

Does Speed Signal Ability?

The Impact of Grade Retention on Wages*

Thomas O. Brodaty[†], Robert J. Gary-Bobo[‡] and Ana Prieto[§]

13 February 2013

Abstract

In a study of education and wages, we define *delay* as the residual of a regression of school-leaving age on the worker's highest degree. Delay is the part of school-leaving age that is not explained by the highest degree. Variability in delay is mainly due to grade retention. Making use of various instruments, we find a robust, significant and negative impact of delay on wages. A year of delay causes a decrease of the student's beginning-of-career wage around 9%, while at the same time, returns to education are positive with values also around 9%. We show that the assumption of fully informed employers is not compatible with this effect. The only reasonable explanation, supported by the data, is the fact that longer delays signal unobserved but negative characteristics to the employers.

*We thank Denis Fougere, Francis Kramarz, Guy Laroque, Edwin Leuven, Bentley MacLeod, Thierry Magnac, Eric Maurin and Jean-Marc Robin for their help and remarks at various stages of this research. We also thank Christian Belzil, organizer of a conference at IZA, Bonn, spring 2007; Kjell Salvanes, Steve Machin, Steve Pischke, Gilles St Paul and Uwe Sunde, organizers of CEPR conferences held respectively in Bergen, Norway, Madrid, Spain, in 2007, and Ammersee, Germany, in 2008; the audiences of these meetings; the members of the Microdata network, as well as seminar audiences in Amsterdam, Leuven, Munich, Paris and Toulouse for a number of very useful remarks. The present manuscript is a deeply revised version of a paper entitled "Does Speed Signal Ability? The Impact of Grade Repetitions on Employment and Wages".

[†]THEMA, University of Cergy-Pontoise, France. email: brodaty@u-cergy.fr

[‡]CREST, ENSAE, 15, boulevard Gabriel Péri, 92245 Malakoff cedex, France. email: robert.gary-bobo@ensae.fr

[§]CNRS.

1 Introduction

Education systems vary in the efficiency and precision of their sorting properties. Degrees convey more or less information to the labour markets, depending on a number of institutional details. In some countries, educational organizations have been designed to play the role of public certification agencies; in these systems, the practice of grade retention is very common: a student is promoted to the next grade only if his or her test results are sufficiently good; students who can't pass are tracked.¹ In contrast, in more egalitarian societies, or in economies promoting mass education, social promotion — i.e., the practice of passing students to the next grade, regardless of their school performance — is dominant.² It seems that education systems may experience historical swings between retention-based designs and social promotion principles. This is clearly the case in the US, where the recent development of grade retention appears to be a by-product of school accountability policies. In many countries, grade repetition practices are hotly debated. There is a concern that the negative statistical results obtained in a number of studies are due to unobservable student selection problems. Recent contributions, proposing an identification strategy that takes care of grade-repetition endogeneity, show that there are some beneficial effects of retention, at least in the short run.³ This literature is discussed further below. Different outcomes can of course be considered, such as self-esteem, health status and crime rates, but labour market outcomes, if not the ultimate judges of the efficiency of grade repetitions, are clearly some of the most important. In the following, we propose a study of the effect of grade repetitions on employment and wages.

The main contribution of the present paper is a causal study, *i.e.*, an IV estimation of the impact of grade repetitions on wages. To the best of our knowledge, our results had not been found before ⁴ (see our discussion of the literature below). Our test is based on an extension of the Mincer equation. 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.

¹France and Germany are good instances of these systems.

²Scandinavian countries and the UK are typical instances.

³See, for instance, Jacob and Lefgren (2004), Neal and Whitmore-Shanzenbach (2010), for an evaluation of the 1996 reforms of the Chicago Public Schools, that ended social promotion.

⁴The closest attempt is due to Eide and Showalter (2001) who studied the impact of grade repetitions on wages in the US. They propose some IV estimates of the effects of grade retention on log-earnings, using the child's birthday and between-state variations in kindergarten entry dates to construct an instrument. They find a positive, but imprecisely estimated and non-significant impact of grade repetitions.

Both variables are instrumented. There is a substantial amount of individual variability in time to degree-completion, conditional on the highest certificate or diploma earned by the student. This variability is mainly due to grade repetitions. We define *delay* as the difference between the individual's school-leaving age and the average school-leaving age of students holding the same highest certificate or degree. Delay is the part of age that is not explained by the highest degree. We then show that *delay* has a negative impact on wages at the beginning of a worker's career. We find that an additional year of education, leading to a higher certificate, yields a 9% increase in wages, on average, while an additional year of delay causes a 9% decrease of the student's wage, also on average⁵. Employers therefore undo what the students try to achieve by repeating grades or retaking exams. Students with a positive amount of delay are perceived as less able by the job market, *ceteris paribus*. It follows that the individuals who decide to repeat a grade or resit an exam a year later, gain nothing on average. It can nevertheless be rational for students to invest one more year in education, trying to make up for a failure, since the investment will *ex post* be profitable for those lying above the regression line, roughly half of the population. If we take into account the utility of degrees in terms of self-esteem and social status, the bet is worth taking for many students, *ex ante*. It is however dubious that the social value of grade repetitions is positive.

Both education and delay may be signals in the sense of Spence and are potentially endogenous in the wage equation. Thus, two instruments at least are needed to identify and consistently estimate their coefficients by means of IV methods. To this end, we show that various instruments can be used, but we mainly employ past school-opening instruments, and distance-to-the-nearest-college, measured when students were entering grade 6. For instance, employers are assumed not to observe variations in the local ratios of schools to school-age population that are shown to have an impact on school-leaving age. Using 3SLS, we then find a robust, significant and negative impact of the delay variable on wages, averaged over the first five years of career. A number of checks show that the result is fairly robust.

The impact of delay is far from negligible and stems from the fact that it conveys information about the unobserved ability of young workers. Thus, another contribution of this paper is to propose a test for the presence of incomplete information in wage formation, or for

⁵To 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 student *delay*. *Education* and *delay* happen to be orthogonal by construction, because delay in our sense is the residual of a regression of school-leaving age on education.

signaling in the sense of Spence (1973), based on the delay variable. Spence’s signaling hypothesis is often opposed to human-capital theories although in fact, one should interpret the signaling view as encompassing the complete information view of wages, *i.e.*, ‘Becker’s’ view (on this point, see, Riley 2001, Weiss 1995, Lange and Topel 2006). If employers imperfectly observe the relevant characteristics of job applicants, a number of observable characteristics can become signals in the sense of Spence and play a role in the determination of wages. In contrast, traditional human capital theories are valid if these informational aspects of wage formation are of limited importance in practice, because employers observe many relevant traits of the employees and learn quickly about their productive characteristics. But the signaling hypothesis is obviously not incompatible with the fact that education increases a student’s future productivity. This is why we view Spence’s theory as an extension of Becker’s theory under conditions of incomplete information, instead of taking it only as a synonym for education as pure screening.⁶ Our claim is that signaling in the sense of Spence is the only likely reason for which delay has a significant and negative impact in the wage equation.

To estimate the models, we used a very rich sample of 12,310 male students, drawn from a survey⁷ 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, and to compute the delay variable.

We rely on several sources of variation to identify the causal effects of education and delay. 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). 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)). These instruments have been criticized for various reasons,

⁶Our approach borrows elements from the recent literature on employer learning (*e.g.*, Farber and Gibbons (1996), Altonji and Pierret (2001), Lange (2007), Arcidiacono, Bayer and Hizmo (2010)). 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. Thus, we have a good chance of finding a signaling effect before it fades out, precisely because our data cover the first years of career of young workers.

⁷ The survey is called *Génération 92*, and produced by a French state-sponsored institution called CEREQ. See section 2 below for details.

and notably because mobility in geographical space is likely to be endogenous. In the present research, 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).⁸ We tried various possibilities, but the instruments exploiting vocational high-school supply variations happened to work better than others.⁹ In some robustness checks, we also use the student’s month of birth, and claim that it is an admissible instrument for delay¹⁰.

These assumed sources of exogenous variation pass the tests of overidentifying restrictions and of weak instruments. These tests are crucial in our context because the question of knowing which variables are observed by employers is ultimately empirical. If the employers did observe our instruments, or variables correlated with our instruments, this knowledge would be reflected in wages, and the overidentifying restrictions of our model would be rejected. At the same time, of course, instruments must be strong, and given the multidimensional nature of the instrumentation problem, we need to do more than simply applying the usual rules of thumb. We thus provide an application of the weak IV tests proposed by Stock and Yogo (2005).

Relationships with the literature

The relevance of degrees (as opposed to years of schooling) has been discussed in the literature on returns to education. A number of authors have identified “sheepskin effects”.¹¹ Degree holders tend to obtain higher wages than the workers with the same number of years of education who failed to pass the final exams. These authors have commented the possible connection of their results with the signaling view, but, either they didn’t really prove that these results can only be attributed to signaling, or they didn’t treat the endogeneity problem.

⁸ 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, the *Base Centrale des Etablissements*.

⁹ 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.

¹⁰ We show below that the month-of-birth variable passes the test of weak instruments in the sense of Stock and Yogo (2005), because it is a good instrument for delay (not for education).

¹¹See, *e.g.*, Hungerford and Solon (1987), Groot and Oosterbeek (1994), Jaeger and Page (1996), Belman and Heywood (1997)

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 view¹². Grade repetitions in primary and secondary school are frequent in some countries and absent in some others; see, Paul (1997). As noted above, in the recent years, grade retention policies became more common in the US. On grade retention in the US, see, *e.g.*, Eide and Showalter (2001), Jacob and Lefgren (2004), Dong (2010). Grade repetitions are also frequent in developing countries (see, *e.g.*, Gomes-Neto and Hanushek 1994). European data are studied by Fertig (2004), Mahjoub (2007), Alet (2010). These recent contributions find short-run positive effects of grade repetition on test scores. 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 France¹³. Our delay variable is the result of an addition of these sources of age variation.

The consequences of incomplete information for labor markets have been explored in various ways¹⁴, but a handful of contributions only have proposed empirical tests of Spence's theory, since the mid-seventies. Empirical tests of Spence's theory are difficult to construct because signaling and human-capital theories, inspired by Becker (1964) and Mincer (1974), predict the same increasing relationship between education and wages. Early attempts are, for instance, due to Wolpin (1977), Riley (1979), Albrecht¹⁵ (1981) and Weiss (1988). Some identification¹⁶ strategies rely on shocks affecting the supply or demand of education; see, *e.g.*, Lang and Kropp (1986), Bedard (2001), Chevalier *et al.* (2004), Hämäläinen and Uusitalo (2008). Closer to the approach proposed below are strategies testing for the presence

¹² 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. 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.

¹³ In this country, 45% of grade 9 male students had already accumulated a year of delay at least in 2002.

¹⁴ On the empirical assessment of adverse selection problems, see, for instance, Gibbons and Katz (1991), Foster and Rosenzweig (1993), Shaw and Lazear (2008).

¹⁵ See also Albrecht and van Ours (2006).

¹⁶ 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 Fang (2006).

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 credential has a signaling value on US labor markets¹⁷.

In the following, Section 2 is devoted to the linear model and to our test of the signaling hypothesis; we discuss the conditions of validity of our instruments and the weak instruments problem, given assumptions relative to the employer’s and the econometrician’s information. Section 3 describes the data and instruments. Section 4 presents estimation results and various robustness checks. We conclude in Section 5. In the appendix, we propose an analysis of the sources of OLS bias in our model, and discuss the possible role of measurement errors. The weak-instruments test results are also presented. We refer to a working paper version of the present article (*i.e.*, Brodaty *et al.*, 2010) for many details on the data, robustness checks, additional estimation results, etc.

2 The Linear Model and Test of the Signaling Hypothesis

We first present our testing strategy for the presence of job market signaling. The test relies on an extended log-wage equation, suggested by economic theory. The log-wage equation can be embedded in a linear system of simultaneous equations, by the addition of auxiliary regressions explaining education and delay. To perform the test, we jointly estimate the log-wage and linear auxiliary equations, using 3SLS, with the help of several instruments. In principle, in its simplest form, the test can be implemented if at least two valid instruments for education and delay can be found.

Following Farber and Gibbons (1996), we distinguish four 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 the student’s highest degree and time-to-degree but are not reported in the employee’s CV).¹⁸

¹⁷The GED (General Educational Development) 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).

