

DYNAMIC SKILL ACCUMULATION, COMPARATIVE ADVANTAGES,
COMPULSORY SCHOOLING, AND EARNINGS

Christian BELZIL
Jörgen HANSEN
Xingfei LIU

December 2011

Cahier n° 2011-29

DEPARTEMENT D'ÉCONOMIE

Route de Saclay
91128 PALAISEAU CEDEX
(33) 1 69333033
<http://www.economie.polytechnique.edu/>
<mailto:chantal.poujouly@polytechnique.edu>

Dynamic Skill Accumulation, Comparative Advantages, Compulsory Schooling, and Earnings*

Christian Belzil

Ecole Polytechnique, ENSAE, and IZA

Jörgen Hansen

Concordia University, CIRANO, CIREQ and IZA

Xingfei Liu

Concordia University

December 6, 2011

*This paper builds on material initially found in the working paper “The Distinction between Dictatorial and Incentive Policy Interventions and its Implications for IV Estimation”. We would like to thank James Heckman, Steve Durlauf, Chris Ferrall, Paul Gomme, Bo Honore, Nikolay Gospodinov, Susumu Imai, Winfried Koeniger, James McKinnon, Hessel Oosterbeek, Paul Devereux, Steve Pischke, Arnaud Maurel, Jean-Marc Robin, Uwe Sunde, Solomon Polachek, Chris Taber for comments and discussions related to this (or a previous) version. We also thank seminar participants at Ecole Polytechnique (France), University of Wisconsin, Queen’s University, McMaster University, IZA, Concordia University, the Malinvaud Seminar at CREST, IWAE 2010 (Catanzaro) and the Institute for Economic Analysis at Autònoma University in Barcelona. The usual disclaimer applies.

Abstract

We show that a calibrated dynamic skill accumulation model allowing for comparative advantages, can explain the weak (or negative) effects of schooling on productivity that have been recently reported (i) in the micro literature on compulsory schooling, ii) in the micro literature on estimating the distribution of ex-post returns to schooling, and (iii) in the macro literature on education and growth. The fraction of the population more efficient at producing skills in the market than in school is a pivotal quantity that determines the sign (and magnitude) of different parameters of interest. Our model reveals an interesting paradox; as low-skill jobs become more skill-enhancing (*ceteris paribus*), IV estimates of compulsory schooling become increasingly negative, and ex-post returns to schooling (inferred from a Roy model specification of the earnings equation) become negative for an increasing fraction of the population. This arises even if each possible input to skill production has a strictly positive effect. Finally, our model provides a foundation for the weak (or negative) effect education on growth measured in the empirical literature.

Key Words: *Compulsory Schooling Reforms, Dynamic Skill Accumulation, Comparative Advantages, Returns to schooling, Education and Growth, Dynamic Discrete Choice, Dynamic Programming.*

JEL Classification: I2, J1, J3.

1 Introductory Remarks

In this paper, we show that a calibrated dynamic skill accumulation (DSA) model allowing for comparative advantages, can reconcile many results reported in the vast and heterogenous literature on estimating the economic benefit of schooling.¹ In particular, it may explain the incidence of very low (possibly negative) estimates that have been reported in the recent literature on mandatory schooling (Devereux and Hart, 2010). The model also sheds light on the reasons why ex-post returns to schooling are found to be negative for a subset of the US population (Carneiro, Hansen, and Heckman, 2003, and Heckman, Humphries, Urzua and Veramendi, 2011). Finally, it indirectly provides a theoretical foundation for the very weak (or negative) correlation between economic growth and education (Bils and Klenow, 2000).²

As will become clear later, the difference between the rates at which labor market experience and education may be used to produce new skills is a pivotal quantity that drives the sign (and magnitude) of IV estimates, and other parameters measuring the effect of schooling on productivity. We show how negative IV estimates, as well as negative ex-post returns inferred from a Roy model specification of the earnings equation, may arise even if each possible input to skill formation (education and different types of labor market employment) has a strictly positive effect.

Our approach is objective. We neither address the relevance of IV estimation strategies, nor its statistical performance.³ Indeed, our approach does not even require imputing pre-estimation objectives to empirical economists using IV methods.⁴ Put differently, our objective is to show how an education

¹Throughout the paper, the terms “return to schooling”, “treatment effect of schooling on earnings”, and “economic benefit of schooling” may be used interchangeably.

²In this literature, economists attempt to capture the causal effect of education on growth rates, after controlling for cross-country differences in capital stock, and well as other factors explaining growth.

³For opposite views regarding IV estimation strategies, the reader can consult Keane (2010), Heckman, Urzua and Vytlacil (2005), Deaton (2008) and Imbens (2009).

⁴For instance, anecdotal evidence suggests that some economists interpret their estimates as a direct conditional effect of schooling for some population (thereby relying explicitly on the IV identifying moment condition), while others may believe that they

policy intervention may at the same time increase educational attainments and reduce average earnings. We use IV methods simply because it is the main tool used in the literature on compulsory schooling.

In the paper, we construct a synthetic dynamic skill accumulation (DSA) data generating process with heterogenous agents. The model is in the spirit of Ben-Porath (1967), and has similarities with other dynamic skill accumulation models analyzed in the literature (Heckman, Lochner and Taber, 1998, Keane and Wolpin, 1997). We claim that a properly constructed DSA model is the natural benchmark to analyze life cycle data on schooling and earnings.⁵

Our model has a relatively simple structure. Forward looking agents are endowed with an individual specific level of academic ability and allocate their time between schooling and work, over an exogenously given time horizon. Individuals can either work in a job that entails a high accumulation rate of human capital or in a job that offers a lower rate. We sometimes refer to the employment status characterized by a low skill-accumulation rate as the “low-skill” job. Our terminology should not be confused with the usual meaning of low-skill vs high-skill jobs, which typically refers to either a skill requirement, or to a pay level. In our model, two individuals with the same level of accumulated human capital earn the same amount. However, their incremental skill acquisition is function of the state occupied.⁶ As is usually done in the structural literature, we do not distinguish between ex-ante and ex-post returns and assume that individuals have full information about the skill accumulation technology.⁷

As in Heckman, Lochner and Taber (1998), we dissociate the earnings process from the learning process. The dynamics arise because the psychic cost of occupying the type of employment that produces more human capi-

estimate some total effect (direct and indirect effects) and ignore “identification” issues.

⁵The classical Ben-Porath model provides the structural foundation of the popular Mincer equation and has also been used in the Economic Growth literature (Lucas, 1988). As well, dynamic skill accumulation models play a central part in the recent literature on child development (Cunha, Heckman, and Schennach, 2011).

⁶Our characterization of the employment status is therefore a discrete multi-state version of the classical Ben-Porath model. The term “job” does not refer to an employer-employee relationship. Instead, it characterizes different states which entail a different investment rate.

⁷Cunha, Heckman and Navarro (2005) analyze a framework where ex(post returns may diverge from ex-ante returns.

tal depends on both accumulated human capital and intrinsic ability (skills reduce the psychic cost of acquiring new skills).⁸ Comparative advantages are introduced in the model by assuming dispersion in academic abilities (in the skill formation technology parameters). In particular, there exists a sub-population endowed with low academic ability, which is more efficient at producing skills in the market (in the low-skill job) than in school. Although it would be easy to introduce dispersion in the rate at which individuals produce new skills in a low skill job, it turns out that our finding is more clearly illustrated with such a simpler parameterization.⁹

To stress the specific implications of compulsory schooling compared to other educational interventions, we analyze three distinct types of interventions; those that stimulate schooling attainments by affecting the net utility of attending school (education subsidies), those that do it by setting a minimum school leaving age (compulsory schooling), and those that create a disincentive to invest in schooling, by subsidizing low-skill employment. All of them are anchored in the DSA model.

