**alternate 1**	**alternate 2**	**age at grade 6 entry**	**delay at grade 6 entry**	**delay accumulated after grade 6**
Mean Wage	Delay	-9.26%	-10.46%	-9.96%	-9.92%	-16.04%	-25.62%	-7.93%
		(-3.29)	(-3.28)	(-3.76)	(-3.74)	(-1.41)	(-1.99)	(-2.77)
	Education	9.12%	9.77%	9.01%	9.02%	6.63%	9.01%	8.45%
		(10.49)	(9.30)	(10.90)	(10.91)	(6.80)	(9.15)	(10.49)
	Overid F-test	**0.3984**	**0.3977**	**0.4674**	**0.4663**	**0.0132**	**0.0421**	**0.2074**
Employment	Delay	-19.29%	-21.43%	0.54%	-18.05%	-45.30%	-62.81%	-16.49%
		(-2.28)	(-2.26)	(0.08)	(-2.25)	(-1.24)	(-1.56)	(-1.93)
	Education	12.98%	14.82%	4.13%	12.35%	7.20%	13.35%	11.57%
		(5.03)	(4.82)	(2.05)	(5.03)	(2.31)	(4.34)	(4.97)
	Overid F-test	**0.2812**	**0.2931**	**0.7660**	**0.2243**	**0.1538**	**0.1991**	**0.1976**

Note. Alternate 1. Delay and employment have been changed. Alternate 2: Delay has been changed but employment variable has not been changed accordingly

Table 6 : Robustness Check IV: Impact of Mobility

	Subsample of immobile individuals	Benchmark	Control* for mobility; whole sample		Instrumented mobility; whole sample	
Mean Wage						
	10,586 observations	**---**	**Move****	**Distance*****	**Move****	**Distance*****
Delay	-5.62% (-1.31)	-9.26% (-3.29)	-8.11% (-2.87)	-8.57% (-3.07)	-7.72% (-2.44)	-8.62% (-3.09)
Education	6.69% (4.22)	9.12% (10.49)	8.37% (7.95)	8.46% (8.74)	7.61% (3.86)	7.11% (4.76)
Mobility	--- ---	--- ---	4.47% (2.09)	0.00014 (2.97)	12.78% (0.76)	0.00067 (1.47)
R^2	0.1226	0.1653	0.2029	0.1936	0.1713	0.1702
p-val. of Fisher Overid. Test	**0.1426**	**0.3984**	**0.3982**	**0.4585**	**0.2212**	**0.3164**
Employment						
	10,586 observations	**---**	**Move****	**Distance*****	**Move****	**Distance*****
Delay	-20.14% (-1.39)	-19.29% (-2.28)	-17.73% (-2.05)	-17.84% (-2.10)	-14.19% (-1.46)	-18.12% (-2.12)
Education	12.47% (2.34)	12.98% (5.03)	13.00% (4.10)	12.76% (4.41)	3.67% (0.61)	6.59% (1.45)
Mobility	--- ---	--- ---	-11.90% (1.85)	-0.00023 (-1.61)	79.31% (1.55)	0.00214 (1.53)
R^2	0.0161	0.0196	0.0212	0.0211	0.0194	0.0194
p-val. of Fisher Overid. Test	**0.4595**	**0.2812**	**0.2258**	**0.2585**	**0.5603**	**0.3501**

Note: Immobile individuals stayed in the same county (i.e., *département*) between grade 6 entry and 1992: 10,586 individuals.
*Mobility dummy or distance variable is added as a control in the four equations.
**Move is a dummy taking value 1 if the individual's county of residence has changed between grade 6 entry and 1992.
***Distance between location at grade 6 entry and location in 1992, in kilometers. Education is added in the mobility equation. Mobility is instrumented with benchmark instruments.