¹⁸In the work of Farber and Gibbons (1996), the role of category-(iv) variables is played by the Armed

2.1 Derivation of the Model

Let s_i denote individual i 's education level and let d_i denote 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. We now drop index i to simplify notation.

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

$$\ln(q) = a_0s + b_0\delta + Xc_0 + \tilde{\theta}_1 + \tilde{\theta}_2, \quad (1)$$

where X is a vector of covariates observed by both the employers and the econometrician, $\tilde{\theta}_1$ is an ability factor observed by the employer, but not by the econometrician, and $\tilde{\theta}_2$ 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. Parameters c_0 are the coefficients on control variables X , observed by employers, students and the econometrician. The list of controls includes a measure of experience, if needed. The productivity-enhancing effect of education is denoted a_0 ; it is in general nonnegative. The *direct* effect of delay on productivity is denoted b_0 . In theory, if experience and education are measured precisely, this latter parameter should be zero. Note that a nonzero b_0 would capture effects that are not already explained by $X, s, \tilde{\theta}_1, \tilde{\theta}_2$. We allow for the possibility that b_0 is nonzero, because delay might capture some aspects of the person's maturity. Delay being a measure of experience, its coefficient would then be nonnegative. We will return to this point later.

It is reasonable to assume that some relevant characteristics of a job applicant are not directly observed by the employer, at least at the beginning of a career. Accordingly, we assume that employers are incompletely informed: they observe only $(X, s, \delta, \tilde{\theta}_1)$. This is enough to generate the possibility of signaling à la Spence. We precisely define "Spence's hypothesis" as the assumption that employers do not observe $\tilde{\theta}_2$. In contrast, we define "Becker's hypothesis" as the assumption that employers observe $(X, s, \delta, \tilde{\theta}_1)$ and $\tilde{\theta}_2$, *i.e.*,

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, or do not observe individual traits that are correlated with the instruments.

under the latter assumption, employers observe all the productivity-relevant characteristics of applicants.

Standard economic theory, as well as signaling theories, suggest that wages w are equal to the expected productivity of employees conditional on employers' information, that is,

$$w = E[q \mid X, s, \delta, \tilde{\theta}_1].$$

Now, assuming that our random factors and variables are normally distributed, $\ln(q)$ conditional on $(X, s, \delta, \tilde{\theta}_1)$ is also normal. Using the well-known formula for the expectation of a log-normal random variable, we thus have,

$$w = \exp\{E(\ln(q) \mid I) + (1/2)Var(\ln(q) \mid I)\},$$

where by definition, $I = (X, s, \delta, \tilde{\theta}_1)$. A special property of normal vectors is that the conditional variance $Var(\ln(q) \mid I)$ doesn't depend on the value of I ; it can therefore be treated as a constant, included in X . Thus, we get the equation,

$$\ln(w) = a_0s + b_0\delta + Xc_0 + \tilde{\theta}_1 + E(\tilde{\theta}_2 \mid I). \quad (2)$$

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

$$E(\tilde{\theta}_2 \mid I) = a_1s + b_1\delta + Xc_1 + f_1\tilde{\theta}_1, \quad (3)$$

where a_1 , b_1 , c_1 and f_1 are theoretical regression coefficients. Substituting this result in the log-wage equation, we finally get the model,

$$\ln(w) = as + b\delta + Xc + \theta_1, \quad (4)$$

where to simplify notation, we define, $a = a_0 + a_1$, $b = b_0 + b_1$, $c = c_0 + c_1$, and $\theta_1 = (1 + f_1)\tilde{\theta}_1$. We obtain a two-dimensional signaling model. Delay appears in the log-wage equation mainly because it is a signal that conveys information about the hidden talent factor $\tilde{\theta}_2$. Coefficient b_1 is the signalling effect of delay. 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_1 . It is not possible to identify a_0 and a_1 (or b_0 and b_1) separately without making strong additional assumptions, but if we can find a consistent estimator of $b = b_0 + b_1$, and if we are ready to assume that the pure productivity effect of delay

b_0 is nonnegative, we can easily test if b_1 is significantly negative. Of course, a full-fledged signaling model should specify the students' preferences and private information. This is done below, in sub-section 2.3. But at this point, we have enough elements to present the testing strategy.

2.2 Testing Signaling vs. Complete Information

Technically, we assume that some variables Z are exogenous sources of variation for s and δ , assumed uncorrelated with the θ s. Variables Z shift the student's costs of education and at the same time, are not correlated with variables observed by the employers, but that the econometrician doesn't observe. To be precise, if we assume

$$\text{plim}\left(\frac{1}{N} \sum_{i=1}^N Z_i' \tilde{\theta}_{i1}\right) = 0, \quad (5)$$

and if Z has at least two components, Z is a valid vector of instruments for s and δ , under Spence's hypothesis. Note that (5) doesn't involve $\tilde{\theta}_2$. 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 applicant's face — and therefore, that vector Z is neither included in the employer's information set, nor correlated with characteristics observed by the employer but that the econometrician doesn't observe. In addition, under Spence's hypothesis, if Z happened to be correlated with $\tilde{\theta}_2$, since $\tilde{\theta}_2$ is not observed by employers, it is not reflected in the wages, and the wage equation can still be estimated consistently with the help of Z . In the application below, we used more than 2 instruments and could thus use the test of overidentifying restrictions to show that none of the chosen instruments can significantly explain the wage equation's residual.

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 $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are fully observable by the employers (but not by the econometrician). Under Becker's hypothesis, we thus have,

$$\ln(w) = a_0 s + b_0 \delta + X c_0 + \tilde{\theta}_1 + \tilde{\theta}_2. \quad (6)$$

If we estimate the regression

$$\ln(w) = a s + b \delta + X c + \nu,$$

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 strengthen our hypothesis and assume in addition,

$$\text{p lim}\left(\frac{1}{N} \sum_{i=1}^N Z_i' \tilde{\theta}_{i2}\right) = 0, \tag{7}$$

then a test of “Spence” versus “Becker” is at hand, using our linear model.

Under Spence’s hypothesis, the IV estimates $(\hat{a}_{IV}, \hat{b}_{IV})$ of the coefficients on s and δ in the log-wage equation are consistent and we find,

$$\text{p lim}(\hat{a}_{IV}) = a_0 + a_1, \quad \text{p lim}(\hat{b}_{IV}) = b_0 + b_1.$$

In contrast, under Becker’s hypothesis, the same IV estimates are consistent but satisfy $\text{plim}(\hat{a}_{IV}, \hat{b}_{IV}) = (a_0, b_0)$. If Becker’s hypothesis holds, we strongly expect to find $\hat{b}_{IV} \geq 0$ 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 $b_1 < 0$ under the prior assumption that $b_0 \geq 0$. We thus have a test of the presence of signaling in the sense of Spence, conditional on the assumption that b_0 is nonnegative.

Note at this point that, when we reject “Becker”, we do not reject the fact that education has real productivity-enhancing effects, that is, $a_0 > 0$. Our formulation of Spence’s model is not incompatible with the fact that education really increases productivity, and we do not test this latter assumption. We only reject the assumption that employers are “fully informed”, in the restricted sense that they observe and price any relevant characteristic of job applicants.

To sum up, under Becker’s hypothesis, employers are supposed to observe all the information relevant for productivity and in this case, the valid instruments are variables that have no direct impact on productivity, *i.e.*, affect productivity only through “inputs” *that the econometrician also observes*, such as degrees.¹⁹ Under Spence’s hypothesis, employers are incompletely informed, therefore, valid instruments can be chosen in a larger set, including the variables that are valid IVs under Becker’s hypothesis, plus variables that employers do not observe, even if they are correlated with productivity-relevant characteristics, provided

¹⁹ Valid IVs should of course not be correlated with things that the employers observe and that the econometrician doesn’t observe.

that the latter characteristics are not observed by employers. In this latter case, variables not observed by employers will not be directly priced in the wage equation, and more precisely, will not be reflected in the residual of the wage equation.

In practice, a difficulty stems from the fact that these requirements can only be met with a reasonable degree of approximation. Perfect information only means substantially more information than in the incomplete information case, and perfect instruments for education probably do not exist. However, we believe that the variables used below, that are mainly describing the student’s geographical environment in early childhood, are likely to be valid IVs for this model. The validity of our assumption that Z is not correlated with things observed by employers, and that we do not observe, is ultimately an empirical question. If Z cannot explain the residuals of the wage equation, then Z cannot be in the employer’s information sets and is irrelevant for the pricing of labor.²⁰

Given that we find that $\widehat{b}_{IV} < 0$, the most plausible interpretation of our results is that a form of Spence’s signaling effect is present in the data. Delay signals negative productivity characteristics to the employer. Our estimates would be compatible with a full-information view only if delay “destroyed” productivity (because then b_0 would be significant and negative). Could this for instance be due to the fact that higher delay in fact measures a form of lower “quality” of the diplomas and certificates? No, because this interpretation contradicts the full-information assumption according to which employers observe all the productivity-relevant aspects of quality directly. In other words, under “Becker’s view,” employers do not need to use delay as a proxy for something that they are supposed to observe directly. Thus, if there are no measurement errors, the only consistent interpretation of the presence of the delay variable in the log-wage equation, under employer full-information, is to view it as a measure of experience, or a measure of maturity. We know from the literature that potential experience (*i.e.* “maturity”) and partial degree completion²¹ (*i.e.*, “credits”) have a positive return. 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. There

²⁰ It is possible to dispense with the assumption that Z is not observed by employers, provided that Z is irrelevant in a certain sense. We can admit that employers observe Z if (i), Z has no direct impact on productivity q , and (ii), Z doesn’t contribute to the explanation of θ_2 , given X, s, δ, θ_1 ; in other words, if the regression function $E(\theta_2 | X, s, \delta, \theta_1, Z)$ doesn’t depend on Z . Finally, if delay is not observed by the employers, then, it is easy to check that we must have $p \lim(\widehat{b}_{IV}) = 0$.

²¹ See, *e.g.*, Cameron and Heckman (1993), Kane and Rouse (1995).

exists a form of measurement error if the education level observed by the econometricians is not the same as that observed by the employers. But if we assume that measurement errors are negligible at least, we have a test of “Becker vs Spence”.

Therefore, we must study the consequences of measurement errors for our testing strategy: this is done in details below. For the time being, we continue to reason as if the econometrician and the employer observed the same measure of education. We now specify the auxiliary equations that can be appended to the log-wage equation. The discussion of this extended model will shed light on the empirical results, and allow for a more precise assessment of the weak instruments problem in our framework.

2.3 The complete signaling model: the potential weak instruments problem

Assume now that students make rational educational investment decisions based on a cost of education function $C(d, s, X, Z)$, where d is school-leaving age (or years of schooling), X , Z and s are as defined above. There are at least two “cost-shifters” with differing impacts on students and we can partition the vector Z as $Z = (Z_1, Z_2)$. We assume that,

$$C(d, s, X, Z) = \frac{b_2 d^2}{2} + s Z_1 a_2 + X c_2. \quad (8)$$

The cost is mainly an increasing and convex function of the time spent in the educational system and of family-background controls. In addition there are some specific interaction effects of Z_1 with s : the cost is smaller when, say, Z_1 is higher, and all the more since ambitions s are higher. Typically, we will use past vocational high-school openings as Z_1 . There could be other interactions and other impacts on C but they would not add anything essential. An important point is that duration d is random, from the point of view of students and employers. We assume the following,

$$d = \mu s + Z_2 \lambda + X c_3 + \theta_3, \quad (9)$$

where $\mu > 0$, $\lambda \neq 0$ and θ_3 is a random factor, potentially correlated with θ_1 and θ_2 .

The crucial assumption is that employers do not observe Z , but that Z is observed by students and the econometrician. Typically, we have in mind that²² $\mu \simeq 1$.

²² Indeed, with our data, when we estimate the above equation while taking the possible endogeneity of s into account (using in fact Z_1 as an instrument), we find that $\mu \simeq 1.03$, not significantly different from 1.