In the paper, we raise and answer the following questions.

- What are the implications of comparative advantages for IV estimates of the treatment effect of schooling, which are obtained from
 - changes in compulsory schooling?
 - education subsidies?
 - low-skill employment subsidies?
- Can a dynamic skill accumulation model accommodate

⁸Actually, the main results reported in the paper would also prevail in the case where schooling has no dynamic effect on post-schooling skill formation. We chose to introduce the dynamic structure in order to be realistic. Equally, the main results would also go through if we assumed, in line with the literature on formal training, that investing in training reduces current earnings.

⁹One implication of our model is that those who have comparative advantages in academic activities (those who are more educated and work in jobs that produce more skills) earn more than those who work in low skill jobs (aside from purely random shocks). However, because one of our focus will be on policies that prevent the bottom tail of the ability distribution to act based on their comparative advantages (compulsory schooling), allowing for some of the non-academic types to earn more than some of the academic types, would only reinforce our findings.

- the incidence of negative IV estimates for compulsory schooling?
- the incidence of negative (ex-post) returns to schooling for a certain fraction of the population?
- the weak effect of education on growth found in the empirical literature?

The main results of the paper may be summarized as follows. The incidence of negative IV estimates, which are generated from educational interventions targeting the bottom tail of the ability distribution, is a natural implication of the existence of comparative advantages. Essentially, IV estimates are bound to lie below zero as long as those individuals, who are more effective at producing skills in the market than in school, are also those affected by the policy. This is the case with compulsory schooling interventions, but generally not with other interventions.

The negative effect of compulsory schooling on earnings is perfectly consistent with the existence of negative ex-post returns, which have been reported in the literature on estimating distribution of returns (Carneiro, Hansen, and Heckman, 2003, and Heckman, Humphries, Urzua and Veramendi, 2011). It is also consistent with the structural dynamic programming literature on schooling and earnings (Belzil and Hansen, 2007, Belzil and Hansen, 2002, and Keane and Wolpin, 1997).¹⁰ The earnings loss displayed by compulsory schooling policies is the logical consequence of a policy that forces those who are endowed with negative returns to take actions that do not reflect their comparative advantages.

Our model reveals an interesting paradox. As we increase (counterfactually) the rate at which low-skill jobs produce new skills (leaving both the distribution of treatment effects of schooling and the return to high-skill jobs unchanged), IV estimates of compulsory schooling become increasingly negative, and ex-post returns to schooling become negative for an increasing fraction of the population. In other words, the incidence of negative IV estimates is the by-product of the difference between the rate at which skills are produced in school and in the market (in low skill jobs). Negative IV

¹⁰In the structural literature, post-schooling accumulation is usually exogenous and modeled through a Mincer type of function (quadratic in experience). Keane and Wolpin (1997) is an exception. The authors endogenize post-schooling choices by allowing individuals to move freely between blue-collar and white-collar occupations.

estimates may arise even if each possible input (education, low-skill job and high-skill job) has a strictly positive effect on skill formation.¹¹

The residual parts of the paper are organized as follows. In Section 2, we lay-out the behavioral model. The 3rd section is devoted to the construction of the treatment and control groups. In the 4th section, we implement IV methods on our benchmark model, using instruments generated from both education subsidies and compulsory schooling. In Section 5, we use various simulations to provide a structural interpretation for the incidence of negative IV estimates of compulsory schooling changes, and provide an intuitive discussion of our main results. In Section 6, we discuss the potential links between our results and different literatures. The final section offers concluding remarks.

2 The Behavioral Model

The model is in the spirit of the classical Ben-Porath model (1967), although it also has some fundamental differences. As do Heckman, Lochner and Taber (1999), we construct a model that separates the schooling decision from the post-schooling accumulation process.¹² In line with the structural dynamic schooling literature (Keane and Wolpin, 1997, Eckstein and Wolpin, 1999, Belzil and Hansen, 2002, and Heckman and Navarro, 2006), we assume that post-schooling skill acquisition is also affected by psychic components. The dynamics of skill formation arises because schooling reduces the psychic cost of learning further skills.

To choose the preference parameters we relied mostly on the structural literature, in order to obtain a realistic range of the relevant parameters

¹¹This finding discloses an obvious commonality with theoretical findings of Heckman and Vytlačil (2005), who show how misspecification of the first stage of a linear IV model, may lead to estimates completely disconnected from the treatment effects of interest. We address this issue in Section 6.

¹²In the Ben-Porath model, the cost of skill accumulation is suffered in terms of current earnings reduction. This assumption, coupled with other technical aspects (continuous choice variable, continuous time, concavity), imply several features that are typically not supported by schooling/earnings data. For instance, earnings do not jump sufficiently beyond school completion (the fraction of time spent learning declines only gradually), and earnings growth rates are highly persistent.

(when possible). Then, we simulated the model and adjusted the parameters until the final values enabled us to match the population characteristics or the population moments that we stated as desirable.

2.1 Model Structure

The baseline model is a stochastic dynamic discrete choice model of labor supply/human capital accumulation over the life-cycle. There are 50 periods to allocate between the 3 mutually exclusive states; schooling (s), work with a low rate of skill accumulation (e), and work with a high rate of skill accumulation (a). We sometimes refer to state e as the “low skill” job, thereby referring to its associated skill accumulation rate (as opposed to a skill requirement), as opposed to an employer-employee relationship.

The choices are summarized in the binary indicators, d_{tk} , where $d_{tk} = 1$ when option k (s, e, a) is chosen at date t . The variables corresponding to the capitalized letters (S_t, E_t, A_t) are used to measure the number of periods accumulated in each state. There is a maximum of 16 years of schooling attainable.

In observational data, the pendant of state e could be full time employment with learning by doing, while state (a) could represent work, with on-the-job training. The distinction between Employment (e) and Work and Training (a) is therefore in the intensity of human capital accumulation (a is the high intensity mode).

Individuals are risk neutral and maximize the expected value of lifetime net earnings, over the entire life-cycle. The state-specific utilities are defined below. As is common in the structural literature, we assume that individuals have full information about the skill accumulation technology.

2.2 School

The utility of individual i , at time t , who attends school (state s), denoted U_{it}^s , is

$$\begin{aligned}
 U_{it}^s &= \alpha_i^s + \alpha_1^s \cdot I(S_t \leq 4) + \alpha_2^s \cdot I(5 \leq S_t \leq 8) + \\
 &\alpha_3^s \cdot I(9 \leq S_t \leq 12) + \alpha_4^s \cdot I(13 \leq S_t \leq 16) + \\
 \alpha_5^s \cdot I(d_{t-1,s} &= 0) + \varepsilon_{it}^S
 \end{aligned} \tag{1}$$

where $I(\cdot)$ is the indicator function. The parameters $\alpha_1^s, \alpha_2^s, \alpha_3^s$ and α_4^s capture the variation in the utility of attending school with grade level. These parameters reflect tuition costs and the like. The parameter α_5^S captures the psychic cost of attending school for those who would have interrupted their education. The term α_i^S represents individual heterogeneity in taste for schooling (academic ability). Finally, ε_{it}^S is a purely stochastic shock.