Table 7 : Linear Model: Complete Results

	Mean Wage		Employment		Education		Delay	
	Coeff.	**t**	**Coeff.**	**t**	**Coeff.**	**t**	**Coeff.**	**t**
Constant	7.2668	49.49	-2.7452	-6.25	14.9502	74.44	-1.1139	-10.20
Delay	-0.0926	-3.29	-0.1929	-2.28	-	-	-	-
Education	0.0912	10.49	0.1298	5.03	-	-	-	-
Father's occupation								
Farmer	-0.0681	-3.37	0.0244	0.41	0.5657	3.62	-0.1644	-1.89
Craftsman	0.0152	1.41	0.0778	2.48	0.0835	0.92	0.0164	0.33
Executive	0.0108	0.77	-0.0989	-2.39	1.1770	12.99	0.1016	2.02
Middle Manager	0.0200	1.66	-0.0390	-1.11	0.7177	8.01	0.0913	1.83
White Collar. **Reference group**								
Blue Collar	0.0173	1.98	-0.0082	-0.32	-0.2909	-4.21	0.0271	0.71
Missing or Deceased	-0.0097	-0.81	-0.0967	-2.77	-0.0153	-0.15	0.0641	1.16
Mother's occupation								
Farmer	-0.0396	-1.93	0.0468	0.78	0.0697	0.41	-0.0638	-0.67
Craftsman	0.0273	1.81	0.0477	1.08	0.0608	0.48	0.0247	0.35
Executive	0.0183	1.34	0.0145	0.37	0.4812	4.34	0.0946	1.54
Middle Manager	0.0149	1.04	0.0628	1.50	0.2550	2.22	0.1842	2.89
White Collar. **Reference group**								
Blue Collar	0.0145	1.53	0.0327	1.18	-0.4311	-5.79	-0.0630	-1.52
Missing or Deceased	0.0005	0.07	0.0160	0.73	0.0513	0.84	0.0779	2.30
Father's education								
High school dropouts. **Reference group**								
Vocational degree	-0.0072	-0.71	-0.0348	-1.18	0.6705	9.43	0.0207	0.52
Advanced vocational degree	0.0004	0.04	-0.0401	-1.37	0.4780	6.43	-0.0317	-0.77
High school graduates	-0.0001	-0.01	-0.1056	-2.18	1.2180	11.17	0.0271	0.45
Father went to College	0.0363	2.01	-0.1379	-2.61	1.6131	14.21	0.1855	2.94
Mother's education								
High school dropouts. **Reference group**								
Vocational degree	-0.0240	-2.41	-0.0108	-0.37	0.6869	10.43	-0.0166	-0.45
Advanced vocational degree	-0.0190	-1.68	0.0067	0.20	0.5537	6.69	-0.0431	-0.94
High school graduates	-0.0330	-2.06	-0.0523	-1.10	0.9919	9.98	-0.1305	-2.36
Mother went to College	-0.0167	-0.76	-0.1606	-2.45	1.1901	9.13	-0.2743	-3.79
Population aged 15-19, 1982	0.0292	4.20	0.0310	1.51	0.1526	2.41	0.0205	0.59
Local population density 1982	0.0000	-0.06	-0.0002	-0.66	0.0044	5.42	0.0010	2.25
Unemployment rate 92-97 (in 1992 county of residence)	-0.0043	-2.69	-0.0163	-3.27	-	-	-	-
Residence in 1992								
Reference (rest of France)								
Paris Area in 1992	0.0697	5.92	0.0241	0.69	-	-	-	-
Unemployment rate 82-87 (in county of residence at grade 6 entry)	-	-	-	-	-0.0525	-3.66	0.0545	6.89
Residence at grade 6 entry								
Reference (rest of France)								
Paris Area at grade 6 entry	-	-	-	-	0.5368	5.47	0.2898	5.45
Distance to college	-	-	-	-	-0.0067	-2.17	-0.0065	-4.09
(Distance to college)^2	-	-	-	-	0.0001	1.39	0.0001	4.19
Stock of vocational high schools 1982	-	-	-	-	0.0171	14.19	0.0048	7.30
Δ stock of vocational high schools 1989-82	-	-	-	-	0.0317	9.77	0.0074	4.38
Number of observations	12,310		12,310		12,310		12,310	
R-Squared	0.1654		0.0197		0.2158		0.0212	
Cross Model Correlation	Mean Wage		Employment		Education		Delay	
Mean Wage	1		0.3134		-0.1922		0.4531	
Unemployment	0.3134		1		-0.1756		0.2346	
Education	-0.1922		-0.1756		1		-0.0279	
Delay	0.4531		0.2346		-0.0279		1	