From this we can now derive delay as employers see it. We assumed that they take delay δ to be the part of age d that is not explained by the highest degree²³, *i.e.*, $\delta = d - E(d|s)$. Under normality assumptions, we can express the regression of d on s as

$$E(d|s) = \kappa_0 + (\mu + \kappa)s, \quad (10)$$

where κ_0 and $(\mu + \kappa)$ are theoretical regression parameters²⁴. From this we derive the delay equation,

$$\delta = -\kappa_0 - \kappa s + Z_2\lambda + Xc_3 + \theta_3. \quad (11)$$

Students are assumed to observe X , Z , and a random factor $\tilde{\theta}_0$. But they do not observe $\tilde{\theta}_1$, $\tilde{\theta}_2$ and θ_3 . Let the student's information set be $J = (X, Z, \tilde{\theta}_0)$. From the students' point of view, school-leaving age and wages are random and therefore risky. Wages as predicted by the student are

$$E[\ln(w) | J, s] = as + bE(\delta | J, s) + Xc + E(\theta_1 | J, s).$$

It follows that students take the signalling effect of delay into account in their education decision, through the term $bE(\delta | J, s)$. Assume that they choose education s so as to maximize,

$$E \left[\frac{\ln(w)}{r} - C(d, s, X, Z) \mid J, s \right], \quad (12)$$

where r is some discount rate. Using the above specifications, it follows that, equivalently, students are assumed to solve,

$$\max_s \{ (1/r)E(\ln(w) | J, s) - sZ_1a_2 - Xc_2 - (b_2/2)[E(d | J, s)^2 + Var(d | J, s)] \}.$$

Now, under normality, $Var(d | J, s)$ doesn't depend on (J, s) and $E(d | J, s)$ is linear with respect to (J, s) . Thus, the first-order conditions for an optimal schooling investment s are of the form,

$$\nu_0 - rZ_1a_2 = \nu_1E(d | J, s), \quad (13)$$

²³Instead of the above definition of delay, we could have assumed that $\delta = d - E(d|s, X, \theta_1)$, since employers observe (d, s, X, θ_1) by assumption. This more sophisticated representation would not change anything essential.

²⁴With our data, if we regress d on s using OLS, we find a very precise estimate of $\mu + \kappa$, equal to 0.7: it follows that $\kappa \simeq -0.3$. Note that κ is the OLS bias of the coefficient on s , due to the fact that employers do not observe Z .

an expression in which ν_0 is a function of a_0, a_1, b_0, b_1 , etc, and ν_1 is a function of r, b_2, μ , etc. It follows that the optimal education level s is a linear function of $(X, Z, \tilde{\theta}_0)$. For some reduced-form parameters h_0, h_1, h_2 , we have the first-stage equation,

$$s = Xh_0 + Z_1h_1 + Z_2h_2 + \theta_0, \quad (14)$$

where to simplify notation, we define, $\theta_0 = h_3\tilde{\theta}_0$, and h_3 is the coefficient of $\tilde{\theta}_0$ in the reduced-form equation for s .

We write delay in reduced form as follows,

$$\delta = Xg_0 + Z_1g_1 + Z_2g_2 + g_3\theta_0 + \theta_3, \quad (15)$$

and $g_0 = c_3 - \kappa h_0$, $g_1 = -\kappa h_1$, $g_2 = \lambda - \kappa h_2$, etc.

Now, we see some potential pitfalls more clearly. The above derivation of the reduced-form equations for s and δ shows that there is a built-in risk of a weak-instruments problem, because s and δ are nearly explained by the same linear combination of the instruments if $\lambda \simeq 0$. This is a real danger, but we use Stock and Yogo's (2005) approach to test for weak instruments, and show that in fact we can find sets of more than two exclusions that are fairly strong, and at the same time pass the test of overidentifying restrictions well.

In Appendix A, we examine the potential problems posed by measurement error. We show that a form of error in the measurement of education (*i.e.*, degrees) can explain the sign of the OLS biases that we find with the model. The main conclusions of this discussion is that (i) the coefficient on delay b can be estimated consistently by linear IV estimators, even in the presence of a measurement error on the education variable; (ii) returns to education are underestimated by OLS while the coefficient on delay is potentially strongly overestimated. Indeed, by means of OLS, we find a small and positive b while, in fact, b is unambiguously negative.

3 Data and Instruments

To perform the estimations presented below we used ‘‘G eneration 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 Education²⁵. G eneration 92 is a sample of 26, 359

²⁵ Articles and descriptive statistics, concerning various aspects of the survey, are available at www.cereq.fr.

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 1st and December²⁶ 31st, 1992. They have left the educational system for the first time, and for at least one year in 1992²⁷. The labor market experience of these individuals is 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 at 11), and in 1992, when they left school. The personal labor-market history of each survey respondent was 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.

3.1 Education levels

To explore the impact of degrees in a linear model, we constructed a synthetic schooling variable, dubbed *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 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;²⁸ (iii) those who passed the national high-school diploma, i.e., the *baccalauréat*,²⁹ have a normal age of 18; (iv) two years of college³⁰ correspond to a normal age of 20, and so on. In the linear model studied below,

²⁶ 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.

²⁷ 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.

²⁸ i.e., the so-called Certificats d’Aptitude and Brevet d’Etudes Professionnelles.

²⁹ 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.

³⁰ 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

this education variable is used instead of the years-of-schooling measure of human capital. We also estimated a nonlinear model, in which education levels are indicated by dummy variables and education choices are given by a discrete choice, Ordered Probit structure, to check for robustness of our results.³¹

3.2 Wages and returns to experience

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 different wage variables with the help of the data. We first purged the wage observations from the effect of potential experience, or alternatively, from the effects of effective experience. For each individual, we observe the first wage, the last wage and the average of observed real wages; all of them are purged from the impact of experience (either potential or effective) using a within regression.³² This yields different wage statistics.

Let w denote the real wage in January 1992 French Francs (using the French consumer price index as a deflator). Let $y = \ln(w)$. Let x denote the measure of experience. We use the standard regression function,

$$y_{it} = \alpha x_{it} + \beta x_{it}^2 + \gamma_i + u_{it},$$

where i indexes individuals and t indexes employment spells. We then estimated the returns to both forms of experience by means of a panel-data *within* estimator of α and β , to take care of the problems due to fixed effects³³. Using our sample, and with potential experience, this yields very precisely estimated values $\hat{\alpha} = 0.066$ (with a t -statistic equal to 23.0) and

the end of each of the college years in French universities, and the DEUG or DUT correspond to the end of grade 14.

³¹ The results of the ordered probit version of the model are presented in the long version of this paper, Brodaty *et al.* (2010).

³² Potential experience is defined as the time elapsed between the month of 1992 during which the individual left the education system (*i.e.*, the school-leaving date) and the wage observation, measured in months, and divided by 12. Effective experience is defined as the number of months of employment between the school-leaving date and the date at which the wage has been observed, divided by 12. Note that we focused on full-time employment spells only.

³³ The within estimator provides us with consistent estimates of α and β , even if the individual-specific effects γ_i are correlated with experience x_{it} .

$\widehat{\beta} = 0.0013$ (with a t -statistic equal to 2.67). These are returns per year of experience. We then reconstruct a purged log(mean-wage) variable y as follows,

$$y_i = \ln \left(\sum_{t \in \text{Spells}(i)} \pi_{it} \exp(y_{it} - \widehat{\alpha}x_{it} - \widehat{\beta}x_{it}^2) \right),$$

where t runs in the set of observed employment spells of individual i and the weight π_{it} is the length of spell t divided by the total duration of the period during which individual i is observed (in months). In other words, our mean wage statistic is the arithmetic average of the full-time real wages earned during full-time employment spells, weighted by their respective spell durations, net of the returns to experience. For each individual, we can also easily define the *first wage*, also corrected for the effects of experience. For descriptive statistics and further details, see the working-paper version of this article, Brodaty *et al.* (2010).

3.3 Delay and grade repetitions

A substantial part of the variance of school-leaving age, conditional on education level or degrees, happens to be due to *repeated grades*. Grade repeaters are quite common, even in college. 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 i 's highest degree is also the highest (and who left school with that degree).³⁴ 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 20.78.³⁵ 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

³⁴ 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.

³⁵ School-leaving age can in fact be measured in months, and then converted back into years (in real numbers). 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 integers, *i.e.* in years, and the results are very similar.

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 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. Figures 1a-1c provide various representations of the distribution of delay for males. Figure 1a is the plain non-parametric estimate of the density. Figure 1b plots the densities, conditional on father education; Figure 1c shows the density of delay, conditional on student education. The overall impression is that the distribution of delay is quite stable and doesn’t depend on education.

3.4 Environment and variations in geographical space

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.³⁶ We know the geographical location of the student’s family at the age of junior high-school entry 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 controls are based on data with a geographical structure. Using a file from the National Geographical Institute (*Institut Géographique 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*, which is the Euclidean distance between the *commune* of residence at the age of entry into grade 6 and the *nearest* college³⁷ (*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.³⁸

A number of other variables are based on inter-county variation, where by county we mean the French *département*.³⁹ With the help of Census and regional macro data from the National Statistical Institute (*INSEE*), we constructed a county-level share of

³⁶Are also observed, for each individual: the number of sisters, the number of brothers, the rank among siblings (*i.e.*, birth order).

³⁷This distance is the Euclidean distance between the two points on the map of France, in kilometers.

³⁸This data source also yields a measure of local population density, which we used as an additional control.

³⁹There are 95 *départements* in France. Communes are a much finer territorial division. So the distance-to-college variable is close to being individual-specific.

population aged 15-19 in 1982, that is used as a control. We constructed an average county-level unemployment rate 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, of the *lycées professionnels* and *lycées techniques*). Interestingly, there is a substantial degree of inter-county variability of the stock of vocational high-schools, *per capita* of 15-to-19-year-olds. We use this variable as an instrument.⁴⁰ 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 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 used 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. There is a

⁴⁰In 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. For more details and figures related to this instrument, see Brodaty *et al.* (2010).

substantial amount of inter-county variation in these stock variations in the sample.⁴¹

In some variants of the model, we used alternative sets of instruments. Our main alternative to school-opening instruments is the individual’s month of birth, which is a good instrument for delay. In some variants we also use a mother-at-home dummy, and the number of siblings as instruments, but they play a secondary role and are mainly used as a robustness check.

4 Estimation Results and Robustness Checks

We thus estimated the following system of three simultaneous linear equations,

$$\begin{aligned}\ln(w_i) &= as_i + b\delta_i + X_i c + \nu_i, \\ s_i &= X_i h_1 + Z_i h_2 + \epsilon_i, \\ \delta_i &= X_i g_1 + Z_i g_2 + \eta_i.\end{aligned}$$

Formally, (a, b, c) , (h_1, h_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 $(\nu_i, \eta_i, \epsilon_i)$ is a vector of random disturbances, with covariance matrix Ω . We keep in mind that the random terms $(\nu_i, \eta_i, \epsilon_i)$ are correlated because they are functions of the correlated ability terms θ . The first equation is the log-wage equation and the last two are, respectively, the reduced forms of the education and delay equations described above.

A first crucial test is to check whether the estimated coefficients $(\widehat{a}_{IV}, \widehat{b}_{IV})$ are significantly different from zero. A second crucial test is to check whether \widehat{b}_{IV} is negative, which, according to our theory, indicates that signaling is taking place. To identify (a, b) , we only need two exclusions from the wage equation, in other words we need two instruments at least. We can use distance-to-college and school opening variables to do this job, or alternatively, use the individual’s month of birth and some other past environmental, or family-background variable. 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. We will show that various combinations of instruments lead to the same conclusions. The model and its variants have been estimated by means of 3SLS (three-stages least squares).

⁴¹The 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; see Brodaty *et al.* (2010) for details.

The education variable s and the delay variable δ have a remarkable property: their empirical covariance is zero.⁴² Let $\{1, \dots, N\}$ 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, the average school-leaving age of students with education level $s = s_i$ can be computed as follows:

$$\tau(s_i) = \frac{1}{N(s_i)} \sum_{j \in B(s_i)} d_j, \quad (16)$$

and of course, $\tau(s_i) = \tau(s_k)$ for all k in $B(s_i)$. Remark that, according to our definition, delay $\delta = d - \tau$ 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 .⁴³

4.1 First stage. Strength and 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 some instruments. Table 1 gives excerpts of the results for a benchmark specification of our linear model.⁴⁴ Education and delay equations have the same specification. The third and fourth columns of Table 1 also present an alternate specification of the first stage, called *variant*. The controls in both versions of the first stage are: mother and father occupation dummies; mother and father education dummies; the population aged 15 to 19 in the county of residence at grade 6 entry, measured in 1982; the population density in 1982 in the town of residence at grade 6 entry.

The instruments for the benchmark are: distance to college at grade 6 entry; the same 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

⁴²For 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 in Brodaty *et al.* (2010).