2.3 The Dynamics of Skill Accumulation

We assume that activities generating skill formation entail some psychic cost.¹³ Individuals who accumulate skills, can do so while achieving a high level of earnings. However, to do so, they must absorb a reduction in net utility.

The utility of work and learning, U_{it}^e , depends only on the wage rate (learning on the job is costless). The utility of work and training, U_{it}^a , is defined as the difference between the wage rate and the monetary equivalent of the psychic cost, $C_{it}^a()$. U_{it}^e , U_{it}^a , and $C_{it}^a()$, are given by the following equations;

$$U_{it}^e = W_{it} \tag{2}$$

$$U_{it}^a = W_{it} - C_{it}^a(S_{it}) \tag{3}$$

$$C_{it}^a() = c_{0i}^a + c_{1a} \cdot S_{it} + \varepsilon_{it}^a \tag{4}$$

and where c_{1a} captures the effect of accumulated schooling on the cost (or disutility) of work and training. We assume that c_{1a} is a negative parameter.¹⁴ The ε_{it}^j 's are stochastic shocks.

¹³In the structural schooling choice literature (Keane and Wolpin, 1997, Eckstein and Wolpin, 1999, and Belzil and Hansen, 2002), it is well known that individual decisions can hardly be rationalized without introducing unobserved heterogeneity affecting the utility of attending school (the consumption value of schooling).

¹⁴Because c_{1a} is negative, we implicitly assume a form of complementarity between schooling and on-the-job human capital accumulation. Heckman, Lochner and Taber (1998) assume a similar learning technology.

2.4 The Earnings Equation

The earnings equation is given by the following expression:

$$\log W_{it} = w_{it} = \alpha + \lambda_i^s \cdot S_{it} + \lambda^e \cdot E_{it} + \lambda^a \cdot A_{it} + \varepsilon_{it}^w \quad (5)$$

where W_{it} represents earnings at time t , α is the intercept term, λ^s is the treatment effect of schooling on earnings, λ^e is the effect of employment, λ^a is the effect of employment with training, and ε_{it}^w is a random shock (described below). At the expense of repeating ourselves, we stress that the current gross earnings are not affected by the state occupied. However, both net earnings (actual earnings minus psychic costs) and the amount of skill investment are function of the type of job occupied.¹⁵

It is convenient to separate schooling, and post-schooling accumulation, and to re-write the wage equation as

$$w_{it} = \alpha + \lambda_i^s \cdot S_{it} + \varphi_{it}$$

where

$$\varphi_{it} = \lambda^e \cdot E_{it} + \lambda^a \cdot A_{it} + \varepsilon_{it}^w \quad (6)$$

2.5 The Bellman Equations

Given the Markovian structure of the model, the solution to the problem is obtained using recursive methods, and optimal choices may be characterized by a Bellman equation (Bellman, 1957).

For each possible choice, there is a specific value function, $V_t^k(\Omega_t)$, equal to

$$V_t^k(\Omega_t) = U_t^k + \beta E \max\{V_{t+1}^1(\Omega_{t+1}), \dots, V_{t+1}^K(\Omega_{t+1}) \mid d_{kt} = 1\}$$

or, more compactly, as

$$V_t^k(\Omega_t) = U_t^k + \beta E V_{t+1}(\Omega_{t+1} \mid d_{kt} = 1)$$

where β is the discount factor, and where Ω_t is the set containing all state variables known by the agent at t .

¹⁵A alternative approach would be to assume that accumulating skills requires formal training, and that current earnings are reduced when the high accumulation state is occupied. However, this would have no influence on the results reported in the paper.

2.6 Comparative Advantages and Random Shocks

To proceed, we follow the structural literature that estimates distributions of treatment effects of schooling (Carneiro, Hansen and Heckman, 2003, and Belzil and Hansen, 2007), and assume that the effect of schooling on log earnings is subject to cross sectional dispersion. We follow more specifically Belzil and Hansen (2007) with respect to the distribution of λ^s , because it is the only paper that allows for a distribution of treatment effects of schooling in conjunction with a separate post-schooling accumulation technology (a Mincer model).¹⁶ The distribution ranges from 0.005 to 0.12, and the population average effect is equal to 0.06.

The heterogeneity distribution, $H_{\nu_i}(\cdot)$, is specified as a multi-variate discrete distribution with R vectors of support points;

$$v_r = \{\alpha_r^s, c_{0r}^a, \lambda_r^s, p_r\} \text{ for } r = 1, 2, \dots, R \quad (7)$$

where p_r is the population proportion of type r . The full distribution is displayed in Table A1 (in appendix).

The vector $\{\varepsilon_{it}^s, \varepsilon_{it}^e, \varepsilon_{it}^a, \varepsilon_{it}^w\}$ is composed of i.i.d. mutually independent random shocks. Each one follows a Normal distribution with mean 0 and variance $\sigma(k)$ for $k = s, e, a, w$.

2.7 Model Calibration and Solution

To implement the models, we experimented with the parameters of the utility of attending school so to obtain desirable features. In the end, we use the following values: $\alpha_1^s = -3$, $\alpha_2^s = -7$, $\alpha_3^s = -12$, $\alpha_4^s = -14$, and $\alpha_5^s = -18$. The return to employment (set to 0.015) and training (set to 0.03) are chosen to reflect the fact that human capital accumulation is more intensive in state a than in state e . They also ensure that the average life-cycle earnings growth will lie between 1% and 2% per year (a well known stylized fact for the US). To introduce some dynamics in the skill accumulation process, we set c_{1a} to 0.50 (each year of schooling reduces the psychic cost of schooling by 50 cents).

In order to allow for a high degree of selectivity on persistent heterogeneity, we set the standard deviations of all random shocks to 0.35

¹⁶For instance, in Keane and Wolpin (1997), the returns to schooling are occupation specific as opposed to individual specific.

The discount factor is set to 0.95 As is relatively common in the literature, we solve the Bellman equations using simulated realizations of the random shocks, for each single type separately. Our solution method is exact to the extent that we solve value functions for each point in the state space (we do not use any approximation or interpolation methods).

3 The Control and Treatment Groups

3.1 Control Group

To generate the control group, we simulate 2,500 realizations of the full vector of random shocks for those types with density 0.05, and 5,000 for the three types that have density equal to 0.10. We therefore end up with 50 years of choices and wage outcomes for 50,000 individuals. An individual is defined as the conjunction of (i) a heterogeneity type and (ii) a specific history of random shocks. Throughout the paper, it is convenient to think of the time horizon as covering choices made between age 15 and 65.

As documented in Belzil and Hansen (2010), simulated data from the control group display all the desired features.¹⁷ Descriptive statistics of the number of periods spent in each state are found in Table A2. The average schooling attainment is equal to 6.5 years, and as normally expected, the incidence of training (high accumulation state) is smaller (6 years on average) than regular employment (37 years on average).

3.2 Treatment Groups

All of those education policies that we consider, have empirical relevance. Practically every developed country has, at one stage, implemented mandatory schooling policies. Many countries have also introduced policies that favor enrollment into higher education.¹⁸ Finally, the third policy, which consists of subsidizing low-skill jobs, raises interest for both technical and

¹⁷Those features include schooling being located in the early phase of the life-cycle, and the incidence of the intensive human capital accumulation state declining with age.