Table 8a : Model B

	Mean Wage		**Employment**		**Education**		**Delay**	
	Coeff.	**t**	**Coeff.**	**t**	**Coeff.**	**t**	**Coeff.**	**t**
Constant	8.5524	195.00	-1.3041	-13.03	-	-	-0.3337	-5.69
Delay	-0.0967	-3.10	-0.1971	-2.28	-	-	-	-
Education (Reference: Dropouts)					-	-	-	-
Vocational degree	0.1437	3.44	0.6022	6.39	-	-	-	-
High school graduates (grade 12)	0.1310	4.92	0.2094	3.40	-	-	-	-
Two years of college (grade 14)	0.1709	8.76	0.2976	6.31	-	-	-	-
Four years of college (grade 16)	0.1388	7.12	0.2037	4.01	-	-	-	-
Graduate studies	0.2786	10.17	0.2436	3.65	-	-	-	-
Father's occupation								
Farmer	-0.0633	-2.97	0.0012	0.02	0.2211	3.42	-0.1658	-1.91
Craftsman	0.0078	0.71	0.0704	2.14	0.0444	1.18	0.0098	0.20
Executive	0.0063	0.36	-0.1182	-2.63	0.4644	12.35	0.0889	1.76
Middle Manager	0.0236	1.76	-0.0507	-1.36	0.2782	7.52	0.0815	1.64
White Collar. **Reference group**								
Blue Collar	0.0120	1.28	-0.0031	-0.12	-0.1262	-4.38	0.0185	0.48
Missing or Deceased	-0.0138	-1.14	-0.0928	-2.55	-0.0075	-0.18	0.0652	1.18
Mother's occupation								
Farmer	-0.0378	-1.82	0.0373	0.60	0.0639	0.90	-0.0408	-0.43
Craftsman	0.0317	2.07	0.0380	0.83	0.0511	0.97	0.0520	0.74
Executive	0.0141	0.98	0.0090	0.21	0.1870	4.03	0.1049	1.70
Middle Manager	0.0163	1.11	0.0587	1.33	0.0979	2.05	0.1886	2.96
White Collar. **Reference group**								
Blue Collar	0.0059	0.59	0.0327	1.13	-0.1531	-4.91	-0.0648	-1.57
Missing or Deceased	-0.0009	-0.11	0.0199	0.85	0.0074	0.28	0.0841	2.48
Father's education								
High school dropouts. **Reference group**								
Vocational degree	0.0051	0.43	-0.0566	-1.79	0.2798	9.44	0.0222	0.55
Advanced vocational degree	0.0061	0.54	-0.0627	-2.02	0.2155	6.96	-0.0355	-0.86
High school graduates	0.0116	0.58	-0.1342	-2.59	0.4985	11.04	0.0308	0.51
Father went to College	0.0309	1.34	-0.1737	-2.94	0.6762	14.26	0.1909	3.02
Mother's education								
High school dropouts. **Reference group**								
Vocational degree	-0.0123	-1.03	-0.0296	-0.96	0.2812	10.23	-0.0116	-0.30
Advanced vocational degree	-0.0077	-0.61	-0.0090	-0.26	0.2276	6.62	-0.0430	-0.92
High school graduates	-0.0253	-1.32	-0.0770	-1.57	0.4075	9.85	-0.1220	-2.19
Mother went to College	-0.0144	-0.55	-0.1868	-2.78	0.5100	9.35	-0.2613	-3.59
Population aged 15-19, 1982	0.0227	3.63	0.0460	2.42	-0.0030	-0.13	-0.0112	-0.37
Local population density 1982	0.0012	0.15	-0.0679	-2.90	0.1528	6.86	0.1401	4.77
Unemployment 92-97	-0.0121	-2.13	-0.0608	-3.39	-	-	-	-
Residence in 1992								
Paris Area in 1992 (Reference: rest of France)	0.0725	7.78	0.0481	1.69	-	-	-	-
Unemployment rate 82-87	-	-	-	-	-0.0935	-4.48	0.1610	5.83
Residence at grade 6 entry								
Reference (rest of France)								
Paris Area at grade 6 entry	-	-	-	-	0.2056	5.61	0.2190	4.71
Distance to college	-	-	-	-	-0.0238	-0.90	0.0205	0.64
(Distance to college)^2	-	-	-	-	-0.0394	-1.56	-0.0748	-2.47
Stock of vocational high schools 1982	-	-	-	-	0.2096	9.37	0.1898	5.92
Δ stock of vocational high schools 1989-82	-	-	-	-	0.1344	6.12	0.1033	3.82
Ordered Probit Cuts								
κ1					-0.4520	-10.06		
κ2					0.7607	16.89		
κ3					1.2099	26.60		
κ4					1.8382	39.47		
κ5					2.1491	45.28		
Estimated Standard Deviations								
standard deviation	0.3042	11.93	0.9165	19.70968	1	-	1.4131	157.01
Estimated Correlation Matrix	Mean Wage		Employment		Education		Delay	
Mean Wage	1							
Unemployment	0.3642 (5.02)		1					
Education	-0.2010 (-1.92)		-0.3010 (-4.07)		1			
Delay	0.4996 (4.72)		0.2853 (2.35)		-0.0261 (-2.75)		1	
Mean Log-Likelihood	-4.57842							
Number of observations	12,310							