⁴³Remark that $\text{plim } \tau(s) = E(d \mid s)$. It is also easy to check that $\text{cov}(\delta, s) = E[(d - E(d \mid s))s] = 0$.

⁴⁴The detailed results are presented in Table F1, Appendix F, of the long version of this article; see, Brodaty *et al.* (2010).

grade 6 entry. The first two columns of Table 1 show that, with the exception of distance-to-college squared in the education equation, all the instruments are highly significant (standard errors are reported below the coefficients, between brackets).

As expected, distance-to-college has a negative impact on education and the *Stock* and $\Delta Stock$ of vocational schools clearly increase education. A higher stock and a smaller distance mean more opportunities to study and smaller costs of education, so the signs of the corresponding coefficients are easily interpretable. Distance-to-college and *Stock* variables influence education in the usual way, these instruments being cost-shifters. To understand why these variables also affect delay and with the same sign, we need to remember that the delay equation is a reduced-form: in fact, delay is a function of education itself, as shown in our discussion of theory in Section 2 above. We know from a regression of delay on education and controls that the IV-estimate of the impact of education on delay is positive and significant, around 0.3. As a consequence, there is no simple story that can be told to explain the sign of the instruments' coefficients in the delay equation. The underlying theory even shows that the sign of these coefficients is ambiguous. But since instruments affect delay for the same reasons that they affect education, there is a potential weak instrument problem. We show in Appendix B that our first-stage specifications pass Stock and Yogo's (2005) demanding test for weak instruments. We also found that the *Stock* and $\Delta Stock$ instruments have a relatively balanced effect in sub-samples, while distance-to-college seems to affect mostly students from highly educated families.⁴⁵

Are these instruments truly exogenous? In that respect, it seems crucial that our instruments pass the tests of over-identifying restrictions very well, but some more fundamental arguments must be put forward. The validity of an instrument must be assessed with respect to the quality of the controls introduced in the equations of interest (the wage equation 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 control for residence in the Paris region, both at the age of grade 6 entry, and the end of studies⁴⁶. We

⁴⁵Appendix C in Brodaty *et al.* (2010) gives further details on the properties of the benchmark instruments; in particular, it shows the impact of instruments in various sub-samples.

⁴⁶We have tried to add more geographical dummies as controls to purge wages and employment rates from regional variations, but this has not proved very useful. The most important and significant geographical

also control the wage equation for county population aged 15 to 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.

The distance-to-college instruments 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. In the present study, given the relatively rich set of controls, the fact that distance-to-college is measured at the age of grade 6 entry (not when students finish high school or during higher education of course) is a real advantage. 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 opportunities⁴⁷. Conditional on family background, our distance-to-college instrument reflects the existence of pre-determined variations in geographical distance, inducing variations in the costs of education. Given the French context and the quality of our data, these variations are not likely to be correlated with individual traits that employers can observe during the hiring process, and that the econometrician does not observe.

The arguments that one can put forward in favor of the per capita stock of vocational high schools in the county (*i.e.*, *Stock*) 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 and in the same 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 may be weak, but not necessarily invalid. Again, given the many controls, *it is not likely that Stock is correlated with individual characteristics observed by the employer, reflected in wages, but unobserved by the econometrician*. For instance, suppose that two similar blue-collar families have sons with similar observable traits who may have faced different environments in terms of local school availability when their respective sons were aged 10; this had an influence on the sons' education and delay. It is likely that the two sons will present the same observable profile to potential employers *except* for education and delay. It is ultimately an empirical

indicator is that of the Paris region, versus the rest of France.

⁴⁷ France is 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. See our discussion of mobility below. In addition, 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.

question to decide if employers observe (and take into account) individual traits that are correlated with our instruments during the hiring process, but if this was the case, these instruments would not pass the test of overidentifying restrictions. In other words, if any one of the chosen benchmark IVs was correlated with the residual of the wage equation, our exclusions would be rejected by the overidentifying restrictions test, meaning that employers do in fact observe something that is correlated with the exclusions. But we will see clearly below that the benchmark exclusions are not rejected⁴⁸. In other words, it could happen that the *Stock* instrument is correlated with unobserved characteristics of the student, but if these latter characteristics are not observed by employers (*i.e.*, if they contribute to the $\tilde{\theta}_2$ factor in Section 2 above), they will not be reflected in starting wages. We also assume that regional variations in 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.

To summarize this discussion, our crucial assumption is that employers, when hiring young workers, do not observe or do not take into account the instruments or the individual characteristics of job applicants that are correlated with these instruments and that the econometrician does not observe. Employers may observe applicant traits that are not observed by the econometrician, but we suppose that these traits are not correlated with our instruments. Given that our IVs are minor aspects of the student’s environment recorded 10 years before the end of schooling, we think that this assumption is reasonable. In any case, it is not rejected by the data.

We can finally use a robustness argument: our results survive substantial changes in the set of IVs. This is why we present many variants. One of these variants is presented in Table 1; it is only an instance of a series of model variants that pass the tests reasonably well, lead to the same main conclusions and do not rely on school-opening instruments.

The first-stage variant presented in the last two columns of Table 1 has three main differences with the benchmark. First, it dispenses with *Stock*, $\Delta Stock$ and *Paris at grade 6*

⁴⁸ The time variation of the per capita stock of vocational high schools between two adequately chosen points in time (*i.e.* $\Delta Stock$) 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 speed of development of the vocational school system which could be due to variations in the policies and lobbying activities of local governments, but are most probably due to random local events. The variability of $\Delta Stock$ 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.

entry. Second, it uses a delay variable measured in months (and divided by 12), whereas the benchmark uses delay measured in integers (*i.e.*, “in years”). To be more precise, the variant of delay is based on the exact month of the year during which the student has left school. We will see below that delay in years and its variant measured in months/12 lead to essentially the same results. The second important difference is the use of a month-of-birth instrument that has a specific impact on delay. Month-of-birth measures a student’s relative maturity in his (early) classes: it takes the value 12 for individuals born in January, is equal to 11 for those who were born in February, and so on. A substantial body of literature has discussed the various reasons for which these variables have an impact⁴⁹, starting with Angrist and Krueger (1991). In the present setting, it is very clear that month-of-birth has an impact on delay, but not on education. The coefficient that we find measures a reduced-form impact⁵⁰. The other IVs used are the mother-at-home dummy, indicating a mother who doesn’t work, and the number of siblings. These family-background instruments can be criticized, on the grounds that family contributions to human capital should be reflected by wages, but in practice, they are fairly strong, and (as seen below) are not rejected by overidentification tests.

4.2 3SLS estimation of the benchmark linear model

We have estimated the linear three-equations model specified and defined above as the benchmark, along with 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, viewed as robustness checks⁵¹.

⁴⁹ On this topic, see, *e.g.*, Plug (2001), Bedard and Dhuey (2006).

⁵⁰ In France, the month of birth has complex effects on educational achievement. It has effects on the duration of schooling through elementary school and pre-school enrollment rules and other effects that are potentially more important. Recent work shows that among other things, a younger relative age increases the probability of grade repetitions (see, Mahjoub 2008, Grenet 2008). The economists’ increasing awareness of the weak instruments problem has been amplified by the discussion on the validity of the quarter of birth as an instrument for education (see, Bound *et al.* 1995, Bound and Jaeger 2000). We find here that the month of birth is a strong instrument for delay and we confirm that it is a weak instrument for education.

⁵¹ The full benchmark results are given in Brodaty *et al.* (2010), Appendix F, Table F1. The benchmark 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. Benchmark instruments are: distance to college at grade 6 entry; distance to college squared, $Stock$, $\Delta Stock$; the indicator of residence in the Paris region at grade 6 entry, the county unemployment rate averaged over years 1982-1987.

Table 2 gives the IV estimates of a and b in the log-wage equations, obtained by means of 3SLS, in two different specifications, with three different wage statistics. The standard errors are given in parentheses, and the coefficients are expressed in percentage (*i.e.*, multiplied by 100). Column A reports the results of a very crude model, estimated without any controls, while column B reports the results of the benchmark itself, with the full set of controls. Each group of two columns corresponds to a given wage statistic: the first two columns report results obtained with the mean wage corrected for potential experience, the middle columns report results obtained with mean wage corrected for effective experience and the rightmost columns correspond to the first wage corrected for potential experience.

3SLS estimates of the coefficient on delay are strongly negative and significant at the 1% level (at least), and the results are very stable from one column to the other. In a column of type B, a year of delay causes a decrease of the mean wage, during the first five years of career, around 9%. At the same time, the effect of education on wages is standard, returns to a year of education are also around 9%. Given that delay and education are orthogonal regressors, these results cannot be due to some form of multicollinearity. 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 limited, giving the impression that family-background controls do not play a major role, but in fact they do, because instruments are rejected when these controls are not included. This is clear if we look at the p -value of the over-identification restrictions F -test. The p -value of the overidentifying restrictions test jumps from 10^{-4} to more than 0.5 when we add the controls, which is very striking. In column A, we reject the fact that the instruments are valid exclusions from the wage equation, while in contrast, in column B, the p -value being above 50%, we really cannot reject the exclusions. This confirms that the judgment on our instruments strongly depends on the quality of our controls. From a purely statistical point of view at least, none of our benchmark instruments is rejected by the overidentifying restrictions test.⁵²

These are our main findings: delay has a negative impact on wages because it signals negative characteristics that employers do not directly observe. We will now present a number of robustness checks. We first changed the list of instruments and then also changed the outcomes: we considered alternative definitions of delay and checked for the impact of

⁵²We also estimated an employment equation. Delay significantly reduces the employment rate in the first five years of career. Results are presented in Brodaty *et al.* (2010, Appendix E).

experience and mobility. We finally estimated an ordered probit version of the model in which education levels are endogenous dummies, to free ourselves from the education variable.

4.3 Robustness check: variants

Table 3a shows the results of a number of variants. The top lines of Table 3a give the 3SLS estimates of coefficients of *delay* and *education* in the mean-wage equation (standard errors are in parentheses). For each variant, we indicate the list of additional controls (added in the wage, education and delay equations) and a check-list of instruments, showing which instruments have been selected (and are excluded from the wage equation). The first column restates the results of the benchmark. It is striking that the results are qualitatively close, the coefficient on delay being always negative and significant, between 6% and 13%. 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 coefficient b . Variant 3 is just Variant 2 with the number of brothers and number of sisters used as additional instruments. 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). In any case, these variants pass the test of over-identifying restrictions very well. But it is not true that any variant would pass this test, as shown by variant 5. Variant 5 is the benchmark with parental education dummies as an additional set of IVs: this variant excludes too 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%). In general, we see that the negative impact of delay on wages is a robust result.

Table 3a also shows that the past local unemployment rate is crucial in passing the weak instruments tests: we cannot reject instrument weakness in variants 2, 3 and 4, because the Cragg-Donald statistic is too low (see Appendix B for details on this test). One or two asterisks indicate that we reject weakness based on a 2SLS bias greater than respectively 10% and 5% of the OLS bias. One or two circles are the corresponding signs for rejection of the Stock and Yogo test based on a maximal size of a Wald test of 2SLS coefficients of respectively 10 and 15% (see, Stock and Yogo 2005).

The study of variants is pursued in Table 3b. Variants 6 to 12 show that we can dispense with distance-to-college, residence in the Paris area and remove one of the *Stock* IVs without losing the main results. Variant 8 is a particularly good compromise: a significant and negative impact of delay, good values of the overidentification F -test, and strong rejection of weak IVs. It seems that only the *Stock* and $\Delta Stock$ play an important role in the significance of coefficient b . 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. Variant 11 is just identified, using *Stock* and past local unemployment only: the effect of delay is still estimated with precision in the wage equation, and we strongly reject weak IVs.

Finally, Table 3c displays the results of some variants based on delay measured “in months/12” (*i.e.*, with fractions of years) instead of delay measured in years (*i.e.*, integer values). Again, we typically find a negative impact of delay, albeit smaller in absolute value, around -4.5% and remarkably good values of the test statistics. Variant 15 proves that we can dispense with the past local unemployment IV if we use month-of-birth as an alternative instrument. The variant presented in Table 1 above is the first stage of Variant 16 in Table 3c.

We conclude from the inspection of variants in Tables 3a, 3b and 3c that our results are quite robust and do not seem to depend too much on a particular instrument.