¹⁸In practice, this may be achieved by implementing tuition subsidies, scholarship funds, or simply constructing universities, colleges, and the like.

economic reasons. Because empirical labor economists practically never consider experiments that reduce schooling attainments, it is interesting to compare the low-skill employment subsidy to those interventions targeting an increase in schooling. At an economic level, an employment subsidy is particularly relevant for countries that want to reduce the “crowding out” in their university system, in favor of professional education programs closely integrated to the labor market. This is the case in Europe, where higher education is heavily subsidized.¹⁹

In order to build the treatment groups, we proceed as we did for the control groups and we simulate 50 years of choices and wage outcomes under each policy intervention. We end up with 100,000 observations (50,000 in control and 50,000 in treatment).

3.2.1 Compulsory schooling

Our definition of compulsory schooling is standard. It increases schooling in the population by setting a minimum age (period) for leaving school. Formally, a policy intervention that dictates school attendance for the first x periods, sets

$$d_{s1i} = d_{s2i} = \dots d_{sxi} = 1 \forall i$$

and implies that individuals start optimizing at date $t = x + 1$.

3.2.2 Education Subsidies

Each education subsidy consists of offering a reward conditional on attaining a specific grade. We divide the schooling spectrum into 4 different levels: Level 1 (grade 1 to 4), Level 2 (grade 5 to 8), Level 3 (grade 9 to 12), Level 4 (grade 13 onwards). For each level, we considered a per-period subsidy of 2 dollars.

3.2.3 Low-Skill Employment Subsidies

In order to enhance comparability between employment and schooling subsidies, we consider an age-specific payment of two dollars paid over the follow-

¹⁹An alternative way to model it would be to allow for the existence of two (or more) parallel education systems, differentiated by their academic content (like in Germany). See Belzil and Poinas, 2011, for a discussion.

ing intervals; 1 to 4, 5 to 8, 9 to 12, and 13 to 16. The payment is conditioned on being in state e .

3.3 The Distinction between Education Subsidies, Employment Subsidies and Compulsory Schooling

The key conceptual difference between mandatory schooling and policies affecting either the cost of schooling, or the utility of work, is that mandatory schooling automatically affects the lower tail of the ability distribution. Subsidies, even when paid at low schooling levels, may affect a much wider subset of the population (Belzil and Hansen, 2010).²⁰

Before comparing IV estimates, it is useful to provide a synthetic economic analysis of those policy interventions. In Table 1, we report the fraction of individuals affected by each policy intervention and two different indicators of the skill differential between those affected and unaffected. The indicators chosen are the average treatment effect of schooling (λ^*), which identifies the individual specific factor (or type), and the ex-ante educational attainment.

3.3.1 Education Subsidies

Upon examining the elements of Table 1, it is relatively straightforward to evaluate the consequences of displacing the subsidy from Level 1 to Level 4. There are two essential features. First, the education subsidies offer a monetary incentive conditional on reaching a specific level. In a sequential framework (when consumption takes time), the effect of the intervention is perceived through an option value, and is therefore affected by the rate of time preference. Second, the actual claim of the incentive payment also depends on individual skill heterogeneity.

Overall, there is evidence that the fraction affected decreases as the payment is delayed to higher levels, although the relationship is not perfectly monotonic. For instance, when set at level 1, the subsidy affects 38.5% of the population. When set at level 4, only 21.1% is affected. We also note that the average ability of those affected is also increasing with the level of

²⁰This may be explained intuitively by the fact that compulsory schooling may be generated by a subsidy (or cost) approaching minus (plus) infinity.

schooling upon which payment is conditioned. This is well illustrated by average treatment effect of schooling of those affected by the subsidy at Level 1, which is equal to 0.047, and is actually well below the population average. In turn, when the subsidy is paid conditional on reaching Level 4, those affected have an average treatment effect close to 0.09, which is well above population average.

3.3.2 Compulsory Schooling

Not surprisingly, implementing a set of compulsory schooling interventions has completely different implications.

First, as the mandatory schooling level is manipulated from one year (its minimum value) to four years, both the fraction of the population affected and its average ability is expected to increase. This simply arises from the fact that increasing the minimum by 1 year affects the extreme (bottom) tail of the skill distribution, and that raising the minimum to 2, 3, and 4 years, gradually recovers a larger set of the population. This is disclosed in Table 1, as the fraction affected ranges between 20.9% (1 year), and 50% (4 years), and the average factor goes from 0.011 to 0.035.

However, the average level of ability of those affected is quite different from what was observed for the education subsidies. To see this, it is informative to compare two distinct policies that affect a similar fraction of the population. For instance, a comparison between the population affected by a Level 1 subsidy (\$2) and a 3-years compulsory schooling policy reveals a huge ability gap, both in terms of the average treatment effect of schooling (0.047 vs. 0.029) and in terms of ex-ante schooling (1.6 years vs 0.72).

3.3.3 Low-skill Employment Subsidies

Although employment subsidies are rarely used as instruments in the applied literature, they raise some theoretical interest. Unlike education subsidies and mandatory schooling policies, work subsidies reduce schooling attainments. However, they are not fundamentally different from education subsidies, in that they achieve their goal by changing the relative price of schooling with respect to labor market work.

As do education subsidies, both the identity of those affected, and the fraction affected change with the age level at which it is implemented. Em-

ployment subsidies paid in early age, affect 30% of the population. The fraction goes to 8% for the subsidy paid between 5 and 8, 14.3% between 9 and 12, and 13.3% between age 13 and 16.

As the employment subsidy is delayed, the average ability of those affected raises significantly. The average effect of schooling among those affected by the subsidy paid between one and four, is equal to 0.054 (a value somewhat below population average), while the subsidy set between 13 and 16 affects individual with very high academic abilities (the average is close to 0.09)

Employment subsidies tend to affect a set of individuals who are relatively more able than those affected by education subsidies. However, the differences are far from being as large as those noted when we compared either of those subsidies to mandatory schooling.

Table 1
An Economic Analysis of Compulsory Schooling,
Education Subsidies, and Employment Subsidies

| | percentage | Average factor (λ^s) | | Education Ex-ante | |
|----------------------|------------|--------------------------------|-------------|-------------------|------------|
| | affected | affected | unaffected. | affected | unaffected |
| Education Subsidies | | | | | |
| Level 1/\$2 | 38.5 | 0.047 | 0.068 | 1.617 | 9.499 |
| Level 2/\$2 | 16.6 | 0.065 | 0.059 | 3.909 | 6.975 |
| Level 3/2\$ | 26.4 | 0.078 | 0.054 | 8.877 | 5.603 |
| Level 4/2\$ | 21.1 | 0.088 | 0.052 | 12.554 | 4.840 |
| | | | | | |
| Employment Subsidies | | | | | |
| Age 1-4 | 29.9 | 0.054 | 0.063 | 2.992 | 8.387 |
| Age 5-8 | 8.0 | 0.069 | 0.059 | 8.511 | 6.624 |
| Age 9-12 | 14.3 | 0.082 | 0.056 | 12.434 | 5.827 |
| Age 13-16 | 13.3 | 0.088 | 0.056 | 13.979 | 5.670 |
| | | | | | |
| Mandatory Schooling | | | | | |
| 1 year | 20.9 | 0.011 | 0.073 | 0.000 | 8.173 |
| 2 years | 32.0 | 0.022 | 0.078 | 0.320 | 9.342 |
| 3 years | 41.3 | 0.029 | 0.082 | 0.721 | 10.512 |
| 4 years | 50.1 | 0.035 | 0.085 | 1.119 | 11.828 |

Note: Education ex-ante refers to the number of years schooling before policy intervention.