Table 8b : Model B with interactions

	Mean Wage		Employment	
	Coeff.	**t**	**Coeff.**	**t**
Delay				
Delay*High-School Dropouts	-8.7%	-2.70	-22.8%	-2.55
Delay*Vocational degree	-10.4%	-3.27	-24.7%	-2.80
Delay*High school graduates (grade 12)	-9.0%	-2.81	-22.3%	-2.51
Delay*Two years of college (grade 14)	-9.0%	-2.82	-15.7%	-1.77
Delay*Four years of college (grade 16)	-10.0%	-3.10	-15.0%	-1.67
Delay*Graduate studies	-10.1%	-3.17	-15.2%	-1.71
Mean Log-Likelihood	-4.57508			
Number of observations	12,310			

Table 9 : Model A

	Mean Wage		Employment		Education		Delay	
	Coeff.	**t**	**Coeff.**	**t**	**Coeff.**	**t**	**Coeff.**	**t**
Constant	8.5401	302.49	-1.2853	-17.45	-	-	-0.2602	-4.86
Delay	-0.1146	-3.52	-0.2189	-2.31	-	-	-	-
Education (Reference: Dropouts)								
Vocational degree	0.1580	5.96	0.5877	8.81	-	-	-	-
High school graduates (grade 12)	0.1426	8.06	0.2167	4.93	-	-	-	-
Two years of college (grade 14)	0.1685	12.21	0.2347	6.33	-	-	-	-
Four years of college (grade 16)	0.1624	11.12	0.3172	5.86	-	-	-	-
Graduate studies	0.2580	13.58	0.0446	1.08	-	-	-	-
Father's occupation								
Farmer	-0.0590	-2.83	0.0182	0.26	0.1913	2.58	-0.0894	-1.19
Craftsman	0.0015	0.14	0.0658	1.89	0.0272	0.64	-0.0433	-0.95
Executive	0.0047	0.34	-0.0988	-2.47	0.5049	11.95	0.0887	1.97
Middle Manager	0.0220	1.83	-0.0416	-1.12	0.3124	7.37	0.0742	1.67
White Collar. **Reference group**								
Blue Collar	0.0083	0.91	-0.0131	-0.51	-0.1403	-4.20	-0.0237	-0.66
Missing or Deceased	-0.0087	-0.68	-0.0866	-2.46	0.0374	0.82	0.0988	2.11
Mother's occupation								
Farmer	-0.0431	-2.14	0.0336	0.45	0.0379	0.48	-0.0777	-1.00
Craftsman	0.0297	2.04	0.0350	0.74	0.0668	1.14	0.0349	0.58
Executive	0.0160	1.20	0.0203	0.46	0.2371	4.69	0.1155	2.25
Middle Manager	0.0141	0.97	0.0585	1.24	0.1612	2.98	0.1502	2.84
White Collar. **Reference group**								
Blue Collar	0.0063	0.60	0.0279	1.02	-0.1868	-5.04	-0.0629	-1.58
Missing or Deceased	0.0036	0.43	0.0273	1.13	0.0566	1.96	0.1070	3.56
Father's education								
High school dropouts. **Reference group**								