4.4 Further robustness checks. Other outcomes, alternative definitions of delay and mobility

We have made further robustness checks, the results of which can only be summarized here. Again, the interested reader will find a wealth of details in the long working-paper version

of this article; see, Brodaty *et al.* (2010). The questions that we ask are the following: is the effect of delay still significant and negative when (i), we change the definition of the dependent variables and (ii), if we control for the experience accumulated before the recorded school-leaving time? Are the results still true if we change the way of measuring delay? Is mobility in geographical space a source of bias?

Alternate specification: Instead of using the mean wage statistic, we can use the first wage or the last wage observed in the 5-year observation period, net of the effects of potential experience after the school-leaving date. We also control for the time spent in summer jobs and internships before the school-leaving date.⁵³ Again, we find that the impact of delay is stable, significant and negative, across these regressions. The specific measure of pre-school-leaving experience, *i.e.*, time spent in summer jobs and internships, has a positive and significant coefficient but doesn't change the crucial coefficients a and b . We conclude from these tests that experience can increase wages while at the same time, delay still reduces wages.

Employment equation: We also estimated an employment equation in which the employment variable is the log of the employment rate, *i.e.*, the ratio of the number of month in employment, divided by the total number of months observed. The employment variable can be treated exactly in the same way as the log-wage, as the dependent variable. According to our theory, students with a positive delay would find a job less easily, being turned down more often by employers, and this would be reflected in the employment rate. It is indeed what we find: the impact of delay on employment is negative and significant.⁵⁴

Alternate definitions of delay: One could also suspect that delay is job search in disguise. So we have changed its definition 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 normal number of years needed for the high-school diploma, but an additional year of delay is added because he has spent one year in college without earning a higher degree.⁵⁵ In the alternate specification, the delay variable has been changed

⁵³Brodaty *et al.* (2010: Table D1, Appendix D) shows the results of these regressions.

⁵⁴Brodaty *et al.* (2010: Table E1, Appendix E) presents the results on the employment rate equation.

⁵⁵In other words, the standard 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.

so that the months spent in school without producing any new diploma are added to the first job-search spell. 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.⁵⁶

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 results show that we find a negative coefficient if delay is replaced with age at grade 6 entry, with delay at grade 6 entry or with delay accumulated after grade 6 entry, although only delay accumulated after grade 6 is truly significant and negative (with a coefficient also around 9%).⁵⁷

Mobility in geographical space: During education years, mobility in geographical space may be endogenous and thus, become a source of bias. In the benchmark model, we use an indicator of residence in the Paris region at two different points in time: 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 the wage equation for residence in the Paris region and local unemployment rates at the age of labor-market entry, but endogenous mobility could perturb the estimates. Note that we estimated variants of the model in which these additional location variables were not used, and in which the impact of delay on wages was negative and significant. We are thus confident that our main result doesn't crucially depend on some problem due to endogenous location choices.

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 impact of delay is still present and negative, of the same order of magnitude (yet smaller), but less significant. Then we defined 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. The *Move* indicator is added as a

⁵⁶Brodaty *et al.* (2010: Table D2, Appendix D) shows the results of these regressions.

⁵⁷Brodaty *et al.* (2010: Table D2, Appendix D) presents the results of these regressions.

control in the linear model's four equations, and we see that it has a positive and significant coefficient: those who moved earn 7% more on average. 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 the coefficients on delay are stable and remain very significant around 9%.⁵⁸

Ordered Probit: Finally, we have replaced the education variable with a set of dummy variables indicating a position on a scale of degrees. This is to free ourselves from the conventions used in the construction of the education variable, based on the notion of normal age. These dummies are endogenous in the wage equation, so we treated the education equation as an Ordered Probit, and estimated it jointly with the delay and the wage equations by Maximum Likelihood, taking care of possible correlations between all error terms.⁵⁹ We find a strong negative impact of delay on wages. The results are essentially similar to those obtained by 3SLS.⁶⁰

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 instruments, with variants of the endogenous variables, and do not depend on the introduction of additional controls for experience and mobility.

4.5 OLS estimates and OLS bias

It is interesting to compare the results obtained with IV estimators (the 3SLS estimates) with the naive OLS estimates of the wage equation and provide an interpretation for the sign of OLS biases on delay and education variables. The first column of Table 4 gives the OLS estimates of a and b : we see that returns to education are slightly underestimated (with a return of 6% per year) while the impact of delay is strongly biased upwards, being close to zero and significant. These findings confirm the analysis proposed in Appendix A,

⁵⁸To measure mobility more finely, we also 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. Controlling for *Distance* doesn't perturb the main results, and *Distance* is significant in the wage equation. For details, see Brodaty *et al.* (2010: Appendix D, Table D3).

⁵⁹This approach is reminiscent of Cameron and Heckman's (1998) Ordered Probit model of education choices.

⁶⁰The complete estimation results are given in Brodaty *et al.* (2010: Table F2, Appendix F).

in which it is suggested that if degrees are measured with error, returns to education are likely to be biased towards zero, that is, underestimated by OLS, while the impact of delay is pushed upwards. We have shown above that, in addition to these sources of bias, the usual “ability bias”, that is, the positive correlation of error terms in the wage and education equations can also contribute to an upward bias of coefficient b . Thus, empirical results do not contradict the analysis of Appendix A. Table 5 shows the correlation matrix of the random vector (ν, ϵ, η) obtained as a side-product of the 3SLS procedure. The sign pattern of the correlations is in accordance with the OLS biases obtained.

Table 4 presents other aspects of our results. Columns 2 and 3 are respectively OLS and 3SLS estimates of the wage equation in which school-leaving age d is the only endogenous regressor (delay and education being removed from the right-hand side). We then find returns to education around 6%, slightly underestimated by OLS. These results are in line with the bulk of the literature on returns to education. Note that our instruments are rejected (the p-value of the overidentification test is very low). This is presumably because delay has been omitted from the right-hand side of the regression. Columns 4 and 5 presents a similar exercise, in which delay is omitted and our education variable is the only endogenous regressor. We then find that returns to education are very similar, also slightly underestimated by OLS, and instruments are rejected. The most interesting variant is when the education variable is replaced with school-leaving age in the wage equation, as in columns 6-7. We then find that returns to education are around 8%, as usual, but the OLS estimate of the coefficient on delay is negative and significant. In these last regressions, it is striking to see that delay has a negative and significant coefficient, even if the model is estimated by OLS. The 3SLS estimates show the same pattern for the OLS bias: returns to education are underestimated and the coefficient on delay is overestimated (or underestimated in absolute value). The estimated value of b seems too large in absolute value, but we cannot reject the instruments in this latter case. These results also show the robustness of our main finding.

5 Conclusion

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 various instruments and notably past school-

openings and distance to college, 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 close to zero. 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 instruments, (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 most likely explanation for these results is that the negative effect of delay on wages is a job-market signaling phenomenon. Employers being incompletely informed, delay, as defined above, conveys information about the young workers' productivity-relevant characteristics. Using an extension of the Mincer log-wage equation, we thus found a way of testing for the presence of signaling in the formation of wages. Human-capital theory under employer full-information does not predict a negative impact of delay: if the full information assumption was (nearly) true, the effect of delay on wages should be approximately zero. The negative coefficient on delay is thus related to incomplete information.

6 References

- Albrecht, James W. (1981), "A Procedure for Testing the Signaling Hypothesis", *Journal of Public Economics*, **15**, 123-132.
- Albrecht, James W. and Jan van Ours (2006), "Using Employer Hiring Behavior to Test the Educational Signaling Hypothesis", *Scandinavian Journal of Economics*, 108, 361-372.
- Alet, Elodie (2010), "Is Grade Repetition a Second Chance?", *manuscript, Toulouse School of Economics*, Toulouse, France.
- Altonji, Joseph, G. and Charles Pierret (2001), "Employer Learning and Statistical Discrimination," *Quarterly Journal of Economics*, 116, 313-350.
- Angrist, Joshua D, and Alan B. Krueger (1991), "Does Compulsory School Attendance Affect Schooling and Earnings", *Quarterly Journal of Economics*, **106**, 979-1014.

- Arcidiacono, Peter, Bayer, Patrick and Aurel Hizmo (2010), "Beyond Signaling and Human Capital: Education and the Revelation of Ability", Duke University, manuscript, forthcoming in *American Economic Journal: Applied Microeconomics*.
- 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.
- Bedard, Kelly, and Elisabeth Dhuey (2006), "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects", *Quarterly Journal of Economics*, **121**, 1437-1472.
- Belman, Dale and John S. Heywood (1997), "Sheepskin Effects by Cohort: Implications of Job Matching in a Signaling Model", *Oxford Economic Papers*, **49**, 623-637.
- Bound, John, Jaeger, David A. and Regina M. Baker (1995), "Problems with Instrumental Variables Estimation when the Correlation Between Instruments and the Endogenous Explanatory Variable is Weak", *Journal of the American Statistical Association*, **90**, 443-450.
- 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.
- Brodaty, Thomas O., Gary-Bobo, Robert J. and Ana Prieto (2010), "Does Speed Signal Ability? The Impact of Grade Retention on Wages", *CREST, INSEE*, July 2010, <http://ces.univ-paris1.fr/membre/Gary-Bobo/index.htm>
- 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 (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 (2003), "Understanding What Instrumental Variables Estimate: Estimating Marginal and Average Returns to Education", University of Chicago, manuscript.
- Chevalier, Arnaud, Colm, Harmon, Walker, Ian, and Yu Zhu (2004), "Does Education Raise Productivity, or Just Reflect It?", *Economic Journal*, **114**, F499-F517.
- 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.
- Dong, Yingying (2010), "Kept Back to Get Ahead? Kindergarten Retention and Academic Performance", *European Economic Review*, **54**, 219-236.
- 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.
- Eide, Eric R. and Mark H. Showalter (2001), "The Effect of Grade Retention on Educational and Labor Market Outcomes", *Economics of Education Review*, **20**, 563-576.
- 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.

- Fertig, Michael (2004), "Shot Across the Bow, Stigma or Selection? The Effect of Repeating a Class on Educational Attainment", *IZA Discussion Paper* no 1266, IZA, Bonn, Germany.
- 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.
- Grenet, Julien (2008), "Le mois de naissance influence-t-il les trajectoires scolaires et la vie professionnelle? Une évaluation sur données françaises", *Paris School of Economics*, manuscript.
- Groot, Wim and Hessel Oosterbeek (1994), "Earnings Effects of Different Components of Schooling: Human Capital Versus Screening", *Review of Economics and Statistics*, **76**, 317-321.
- Hämäläinen, Ulla and Roope Uusitalo (2008), "Signalling or Human Capital: Evidence from the Finnish Polytechnic School Reform", *Scandinavian Journal of Economics*, **110**, 755-775.
- Hungerford, T., and Gary Solon (1987), "Sheepskin Effects in the Returns to Education", *Review of Economics and Statistics*, **69**, 175-177.
- Jacob, Brian, A. and Lars Lefgren (2004), "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis", *Review of Economics and Statistics*, **86**, 226-244.

- 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.
- 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", chap. 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.
- Mahjoub, M. Badrane (2007), "Grade Repetition as a Treatment", *Paris School of Economics*, manuscript.
- Mincer, Jacob (1974), *Schooling, Experience, and Earnings*, Columbia University Press, New York.
- Neal, Derek, and Diane Whitmore Schanzenbach (2010), "Left Behind by Design: Proficiency Counts and Test-Based Accountability", *Review of Economics and Statistics*, **92**, 263-283.
- Paul, Jean-Jacques (1997), "Le redoublement à l'école: une maladie universelle?", *International Review of Education*, **43**, 611-627.
- Plug, Erik J. S., (2001), "Season of Birth, Schooling and Earnings," *Journal of Economic Psychology*, **22**, 641-660.
- 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.