4 IV Estimates using Various Policy Interventions

We now proceed with cross-section IV estimation. To do so, we merge the control group with each specific treatment groups. So, each IV estimate is computed from a sample containing 100,000 individuals (50,000 for treatment, 50,000 for control).

To be realistic, and in order to replicate what is usually achieved in the empirical literature, we use data on earnings measured around the middle of the life cycle interval devoted to labor market work. In the current model, we set it to period 35.

In appendix Table A3, we report OLS regressions using simulated earnings measured at 35. The estimate, equal to 0.11, is well above the population parameter (0.06).²¹ This also confirms that the model is consistent with the popular notion of positive Ability Bias, although the difference between OLS estimates and the population average is not solely due to unobserved ability.

Using standard vector notation, and dropping time subscripts, the IV estimator is defined as

$$\hat{\lambda}_{s,IV} = (Z'S)^{-1}Z'W \quad (8)$$

where W is a vector of log earnings, S is a vector of schooling attainments, and Z is a treatment-control indicator.

Because our estimates are obtained for a cross-section of outcomes measured at a given date, our estimates are implicitly conditioning on age.²² For the moment, we disregard identification, and remain agnostic with respect to the population parameter that may be targeted by IV estimation.

The IV estimates are reported in Table 2. Along with standard errors, we report the first stage F-Statistic, and the correlations between Z and W , as well as Z and S .

Before providing an economic discussion, we review briefly estimates found in Table 2. The IV estimates obtained for various education subsidies, range from 0.033 (Level 1) to 0.080 (Level 4). All of those estimates

²¹An OLS estimate of 0.11 is totally comparable to what would be obtained using different cross-sections of the NLSY. See Belzil and Hansen, 2002.

²²In the applied literature, most economists favor conditioning on age (as opposed to experience) since work experience is usually considered endogenous.

are relatively precise.²³ As normally expected, and as we move from Level 1 to Level 4, the estimates increase with the average treatment effect of those affected.

The results obtained for the instruments generated by employment subsidies display a pattern that is not so different from those obtained from the education subsidies, although they are slightly higher. For instance, the estimates range from 0.038 to 0.096, and are all very significant.

We now turn to the IV estimates obtained from compulsory schooling. Those are found in the lower portion of Table 2. The most striking result is the incidence of negative IV estimates. This is the case for both the one-year and the two-years compulsory schooling policies. The estimates, equal to -0.009 and -0.002, are actually comparable to many of those estimates reported in the recent literature on compulsory schooling.²⁴ All of them are small in absolute value, although the compulsory schooling policy that sets a minimum of 4 years generates an IV estimate (0.009) that is significantly different from 0. As a basis for comparison, the treatment effect for the sub-population affected by a four-year mandatory schooling intervention is equal to 0.035 (Table 1).

²³We compute a standard error using 100 bootstrap replications.

²⁴As noted in Belzil and Hansen (2010), in which earnings are drawn randomly between period 15 and 55, the negativity of compulsory IV estimates prevail also when additional controls for experience are introduced.

Table 2
IV Estimates for Various Policy Interventions

| | IV (st.error) | F Statistic | $Corr(Z, S)$ | $Corr(Z, W)$ |
|------------------------|----------------|-------------|--------------|--------------|
| Educ. Subsidies | | | | |
| Level 1/\$2 | 0.033 (0.003) | 550.0 | 0.074 | 0.0273 |
| Level 2/\$2 | 0.045 (0.002) | 710.9 | 0.084 | 0.0420 |
| Level 3/\$2 | 0.055 (0.002) | 637.5 | 0.080 | 0.0494 |
| Level 4/\$2 | 0.080 (0.004) | 135.9 | 0.037 | 0.0326 |
| Emp. Subsidies | | | | |
| Age 1-4 | 0.043 (0.004) | 212.7 | -0.046 | -0.020 |
| Age 5-8 | 0.054 (0.008) | 57.7 | -0.024 | -0.013 |
| Age 9-12 | 0.077 (0.006) | 79.1 | -0.028 | -0.022 |
| Age 13-16 | 0.128 (0.011) | 35.2 | -0.019 | -0.024 |
| Mand. Schooling | | | | |
| 1 year | -0.009 (0.017) | 33.9 | 0.018 | -0.0018 |
| 2 years | -0.002 (0.006) | 231.4 | 0.048 | -0.0012 |
| 3 years | 0.003 (0.002) | 800.5 | 0.089 | 0.0030 |
| 4 years | 0.009 (0.002) | 2253.3 | 0.148 | 0.0137 |

5 Comparative Advantages and Negative IV Estimates

We now show how the incidence of negative IV estimates relates to the skill formation technology.

5.1 Simulating Changes in Low-skill Job Productivity

At this stage, it is informative to analyze two more versions of the model that differ only up to the value of λ^e , namely a case where $\lambda^e = 0.005$, and another one where $\lambda^e = 0.030$. We compare those two cases to our benchmark model (in which $\lambda^e = 0.015$). It is important to note that, as we move from $\lambda^e = 0.005$ to $\lambda^e = 0.030$, we raise the productivity of one input to skill formation (namely low-skill employment), without changing the rest.

Before proceeding further, and because our model entails an indirect effect of schooling on the likelihood of working in a high skill job, we also evaluate the total effect. Basically, the indirect effect should measure the portion of the yearly growth rate in earnings that is caused by schooling. To approximate it, we evaluate the type (individual) and time specific quantity, $\lambda_i^I(t)$, which we define as

$$\lambda_i^I(t) = \frac{\partial \{\lambda^e \cdot E_{it} + \lambda^a \cdot A_{it}\}}{\partial S_{it}} \Big|_{type\ i} \quad (9)$$

and which is measured at the same period as wage outcomes (period 35).²⁵ The type specific values, which are reported in Appendix (Table A1), range between 0.003 and 0.005 (depending on type). This means that, generally speaking, the total effect of schooling exceeds λ_i^s by half a percentage point (or less).

To quantify the degree of comparative advantages that characterizes the data generating processes, we consider the following index, denoted $M(\cdot)$

$$M(\lambda^e) = \frac{1}{N} \sum_{i=1}^N I(\lambda_i^s + \lambda_i^I < \lambda^e) \quad (10)$$

²⁵To do so, we record simulated values of the numerator (for each individual belonging to a given type) and regress it on schooling. Strictly speaking, the indirect effect should be non-linear in schooling. However, for practical purposes, we approximate it with a linear relationship.

where $I(\cdot)$ is the indicator function, N is the population size, and where the time subscript has been dropped. Basically, $M(\lambda^e)$ measures the fraction of the population who is more effective at producing skills in the market (in low-skill jobs) than in school. As we move from $\lambda^e = 0.005$ to $\lambda^e = 0.030$, we gradually increase this fraction. The corresponding fractions are given in Table 3.²⁶ To illustrate the role of comparative advantages, we computed the resulting IV estimates for all of those 3 cases. The results are found in Table 3.

When the treatment effect of basic work experience is set to 0.005, the set of individuals who produce more skills in a low-skill job is actually empty since $M(\lambda^e = 0.005) = 0$. In such a case, there is no way the policy can affect a subset of the population more effective at producing skills in a low-skill job, and mandatory schooling policies tend to have a neutral (or positive) effect on earnings. For instance, this is illustrated by the one-year mandatory schooling policy, which leads to an IV estimate of 0.004. Obviously, given the low absolute value of this estimate, its degree of significance raises no interest.