Vocational degree	0.0001	0.01	-0.0541	-1.91	0.2847	8.58	-0.0017	-0.05
Advanced vocational degree	0.0013	0.12	-0.0612	-2.10	0.1958	5.56	-0.0565	-1.49
High school graduates	0.0100	0.61	-0.1187	-2.48	0.5255	10.22	0.0506	0.93
Father went to College	0.0214	1.23	-0.1576	-3.00	0.7349	14.40	0.1219	2.26
Mother's education								
High school dropouts. **Reference group**								
Vocational degree	-0.0160	-1.61	-0.0258	-0.94	0.2767	8.92	-0.0181	-0.54
Advanced vocational degree	-0.0094	-0.82	-0.0052	-0.16	0.2167	5.50	-0.0305	-0.72
High school graduates	-0.0257	-1.78	-0.0616	-1.37	0.3811	8.21	-0.0778	-1.55
Mother went to College	-0.0157	-0.77	-0.1635	-2.66	0.4349	7.21	-0.2045	-3.36
Population aged 15-19, 1982	0.0189	2.86	0.0374	1.90	-0.0244	-0.93	-0.0444	-1.65
Local population density 1982	0.0048	0.59	-0.0557	-2.31	0.2215	8.43	0.1542	5.70
Unemployment 92-97	-0.0125	-2.28	-0.0620	-3.44	-	-	-	-
Location in 1992								
Paris Area in 1992 (Reference: rest of France)	0.0735	8.51	0.0588	1.80	-	-	-	-
Unemployment 82-87	-	-	-	-	-0.0477	-1.96	0.1157	4.81
Location at grade 6 entry								
Reference (rest of France)								
Paris Area at grade 6 entry	-	-	-	-	0.2918	7.17	0.1955	4.84
Distance to college	-	-	-	-	-0.0326	-1.04	-0.0067	-0.24
(Distance to college)^2	-	-	-	-	-0.0766	-2.62	-0.0734	-2.72
Stock of vocational high schools 1982	-	-	-	-	0.2938	10.94	0.1796	6.74
Δ stock of vocational high schools 1989-82	-	-	-	-	0.1733	7.05	0.0896	3.92
Ordered Probit Cuts								
c1					34.3373	10.75		
c2					11.4594	8.30		
c3					17.1376	11.10		
c4					11.7871	9.93		
c5					22.6732	8.78		
Estimated Standard Deviations								
standard deviation	0.3194	11.61	0.9188	18.64	1	-	1.4140	235.67
Estimated Correlation Matrix	Mean Wage		Employment		Education		Delay	
Mean Wage	1							
Unemployment	0.3893 (5.27)		1					
Education	-0.2245 (-3.73)		-0.2686 (-4.91)		1			
Delay	0.5563 (5.67)		0.3181 (2.43)		-0.0301 (-3.27)		1	
Mean Log-Likelihood	-4.57751							
Quong Vuong's test (against model B)	1.11							