- Shaw, Kathryn L. and Edward P. Lazear (2008), "Tenure and Output", *Labour Economics*, **15**, 705-724.
- Spence, Michael (1973), "Job Market Signalling", *Quarterly Journal of Economics*, **87**, 355-374.
- Staiger, Douglas and James H. Stock (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica*, **65**, 557-586.
- Stock, James H., and Motohiro Yogo (2005), "Testing for Weak Instruments in Linear IV Regression", in D.W.K. Andrews and J.H. Stock eds., *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas Rothenberg*, Cambridge University Press, Cambridge, U.K.
- 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.
- Weiss, Andrew (1988), "High School Graduation, Performance and Wages", *Journal of Political Economy*, **96**, 785-820.
- 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

7.1 A. Measurement error and other sources of OLS bias

We now answer potential objections related to measurement error. It is likely that education, and more specifically degrees, are measured with error. Employers probably observe a much finer description of degrees than the econometrician, even under Spence's hypothesis. Given that we aggregated some degrees, the coefficient on delay in a naive OLS regression is likely to capture the fact that delay is negatively correlated with the imperfectly observed quality or prestige of the degree. Note that delay could also capture the positive effect of unobserved partial degree completion, known to be equally nonnegligible. The sign of b would then be ambiguous, even if signaling is at work. Assume then to fix ideas that education s is not well observed: the econometrician in fact observes $s^* = s + \theta_4$, where θ_4 is a random error, while the student and the employer observe s . We assume that $E(\theta_4) = 0$, and $Cov(\theta_4, \theta_i) = 0$ for all $i \neq 4$. This error means that delay, as observed by the econometrician, is contaminated. The econometrician in fact observes δ^* where by definition, $\delta^* = d - E(d|s^*)$. Under normality again, we can express the regression of d on s^* as $E(d|s^*) = \kappa_0^* + (\mu + \kappa^*)s^*$. Using (11) and substituting $s = s^* - \theta_4$, we can rewrite δ and δ^* as follows,

$$\begin{aligned}\delta^* &= -\kappa_0^* - \kappa^*s^* + Z_2\lambda + Xc_3 + \theta_3 - \mu\theta_4 \\ \delta &= -\kappa_0 - \kappa s^* + Z_2\lambda + Xc_3 + \theta_3 + \kappa\theta_4\end{aligned}$$

Expressing δ as a function of δ^* and substituting the result in the wage equation yields the log-wage regression, as seen from the point of view of the econometrician⁶¹,

$$\ln(w) = [a + b(\kappa^* - \kappa)]s^* + b\delta^* + Xc + b(\kappa_0^* - \kappa_0) + \theta_1 + [b(\mu + \kappa) - a]\theta_4. \quad (17)$$

This clearly shows that coefficient b is still consistently estimated by means of the linear IV estimator based on Z .

We can now study the OLS biases in this model with two endogenous regressors, s and δ . Suppose to simplify the analysis that there are no controls, *i.e.*, $X = 0$, suppose that s^* is centered. We have $E(s^*) = E(\delta^*) = 0$ and $s^* \perp \delta^*$, since $d - E(d|s^*) \perp s^*$. From this

⁶¹Note that if there was also a measurement error on d , say, $d^* = d + \theta_5$ where θ_5 has a zero mean and is uncorrelated with other variables, then the appropriate definition of delay would be $\delta^{**} = d^* - E(d^*|s^*)$. But θ_5 being independent of s^* , we have $E(d^*|s^*) = E(d|s^*)$ and it follows that $\delta^{**} = \delta^* + \theta_5$. The log-wage regression would have an additional noise term $-b\theta_5$, and our discussion would be essentially the same.

we easily derive $\kappa_0^* = \kappa_0 = 0$. The model under study can now be rewritten,

$$\begin{aligned}\ln(w) &= [a + b(\kappa^* - \kappa)]s^* + b\delta^* + \nu, \\ s^* &= Z.h + \varepsilon, \\ \delta^* &= Z.(\lambda - \kappa^*h) + \eta,\end{aligned}$$

where the random error terms are,

$$\begin{aligned}\nu &= \theta_1 + [b(\mu + \kappa) - a]\theta_4 \\ \varepsilon &= \theta_0 + \theta_4 \\ \eta &= \theta_3 - (\mu + \kappa^*)\theta_4 - \kappa^*\theta_0.\end{aligned}$$

Regressors being orthogonal, the OLS estimates of the wage regression

$$\ln(w) = \alpha s^* + \beta \delta^* + \nu$$

are easy to compute. Denote for simplicity, $\sigma_i^2 = Var(\theta_i)$, $\sigma_x^2 = Var(x)$, $\sigma_{xy} = Cov(x, y)$ and $\sigma_{ij} = Cov(\theta_i, \theta_j)$, for $i, j = 0, 1, \dots, 4$. Using $Cov(Z, \theta_i) = 0$ for all i , some standard computations yield the results,

$$p \lim(\widehat{\alpha}_{OLS}) = a + b(\kappa^* - \kappa) + \frac{\sigma_{\nu\varepsilon}}{\sigma_s^2 + \sigma_4^2}, \quad (18)$$

$$p \lim(\widehat{\beta}_{OLS}) = b + \frac{\sigma_{\nu\eta}}{Var(\delta^*)}. \quad (19)$$

So, the OLS estimators of a and b are biased, as expected. To obtain a more precise view of the OLS biases as a function of measurement-error variance σ_4^2 and other correlations, we compute the values of κ^* and κ . By definition of the theoretical regression coefficients, we have,

$$\mu + \kappa = \frac{\sigma_{ds}}{\sigma_s^2}, \quad \mu + \kappa^* = \frac{Cov(d, s^*)}{\sigma_s^2 + \sigma_4^2} = \frac{\sigma_{ds}}{\sigma_s^2 + \sigma_4^2}.$$

Simple covariance computations also yield, $\sigma_{\nu\varepsilon} = \sigma_{10} + [b(\mu + \kappa) - a]\sigma_4^2$. Substituting these results in expressions (18)-(19) above, we finally obtain,

$$p \lim(\widehat{\alpha}_{OLS}) = a + \frac{\sigma_{10} - a\sigma_4^2}{\sigma_s^2 + \sigma_4^2}. \quad (20)$$

We assume that $\sigma_{10} = Cov(\theta_0, \theta_1) \geq 0$ because θ_0 and θ_1 are two different perceptions of the individual's "talent" or unobserved ability. In the absence of measurement error, the

OLS bias on returns to education a should therefore be positive: it would be the usual “ability bias”. But measurement error tends to cause a counteracting underestimation of the returns. In fact, our empirical results below show that the OLS estimate of α is typically slightly smaller than its IV estimate, so that returns to education seem to be underestimated by OLS. Measurement error is therefore a likely explanation. It is easy to check that an increase in the variance of errors σ_4^2 will typically yield a lower OLS estimate of α . This is also true, with our data, in numerical experiments. We have tried to add random errors to our measure of education and this had the effect of reducing the estimated returns to education⁶².

If we now compute the OLS bias on the delay effect b , we find,

$$\text{plim}(\widehat{\beta}_{OLS}) = b + \frac{1}{\text{Var}(\delta^*)} [\sigma_{13} - \kappa^* \sigma_{10} + (\mu + \kappa^*)(a - b(\mu + \kappa))\sigma_4^2], \quad (21)$$

where,

$$\text{Var}(\delta^*) = \text{Var}(Z \cdot (\lambda - \kappa^* h)) + \text{Var}(\theta_3 - \kappa^* \theta_0) + (\mu + \kappa^*)^2 \sigma_4^2. \quad (22)$$

From these expressions, we see that there are several sources of bias in the estimation of delay effects. It is likely that $\sigma_{13} = \text{Cov}(\theta_1, \theta_3) \leq 0$, given that θ_3 increases delay, but if maturity plays a positive role, σ_{13} may be small in absolute value; in contrast, κ^* is typically negative and converges towards $-\mu$ as σ_4^2 grows without bound. The term $-\kappa^* \sigma_{10}$ is likely to be positive and contributes to an upward bias. Measurement error has a positive impact since $\mu + \kappa^* > 0$, $b < 0$, $a > 0$ and $\mu + \kappa > 0$, so that $(a - b(\mu + \kappa)) > 0$. We conclude that b should be overestimated by OLS for various reasons, including measurement error on s . Indeed, this is what we find with the data: the OLS estimate of b is typically near zero and even slightly positive, while the IV estimate of b is significantly negative, near -9% . In numerical experiments, adding noise to the education variable s^* typically increases the OLS estimate of b . Again, we see that the measurement error explanation is consistent with the facts, but the upward bias on b may also naturally be due to σ_{10} . Finally, $\text{Var}(\delta^*)$ is numerically much smaller than $\text{Var}(s^*)$: this also contributes to a small downward bias on education and a large bias on delay.

⁶² Since we also find that $b < 0$, we have $b(\kappa^* - \kappa) > 0$, and it follows that the IV estimates of α should still overestimate the returns to education by the amount $b(\kappa^* - \kappa)$, in the presence of measurement error. Overestimation is modest if measurement error problems are not too severe, because in this case, $\kappa^* \simeq \kappa$.

7.2 B. Application of Stock and Yogo’s test for weak instruments

Consider the results of the first-stage, presented in Table 1. The F -test for the joint significance of the instruments is far above the rule-of-thumb level of 10 in both equations⁶³. In spite of this, there is a risk of instrument weakness if the same linear combination of instruments in fact explains both education and delay. So we computed the tests recommended by Stock and Yogo (2005) for weak instruments, based on the Cragg-Donald statistic. If the two endogenous variables s and δ can be nearly explained by the same combination of instruments, the Cragg-Donald statistic takes a low value, the IV estimates are biased and the standard errors of these estimates are underestimated (*i.e.*, using 2SLS, the null rejection rate based on t tests at the nominal 5% level could in fact be 10% or more). But our benchmark instruments are not weak: we reject the null hypothesis that the bias of the 2SLS coefficients is more than 10% of the OLS bias (with a risk of 5%). The value of the Cragg-Donald statistic is equal to 13.41 for the benchmark, while Stock and Yogo’s critical values are between 8 and 11 for this version of their test⁶⁴. We also reject the hypothesis that the null rejection rate of the 5% Wald test concerning the 2SLS coefficients of the first stage is in fact 15% or more (the critical value, also based on the Cragg-Donald statistic, is 12.33). This is quite reassuring, given that Stock and Yogo’s tests are demanding. We’ll see below that some variants of the model exhibit much higher values of the Cragg-Donald statistic and therefore less risk of weak IVs than the benchmark⁶⁵.

How come that our school-opening and distance instruments are not weak? To see this, we can look at the other instruments used in the benchmark specification. Residence in the Paris area at grade 6 entry has a large impact on both education and delay. As will be seen below, variants in which the Paris region dummies are not used work well: they are not essential. In contrast, it is interesting to note that past local unemployment rates *reduce* education and *increase* delay. Unemployment can have a depressing effect on

⁶³ See Staiger and Stock (1997).

⁶⁴ If the Cragg-Donald statistic is higher than some critical value, we reject the null assumption that instruments are weak. This means that with the chosen instruments, we reject the fact that they have the potential to lead to a bias of 2SLS estimates relative to OLS estimates of more than 10%, with a risk of rejecting the null wrongly of 5%. If the maximum bias of OLS estimators of returns to education is, say, 10 percentage points, and the relative 2SLS bias is 0.1, then, the maximum bias on returns to education estimated by 2SLS is one percentage point.

⁶⁵ See Brodaty *et al.* (2010), Table C3, Appendix C. This table gives the Cragg-Donald test values and Stock and Yogo’s critical values for a number of variants of the model, including the benchmark and variant of Table 1.

education through different channels. It is very likely that in regions of high unemployment, the weight of liquidity constraints is greater, because the student's parents are more likely to be poorer. At the same time, the reduced-form impact of local unemployment on delay can be understood as the result of smaller opportunity costs of education in regions of high unemployment. In these regions, there are less attractive job market opportunities and the incentives to earn a degree quickly are reduced. Note that current local unemployment is used as a control in the wage equation, for otherwise, the past local unemployment rate would be a bad instrument. The latter variable thus plays a major role in raising the Cragg-Donald statistic because it has opposite effects on education and delay: in the benchmark specification, it helps avoiding the weak instruments problem (as will be seen below in the discussion of variants).