As we progress toward a higher fraction of people more productive in the market, the potential for negative IV estimates starts to set in. For instance, with $\lambda^e = 0.015$ (our benchmark model which has already been analyzed), the fraction of the individuals more effective at work is now 0.10. However, this does not mean that only individuals belonging to this group will be affected, since stochastic shocks also play a role. Indeed, and as already noted in Table 1, the identity of those affected by a one-year policy (endowed with an average $\lambda^s = 0.011$) is consistent with a total effect of schooling approximately equal to 0.015. This total effect is practically equal to λ^e . So, in such a case, the IV estimates are predicted to approach zero (recall that the estimates are equal to -0.009 and -0.002 and are insignificant). Obviously, as the minimum number of years is increased (to three or four years), the policy affects an increasingly large number of individuals who are more effective in school. As a consequence, IV estimates become more and more positive (recall that they are equal to 0.003 and 0.009).

Let's now consider the model in which 25% of the population is more effective in the market ($\lambda^e = 0.030$). Because compulsory schooling affects

²⁶Obviously, we could also measure $M(\cdot)$ in terms of λ_i^s as opposed to $\lambda_i^s + \lambda_i^I$, since $M(\cdot)$ would still be increasing in λ^e .

the bottom tail, the majority of those affected is now more likely to be composed of individuals who belong to this group. The claim is verified upon examining the corresponding IV estimates. IV estimates remain negative up to four years. For instance, a one-year mandatory schooling policy generates an IV estimate equal to -0.017, and highly significant. Equally, the two-years and three-years compulsory schooling policies generate IV estimates that are equal to -0.010 and -0.005, and are also significantly different from zero.

This line of reasoning may also help explaining the positivity of IV estimates generated from education and work subsidies. For instance, the identity of those affected by work subsidies (the middle portion of Table 1) discloses that subsidizing low-skill jobs also entails a earnings depression, since those affected tend to do better in academic activities. However, because the policy also reduces schooling, IV estimates remain positive. It is interesting to note that, in this specific case, IV does not display the actual sign of the effect of the policy on earnings.²⁷

In order to verify that the negativity of IV is specific to compulsory schooling, we have also re-examined IV estimates generated by those subsidies implemented at low grade levels, when $\lambda^e = 0.005$, and when $\lambda^e = 0.030$. To do so, we have implemented both the Level 1 and the Level 2 education subsidies at each possible degree of comparative advantages used for mandatory schooling. The results are in the last two columns of Table 3. None of the IV estimates generated from those education subsidies, ranging between 0.03 and 0.06 (and highly significant) are even close to negativity. This is explained by the fact that none of those policies actually affect the extreme bottom tail, as was already noticed in Table 1.

As indicated earlier, we have ignored issues surrounding the precision of IV estimates. This was dictated by the “economic” nature of our analysis. The difference between a policy engendering a significantly negative effect, and one engendering no effect, is mostly due to the identity of those affected. The larger is the fraction of individuals who have comparative advantages in low skill jobs (with respect to the population of compliers), the more significant (negative) will be the effect. However, in a stochastic model, the presence of random shocks is sufficient to prevent a perfect coincidence. So, for this reason, the degree of statistical significance of those policies do not

²⁷In the applied literature, it is common for empiricists to claim that IV estimates a “policy effect”.

raise immediate interest.²⁸

5.2 An Interesting Paradox

Our analysis reveals an interesting paradox. As we increase the rate at which low-skill jobs produce new skills (leaving both the distribution of treatment effects of schooling, and the return to high-skill jobs unchanged), IV estimates of compulsory schooling become increasingly negative. In particular, negative IV estimates may arise even if each possible input (education, low-skill job and high-skill job) has a strictly positive effect on skill formation.²⁹

As an illustration, if technological changes increase the production of human capital induced by working in jobs that do not require academic ability, we would expect mandatory schooling policies administered to successive cohorts to display more and more negative values.

Strictly speaking, this paradox could also arise in the context where there is no heterogeneity in the treatment effect of schooling. Suppose that all individuals share a common and convex the log-wage relationship, with marginal (local) returns close to 0 in early grade levels and positive returns at higher education levels (Belzil and Hansen, 2002).³⁰ Suppose also that all individuals face the same (strictly positive) return to accumulated experience, as in a Mincer model. It follows that forcing individuals who have a high cost of schooling (who would normally obtain low levels of schooling) to stay in school for an extra year in school, would also reduce earnings as long as their post-intervention schooling level remains in the region where the local returns are close to 0.

²⁸It should however be noted that, generally speaking, IV estimates may be imprecise, even though its numerator (the correlation between wages and the policy shift indicator) is itself significantly different from 0. This actually turns out to be the case for many mandatory schooling IV estimates reported in Table 3.

²⁹It should also be clear that, although our results have been obtained from a framework where the treatment effect of schooling is the only skill accumulation technology parameter, similar results could be obtained while assuming cross-sectional heterogeneity in λ^e . This is the case, for instance, in Belzil and Hansen (2010).

³⁰The term marginal (or local) return refers to the increment in earnings (log) from one grade level to the next (see Belzil and Hansen, 2002).

5.3 An Intuitive Explanation

Our results may also be presented from a purely intuitive angle. To do so, assume away any indirect effect of schooling, and note that $E(\log w_{it} \mid Z_i = 1) - E(\log w_{it} \mid Z_i = 0)$, depends entirely on the earnings of those affected by the intervention, since for those unaffected, there is no difference (on average) between pre-intervention and post-intervention earnings. It follows that

$$\begin{aligned}
 & E(\log w_{it} \mid Z_i = 1) - E(\log w_{it} \mid Z_i = 0) = \\
 & \frac{1}{N} \sum_{i \in \check{A}} (\lambda_{i \in \check{A}}^s \cdot \{S_{i \in \check{A}, t} \mid Z_i = 1 - S_{i \in \check{A}, t} \mid Z_i = 0\} + \\
 & \quad \lambda^e \cdot \{E_{i \in \check{A}t} \mid Z_i = 1 - E_{i \in \check{A}t} \mid Z_i = 0\} + \\
 & \quad \lambda^a \cdot \{A_{i \in \check{A}t} \mid Z_i = 1 - A_{i \in \check{A}t} \mid Z_i = 0\})
 \end{aligned} \tag{11}$$

where N is population size, and where \check{A} denotes the set of individuals affected by Z .

Any policy that raises schooling, implies that

$$\lambda_{i \in \check{A}}^s \cdot \{S_{i \in \check{A}, t} \mid Z_i = 1 - S_{i \in \check{A}, t} \mid Z_i = 0\} > 0 \tag{12}$$

and that either

$$\lambda^e \cdot \{E_{i \in \check{A}t} \mid Z_i = 1 - E_{i \in \check{A}t} \mid Z_i = 0\} < 0 \tag{13}$$

or

$$\lambda^a \cdot \{A_{i \in \check{A}t} \mid Z_i = 1 - A_{i \in \check{A}t} \mid Z_i = 0\} < 0, \tag{14}$$

or both (13) and (14).

However, when policy designers implement mandatory schooling, which affects a very selective subset of the population, individual decisions are disconnected from comparative advantages. Those individuals who are endowed with low returns to academic activities (those for whom $\lambda_{i \in \check{A}}^s \cdot \{S_{i \in \check{A}, t} \mid Z_i = 1 - S_{i \in \check{A}, t} \mid Z_i = 0\} \approx 0$) are forced to delay post-schooling skill accumulation. In the particular case where the policy affects the extreme lower tail of the ability distribution (those who may never or rarely invest in jobs requiring academic skills), it is natural to expect $A_{i \in \check{A}t} \mid Z_i = 1 = A_{i \in \check{A}t} \mid Z_i = 0$.