TABLES AND FIGURES

Figure 1a: Distribution of Delay

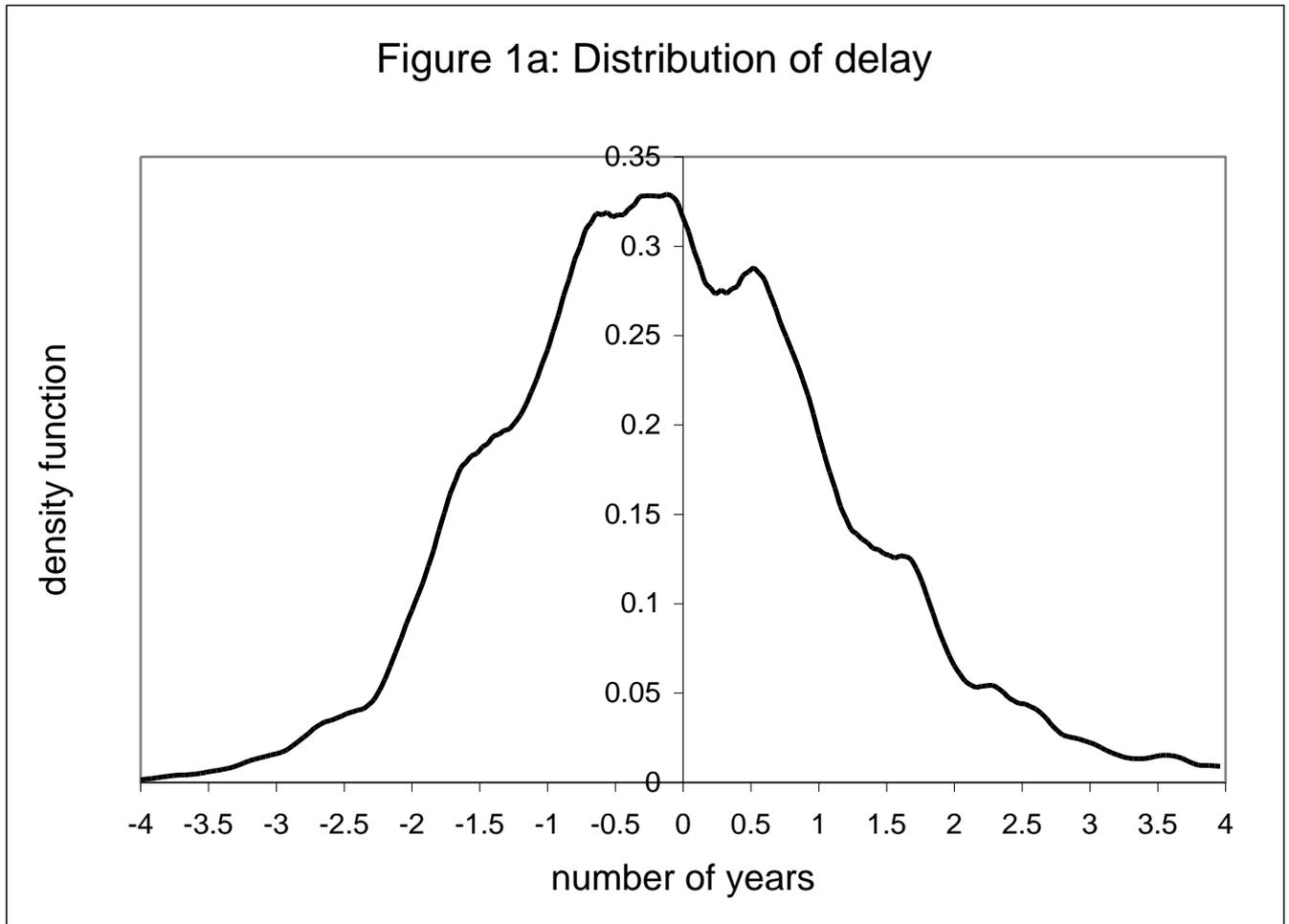


Figure 1b: Conditional Distribution of Delay (Impact of Father's Education)

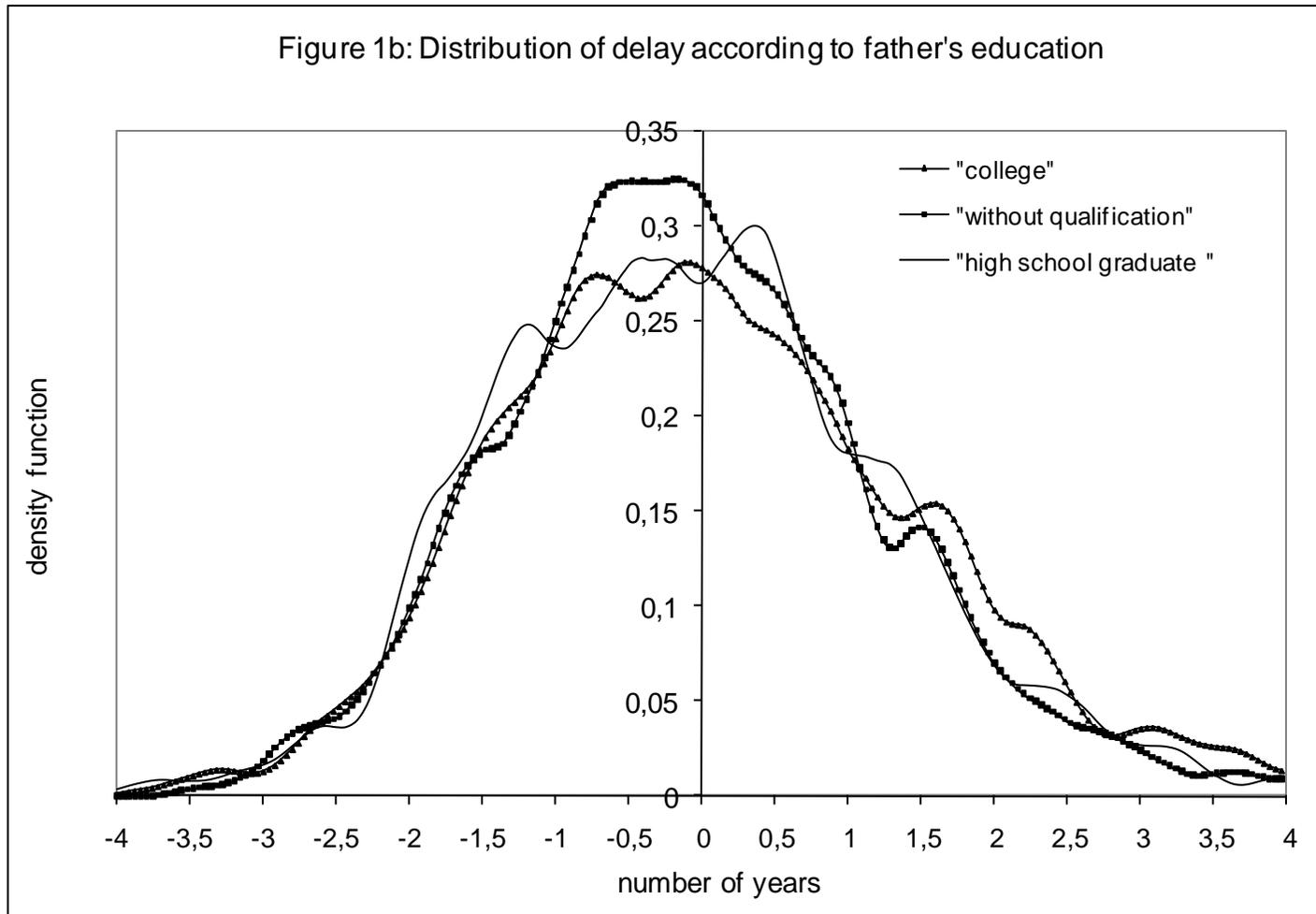


Figure 1c: Conditional Distribution of Delay (Impact of Son's Education)

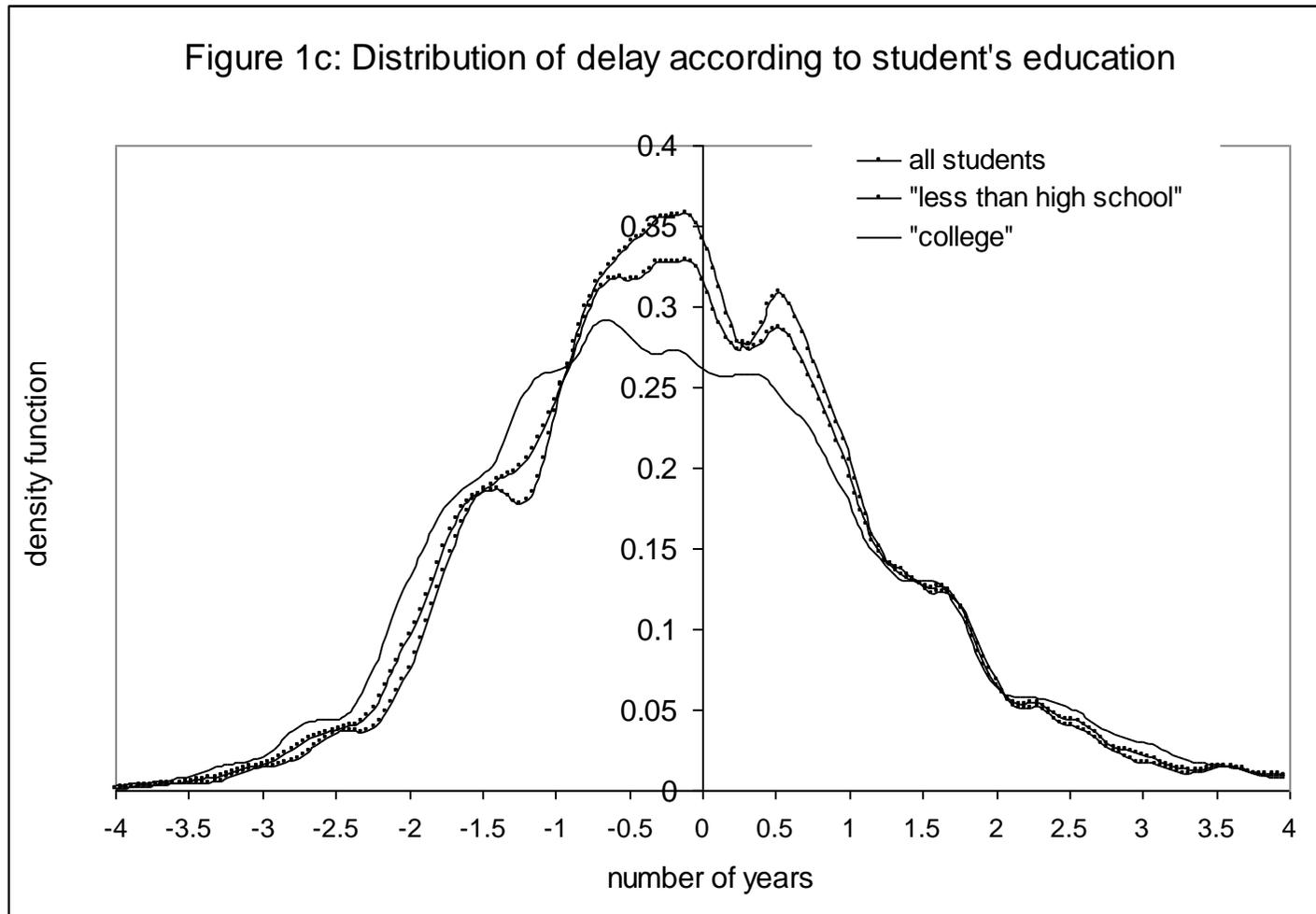


Table 1: First Stage of the Benchmark Linear Model

	Benchmark		Variant	
	Education	Delay (in years)	Education	Delay (in months)
Stock of vocational high schools 1982	0.0171*** (0.0012)	0.0048*** (0.0007)		
Δ stock of vocational schools 1989-82	0.0317*** (0.0032)	0.0074*** (0.0017)		
Paris Area at grade 6 entry	0.5368*** (0.0981)	0.2898*** (0.0532)		
Local unemployment rate before 1987	-0.0525*** (0.0143)	0.0545*** (0.0079)	-0.0431*** (0.0123)	0.0488*** (0.0065)
Month of birth			-0.0054 (0.0068)	0.0347*** (0.0036)
Mother at home			0.4632*** (0.1043)	0.0596 (0.0547)
Number of siblings			-0.1288*** (0.0159)	0.0322*** (0.0085)
Distance to college in 1982	-0.0067** (0.0031)	-0.0065*** (0.0016)	-0.0067** (0.0031)	-0.0061*** (0.0017)
(Distance to college in 1982)²	0.00005 (0.388×10 ⁻⁴)	0.00008*** (0.198×10 ⁻⁴)	0.0001* (0.290×10 ⁻⁴)	0.0001*** (0.199×10 ⁻⁴)
F test of instruments	31.2	18.06	14.37	26.48
R²	0.21	0.021	0.206	0.023
Cragg-Donald Statistic		13.41		12.11
Number of Observations			12,310	

Note: Columns 1 and 3 (resp. columns 2 and 4) are the estimated coefficients of OLS regressions of the education variable (resp. the delay variable) on the listed instruments and controls (standard errors in parentheses). The controls in both versions of the first stage are: mother and father occupation dummies; mother and father education dummies; the population aged 15 to 19 in the county of residence at grade 6 entry, measured in 1982; the population density in 1982 in the town of residence at grade 6 entry. *** (resp. ** or *) indicate significance at the 1% (resp. 5% or 10%) level. The Cragg-Donald statistic is used to test for weak instruments (see notes of Table 3b and Table C3 in appendix C for explanations).

Table 2: 3SLS Estimation of Log-Wage Equation (Benchmark)

	Mean Wage corrected for potential experience		Mean Wage corrected for effective experience		First Wage corrected for potential experience	
	A	B	A	B	A	B
Controls	No	Yes	No	Yes	No	Yes
Delay (in years)	-12.62%*** (0.0229)	-9.91%*** (0.0270)	-9.29%*** (0.0207)	-7.83%*** (0.0258)	-10.92%*** (0.0239)	-8.83%*** (0.0310)
Education	12.37%*** (0.0065)	9.26%*** (0.0084)	11.58%*** (0.0059)	8.37%*** (0.0079)	10.70%*** (0.0068)	8.35%*** (0.0095)
R ²	0.0316	0.1719	0.0353	0.1817	0.0217	0.1100
p-val. of Fisher Overid. Test	0.0001	0.6180	0.0001	0.5091	0.0001	0.8565

Note: Three-Stage Least Squares estimation of the log-wage equation for three different definitions of the dependent variable. Correction for the effect of potential (resp. effective) experience is based on a within estimation of a regression of log-wages on potential (resp. effective) experience and potential (resp. effective) experience squared. The residuals of these regressions are then averaged with weights proportional to employment spells, to compute the corrected mean-wage variables. We have checked that the results are essentially the same if the impact of experience is neglected. The last two columns on the left of Table 2 give the result when the first wage (corrected for potential experience) is used instead of a mean of wages observed during the first 5 years of career. Column A reports the results of a crude regression without any controls (note that in this case the p-value of the Fisher Over-identification test is low and we unambiguously reject the instruments). In contrast, in B columns, controls are added to the regression and the p-value of the F-test of over-identifying restrictions jumps upwards to high values, above 50%. We therefore cannot reject the instruments. The benchmark 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. Benchmark instruments are: distance to college at grade 6 entry; the same distance to college squared, Stock, Δ Stock; the indicator of residence in the Paris region at grade 6 entry, the county unemployment rate averaged over years 1982-1987. Standard errors are in parentheses. *** (resp. ** or *) indicate significance at the 1% (resp. 5% or 10%) level. Coefficients on delay and education have been expressed in percentage (*i.e.*, multiplied by 100).

Table 3a: Robustness Check Ia: Variants

		Benchmark	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5
Mean Wage purged from effects of potential experience	Delay (in years)	-9.91%*** (0.0270)	-10.29%*** (0.0289)	-12.97%** (0.0638)	-7.48%*** (0.0216)	-6.54%** (0.0311)	-7.19%*** (0.0202)
	Education	9.26%*** (0.0084)	9.79%*** (0.0098)	7.63%*** (0.0255)	5.64%*** (0.0081)	6.29%*** (0.0113)	8.29%*** (0.0035)
	p-value of overidentification F-test	0.6180	0.5831	0.5502	0.6962	0.5488	0.0001
Additional controls of equation of interest	Paris Area 1992 (1)	Yes	Yes				Yes
	Average local unemployment rate 92-97 (1)	Yes	Yes				Yes
	Family structure (16)		Yes				
	Age at grade 6 entry (4)		Yes				
Instruments	Distance to college in 1982 (2)	Yes	Yes	Yes	Yes	Yes	Yes
	Stock of vocational high schools 1982 (1)	Yes	Yes	Yes	Yes		Yes
	Δ stock of vocational schools 1989-82 (1)	Yes	Yes	Yes	Yes		Yes
	Average local unemployment rate 82-87 (1)	Yes	Yes				Yes
	Paris Area at grade 6 entry (1)	Yes	Yes				Yes
	Number of siblings (10)				Yes	Yes	
	Mother at home (1)					Yes	
	Parental education (8)						Yes
Number of instruments	6	6	4	14	13	14	
Cragg-Donald Statistic	13.41* ^o	12.26* ^o	2.11	5.08	5.00	10.48	

Note: Tables 3a-3b-3c present variants of the 3SLS estimation of the log-wage equation. Standard errors are in parentheses. Stars on coefficient estimates *** (resp. ** or *) indicate significance at the 1% (resp. 5% or 10%) level. Coefficients on delay and education have been multiplied by 100 to be expressed in percentage. Each column is a variant of the regression of log-mean wage, purged from the effects of potential experience on delay, education and controls. 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) plus some additional variables listed as “additional controls”. The list of instruments used for estimation is given in each column: ‘Yes’ indicates that the corresponding instrument(s) is (are) included in the first-stage regression (and excluded from the wage equation). The last line gives the Cragg-Donald statistic, used to test for weak instruments (see note of Table 3b for explanations).

Table 3b: Robustness Check Ib: Other Variants

		Variant 6	Variant 7	Variant 8	Variant 9	Variant 10	Variant 11	Variant 12
Mean Wage purged from effects of potential experience	Delay (in years)	-12.51%*** (0.0330)	-17.18%*** (0.0629)	-9.41%*** (0.0291)	-5.73% (0.0384)	-12.14%*** (0.0359)	-12.38%*** (0.0364)	-12.71%*** (0.0335)
	Education	8.86%*** (0.0098)	11.18%*** (0.0208)	9.11%*** (0.0085)	12.07%*** (0.0128)	8.72%*** (0.0100)	9.14%*** (0.0115)	9.32%*** (0.0112)
	p-value of overidentification F-test	0.8834	0.5321	0.3300	0.0996	0.5255	---	0.9154
Additional controls of equation of interest	Paris Area 1992 (1)		Yes	Yes	Yes			
	Average local unemployment rate 92-97 (1)	Yes		Yes	Yes	Yes	Yes	Yes
Instruments	Distance to college in 1982 (2)	Yes	Yes		Yes			Yes
	Stock of vocational high schools 1982 (1)	Yes	Yes	Yes		Yes	Yes	Yes
	Δ stock of vocational schools 1989-82 (1)	Yes	Yes	Yes		Yes		
	Average local unemployment rate 82-87 (1)	Yes		Yes	Yes	Yes	Yes	Yes
	Paris Area at grade 6 entry (1)		Yes	Yes	Yes			
Number of instruments		5	5	4	4	3	2	3
Cragg-Donald Statistic		15.68*** ^o	3.59	18.82*** ^{oo}	20.96*** ^{oo}	24.39*** ^{oo}	34.14 ^{oo}	18.23 ^{oo}

Note on the Cragg-Donald statistic: If the two endogenous variables education and delay can be nearly explained by the same combination of instruments, the Cragg-Donald statistic takes a low value, the IV estimates are biased and the standard errors of these estimates are underestimated. If the Cragg-Donald statistic is higher than some critical value, we reject the null assumption that instruments are weak. Our benchmark instruments are not weak: we reject the null hypothesis that the relative bias of the 2SLS coefficients is more than 10% of the OLS bias (with a risk of 5%). The value of the Cragg-Donald statistic is equal to 13.41 for the benchmark, while Stock and Yogo's critical values are between 9 and 15 for this version of their test. This means that with the chosen instruments, we reject the fact that they have the potential to lead to a bias of 2SLS estimates relative to OLS estimates of more than 10%, with a risk of rejecting the null wrongly of 5%. We also reject the hypothesis that the null rejection rate of the 5% Wald test concerning the 2SLS coefficients of the first stage is in fact 15% or more (the critical value, also based on the Cragg-Donald statistic, is 12.33). This is quite reassuring, given that Stock and Yogo's tests are demanding. Table C3, in Appendix C gives the Cragg-Donald test values and Stock and Yogo's critical values for the variants of Tables 3a-3c, including the benchmark and the variant of Table 1. Stars ** and * refer to the first Stock and Yogo test of weak instruments, based on 2SLS bias relative to OLS bias. Circles ^o and ^{oo} refer to the second Stock and Yogo test of weak instruments, based on the Wald test of 2SLS coefficients. One star * (resp. two stars **) means that the Cragg-Donald statistic is high enough to reject a bias of more than 10% (resp. 5%) of the OLS bias at the significance level of 5%. One circle ^o (resp. two circles ^{oo}) means that the Cragg-Donald statistic is high enough to reject the null hypothesis that the null rejection rate of the 5% Wald test concerning the 2SLS coefficients of the first stage is in fact 15% or more (resp. 10% or more).

Table 3c: Robustness Check Ic: Other Variants

		Variant 13	Variant 14	Variant 15	Variant 16
Mean Wage purged from effects of potential experience	Delay (in months)	-4.03%*** (0.0134)	-5.13%*** (0.0150)	-3.71%** (0.0177)	-3.51%** (0.0165)
	Education	7.63%*** (0.0056)	8.21%*** (0.0068)	4.09%*** (0.0100)	7.26%*** (0.0106)
	p-value of overidentification F-test	0.605	0.2021	0.1653	0.5593
Additional controls of equation of interest	Paris Area 1992 (1)	Yes	Yes		Yes
	Average local unemployment rate 92-97 (1)	Yes	Yes		Yes
Instruments	Distance to college in 1982 (2)	Yes	Yes	Yes	Yes
	Stock of vocational high schools 1982 (1)	Yes	Yes	Yes	
	Δ stock of vocational schools 1989-82 (1)	Yes	Yes	Yes	
	Average local unemployment rate 82-87 (1)	Yes	Yes		Yes
	Paris Area at grade 6 entry (1)	Yes	Yes		
	Number of siblings (1)				Yes
	Number of siblings dummies (10)	Yes			
	Month of birth (1)	Yes	Yes	Yes	Yes
Mother at home (1)	Yes			Yes	
Number of instruments		18	7	5	6
Cragg-Donald Statistic		12.29*	21.83***°	14.22***°	12.11*°

Note: See explanations in the footnotes of Tables 3a and 3b.

Table 4: OLS Biases and Variants with School-Leaving Age

Dependent Variable: Mean Wage purged from effects of potential experience							
	1	2	3	4	5	6	7
	OLS	OLS	3SLS	OLS	3SLS	OLS	3SLS
School-Leaving Age	---	5.84%*** (0.0010)	6.96%*** (0.0068)	---	---	8.17%*** (0.0012)	10.96%*** (0.0096)
Education	6.10%*** (0.0009)	---	---	6.09%*** (0.0009)	7.47%*** (0.0064)	---	---
Delay (in years)	0.08%*** (0.0016)	---	---	---	---	-6.88%*** (0.0019)	-18.18%*** (0.0302)
R ²	0.4004	0.3427	0.0212	0.3992	0.2198	0.4061	0.1894
p-val. of Fisher Overid. Test	---	---	0.0001	---	0.0052	---	0.6676

Note: This table presents regressions of the log-mean wage, corrected for the effect of potential experience on the benchmark set of controls and various subsets of endogenous variables: school-leaving age, education and delay. The first column shows the results of an OLS estimation of the wage equation: we find a small positive coefficient on delay (upward bias) and a slightly underestimated return to education (downward bias). Both biases can be explained by measurement error affecting education, as suggested in Section 2 of the paper, but there are other possible explanations. All 3SLS estimations of Table 4 rely on the benchmark instruments and controls described in Table 1 and 2 above. The rightmost columns of the table present the results obtained when education is replaced by school-leaving age in the wage equation. The OLS coefficient of delay in the next to last column is negative and significant. Standard errors are in parentheses. *** (resp. ** or *) indicate significance at the 1% (resp. 5% or 10%) level. Coefficients on delay and education have been expressed in percentage (*i.e.*, multiplied by 100).

Table 5: Correlation Matrix of Residuals (Benchmark Model)

	Mean Wage	Education	Delay
Mean Wage	1	-0.2338	0.4793
Education	-0.2338	1	-0.0279
Delay	0.4793	-0.0279	1

Note: Table 5 presents the cross-correlation matrix of the residuals of the benchmark wage, education and delay equations, estimated by 3SLS.