The effect of compulsory schooling on earnings then depends on the negativity of $E_{i \in \check{A}t} | Z_i = 1 - E_{i \in \check{A}t} | Z_i = 0$. So, at any given point in time, the difference between treatment and control in the total amount of post-schooling skills accumulated, namely

$$\lambda^e \cdot \{E_{i \in \check{A}t} | Z_i = 1 - E_{i \in \check{A}t} | Z_i = 0\} + \lambda^a \cdot \{A_{i \in \check{A}t} | Z_i = 1 - A_{i \in \check{A}t} | Z_i = 0\} \quad (15)$$

is likely to be negative.

Table 3
IV Estimates with Different Degrees of Heterogeneity
in Skill Formation Technology

| | | IV estimates (st.error) | | | | | |
|--------------|----------------|-------------------------|---------|---------|---------|---------------------|-------------|
| | | Mandatory Schooling | | | | Education Subsidies | |
| | | 1 year | 2 years | 3 years | 4 years | Level 1/\$2 | Level 2/\$2 |
| Model | | | | | | | |
| λ^e | $M(\lambda^e)$ | | | | | | |
| 0.005 | 0.0 | 0.004 | 0.010 | 0.015 | 0.020 | 0.043 | 0.057 |
| | | (0.023) | (0.008) | (0.004) | (0.003) | (0.004) | (0.003) |
| 0.015 | 0.10 | -0.009 | -0.002 | 0.003 | 0.009 | 0.033 | 0.045 |
| | | (0.017) | (0.006) | (0.003) | (0.002) | (0.003) | (0.002) |
| 0.030 | 0.25 | -0.017 | -0.010 | -0.005 | -0.0003 | 0.026 | 0.037 |
| | | (0.005) | (0.005) | (0.002) | (0.002) | (0.003) | (0.002) |
| | | | | | | | |

Note: The fraction $M(\cdot)$ denotes the proportion of individuals who are more productive at producing human capital in the market than in school.

6 Placing our Results in Perspective

Before concluding, it is important to place our analysis in a wider perspective. As we claimed in introduction, our results provide a theoretical foundation for a wide set of empirical results that have recently been reported in the compulsory schooling literature, in the structural literature on estimating distributions of returns to schooling, and in the literature on education and growth.

However, at a more econometric level, our results may also be seen as complementary to those established by Heckman and Vytlačil (2005), and discussed in Heckman, Vytlačil and Urzua (2007), and Heckman (2010). So, for the sake of completeness, we briefly address the issue at the end of the section.

6.1 The Recent Literature on Compulsory Schooling

One of the most striking developments in the recent empirical literature is the incidence of low, and even negative, IV estimates obtained from compulsory schooling reforms. Devereux and Hart (2010) analyze the increase in the compulsory schooling age from 14 to 15, which took place in the UK, in 1947. The authors report IV estimates that are actually close to 0.³¹ In a more recent working paper, Chib and Jacobi (2011), using UK General Household Surveys, also analyze this policy change. Although their approach is slightly different from Devereux and Hart (2010), they also find evidence of very low IV estimates. Indeed, some of the estimates reported are actually negative (around -0.003).³²

Pischke and von Wachter (2008) also report estimates of the return to raising the minimum school leaving age in the former West Germany that are very close to zero. The incidence of very low (and negative) IV estimates is also reported in Grenet (2010), who analyzes the effects of a reform that raised the minimum schooling age in France in 1967, using the French Labor Force Survey.³³

³¹Cameron and Taber (2004) present several arguments that may explain why low estimates of the returns to schooling tended not to get reported for a long time.

³²Chib and Jacobi (2011) use Bayesian fuzzy regression discontinuity methods.

³³Other relevant studies are surveyed in Devereux and Hart (2010).

Despite the multiplicity of low returns, none of those studies actually provide any structural (theoretical) interpretation of the incidence of negative IV estimates. As of now, the prevalence of low (or negative) IV estimates is regarded as a puzzle.³⁴ As should be clear by now, the incidence of low (possibly negative) IV estimates generated from educational interventions targeting the bottom tail of the ability distribution, is a natural implication of the existence of comparative advantages. Essentially, IV estimates are bound to lie below zero as long as those individuals, who are more effective at producing skills in the market than in school, are also those affected by the policy. This is the case with compulsory schooling interventions, but generally not with other interventions.

6.2 Comparative Advantages and the Distribution of Returns to Schooling

In the structural literature on estimating distributions of returns to schooling (Carneiro, Hansen, and Heckman, 2003, and Heckman, Humphries, Urzua and Veramendi, 2011), earnings equations are usually specified as a Roy model. Using our notation, and in the specific case where the model does not incorporate any control for age or experience, the object estimated by the econometrician boils down to the population distribution of the following quantity:

$$\text{Ex-post returns}_i = \frac{\partial\{\lambda_i^s \cdot S + \lambda^e \cdot E + \lambda^a \cdot A\}}{\partial S} \quad (16)$$

Carneiro, Hansen, and Heckman (2003) and Heckman, Humphries, Urzua and Veramendi (2011), report a substantial fraction of individuals who experience negative ex-post returns to schooling.³⁵ Depending on the model specification considered, the fraction is close to 20%. It is interesting to note that this number is also consistent with results reported in Belzil and Hansen

³⁴This may be partly explained by the fact that empirical labor economists often interpret their estimates within the static Becker model (Becker, 1964), in which individuals no longer accumulate skills beyond schooling. The Becker model is used by Card (1999) in his survey of the IV literature on returns to schooling.

³⁵The term “ex-post returns” is used to characterize the parameters of the actual earnings data generating process. In the presence of imperfect information, individuals may base their schooling decisions on a different set of parameters which reflect the information available at the time of schooling decisions.

(2007) who find that between 20% and 25% of the population analyzed (US males sampled from the NLSY 79) is endowed with an individual specific treatment effect of schooling inferior to the early career return to general experience. Although the model is set within a classical Mincer framework (the post-schooling accumulation rate is exogenous), this is an indication that the parameters used to calibrate our model are actually realistic.

As seen earlier, our model provides a structural foundation for the existence of negative ex-post returns to schooling, even if each possible input (education, low-skill job and high-skill job) has a strictly positive effect on skill formation. Although the set of individuals affected by compulsory schooling may not be exactly coincident with the set of individuals characterized by $M(\cdot)$ because individuals are subject to random shocks, negative IV estimates obtained in the compulsory schooling literature are the mirror image of the existence of negative ex-post returns to schooling within a Roy model. That is as we increase (counterfactually) the rate at which low-skill jobs produce new skills (*ceteris paribus*), ex-post returns to schooling become negative for an increasing fraction of the population.

6.3 Education and Growth

Although our model is not directly concerned with growth, it still provides a theoretical foundation for the very weak (or negative) correlation data between economic growth and education, documented in the empirical literature. In this literature, economists attempt to capture the causal effect of education on growth rates, after controlling for cross-country differences in capital stock, and well as other factors affecting it. Following Bils and Klenow (2000), many economists have reported very weak, and sometimes negative, estimates of the effect of education on growth.³⁶

The empirical analysis is typically achieved using panel data of countries that have experienced, for the most part, expansions in their education system. The effect of education is identified by the co-movements of schooling levels of successive cohorts, and their associated growth rates in GNP.

However, returning to our model, the source of a given education expansion may have a non-neutral effect on growth. While countries that have experienced changes in education through incentive-based policies, are likely to

³⁶The literature is surveyed in Sunde and Vischer (2011).

experience positive changes in earnings across cohorts, this is not necessarily true of those that have increase schooling by changes in compulsory schooling. Those countries who have experienced expansion in average schooling levels by successive changes in compulsory schooling, are much more likely to experience a modest growth rate, since economic growth induced by other factors (technology, population growth, etc..) is likely mitigated by those education policies. In the extreme case where a large fraction of the population is more effective at producing skills in the market, negative effect cannot be ruled out.

As far as we know, the mechanisms underlying education expansion across cohorts have not been considered in the empirical growth literature. Taken as such, our model leads us to believe it should. After all, even if it does not incorporate capital investments and technological changes, it nevertheless singles out an effect of education policies that holds, other things held constant. The effect of education on growth has been at the forefront of the empirical literature for many years and, most likely, it will continue to be so in the future.

6.4 The Literature on IV Estimation of Treatment Effects

Although the focus of the paper was not on point estimation of the skill formation technology parameters, it is possible to draw some analogies with the literature on estimating treatment effects. In a series of papers, which include Heckman and Vytlacil (2005) and Heckman, Vytlacil and Urzua (2007), the authors express IV as a weighted average of Local average Treatment effects and show that within a Roy model, potential misspecification of the first stage model may induce negative IV weights, which in turn may lead to IV estimates having the opposite sign of the treatment effect of interest.

To some extent, the incidence of negative IV estimates coexisting with strictly positive effects of each input to skill formation, displays some commonalities with those issues raised by Heckman, Vytlacil and Urzua. However, the problem is not about choosing the right specification of the first-stage model, but about the specification of the outcome equation. Because our outcome equation is more general than the one analyzed in Heckman and Vytlacil (2005) and in Imbens and Angrist (1994), the independence as-

sumption (necessary for identifying a treatment effect) is violated.³⁷ Indeed, none of the existing treatment effects (treatment for the treated, treatment for the untreated, population average treatment effect, or local average treatment effect) may be viewed as the estimand corresponding to negative IV estimates.³⁸

To summarize, IV actually estimates a quantity that depends directly on the difference between two different structural parameters, and not on their level. So, the paradox described in the previous sub-section is a consequence of the failure of the IV identifying moment condition, which itself follows from the misspecification of the earnings equation.

7 Concluding Remarks

We have examined the implications of the existence of comparative advantages in human capital production, within a life cycle framework. Our model helps reconciling a large number of findings in the vast and heterogeneous literature devoted to the economic benefit of schooling. As illustrated in the paper, changes in the rate at which skills are produced in the labor market may have a pivotal force on (i) the sign of IV estimates generated from education interventions that target the bottom tail of the skill distribution, on (ii) the measured density of a given population endowed with negative ex-post returns, and on (iii) the effect of education on economic growth.

The implications of comparative advantages go far beyond what we covered in the paper. Many countries, in which higher education is heavily subsidized, are currently contemplating education reforms that may favor the development of professional education system, as a substitute for higher education enrollment.³⁹ We claim that the effectiveness of those policies is also highly dependent on the actual fraction of the relevant population endowed with comparative advantages in labor market work.

³⁷Vytlacil (2000) proved that the LATE framework of Imbens and Angrist (1994) is equivalent to a non-parametric version of the Roy model.

³⁸For instance, in Belzil and Hansen (2011), we perform a similar analysis on a cross-section of earnings that is representative of most data sets used empirical work. As earnings are measured at different points in time between early career and retirement, we can also perform IV regressions that control for experience. The results remain practically identical.

³⁹This is the case for Switzerland and France, among others. See Belzil and Poinas (2011) for a discussion

More generally, evaluating the effectiveness of any education policy requires the knowledge of the full skill formation technology, and in particular, knowledge of the fraction of the population who are more effective at producing skills in the market than in school. Surprisingly, this objective has been ignored by most empirical labor economists who, for more than 30 years, have focussed almost entirely on the point estimation of the treatment effect of schooling using natural experiments and the like. On the other hand, the estimation of life cycle human capital accumulation behavior models allowing for comparative advantages is still in its infancies. We believe that its future development is essential to the comprehension of a wide class of education policies.

References

- [1] Becker, Gary (1964), *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, New-York: National Bureau of Economic Research
- [2] Bellman, Richard (1957) “*Dynamic Programming*” Princeton, New-Jersey, Princeton University Press.
- [3] Belzil, Christian and Hansen, Jörgen (2011) “Education Policy Interventions, Earnings Sampling, and IV Estimation of the Return to Schooling”, Discussion paper.
- [4] Belzil, Christian and Hansen, Jörgen (2010) “The Distinction between Dictatorial and Incentive Policy Interventions and its Implication for IV Estimation”, IZA Discussion Paper 4835.
- [5] Belzil Christian and Francois Poinas (2011) “Academic vs. Professional Education in France and Germany: Modeling Selectivity and Outcomes”, Report presented to the Swiss Leading House, University of Geneva
- [6] Belzil, Christian (2007) “The Return to Schooling in Structural Dynamic Models: A Survey of the Literature” *The European Economic Review*, vol. 51, no 5, 1059-1105
- [7] Belzil, Christian and Hansen, Jörgen (2007) “A Structural Analysis of the Correlated Random Coefficient Wage Regression Model“, *Journal of Econometrics*, vol 140, 2, October, 827-848
- [8] Belzil, Christian and Hansen, Jörgen (2002) “Unobserved Ability and the Return to Schooling” *Econometrica*, 70, 575-91.
- [9] Ben Porath, Yoram (1967): “The Production of Human Capital and the Life Cycle of Earnings.” *Journal of Political Economy*, 75(4), pp. 352-365.

- [10] Bilal, Mark, and P. J. Klenow (2000): “Does Schooling Cause Growth?,” *American Economic Review*, 90(5), 1160-1183.
- [11] Cameron, S. V. and Christopher Taber (2004) “Estimation Of Educational Borrowing Constraints Using Returns To Schooling” *Journal of Political Economy*, 112, 132-182.
- [12] Card, David (1999) “The Causal Effect of Education on Earnings” *Handbook of Labor Economics*, edited by David Card and Orley Ashenfelter, North-Holland Publishers.
- [13] Carneiro, P. and James Heckman (2002) “The Evidence on Credit Constraints in Post-Secondary Schooling” *Economic Journal*, 112 (482), 705-734)
- [14] Carneiro, P., K. Hansen and J. Heckman (2003) “Estimating Distributions of Counterfactuals with an Application to the returns to Schooling and measurement of the Effects of uncertainty on Schooling Choice” *International Economic Review*, 44, 361-422.
- [15] Cunha, F., J. Heckman and Salvador Navarro (2005) ”Separating Uncertainty from Heterogeneity in Life Cycle Earnings” *Oxford Economic Papers* 57, 191–261
- [16] Cunha, F., J. Heckman and Schennach, S. (2011) “Estimating the Technology of Cognitive and Noncognitive Skill Formation ” *Econometrica*, 79, 575-91.
- [17] Deaton, Angus (2008) “Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development” *Keynes Lecture*, British Academy, October 9, 2008
- [18] Paul Devereux and Robert A. Hart (2010) “Forced to Be Rich? Returns to Compulsory Schooling in Britain” IZA Discussion Paper 3305, forthcoming in *The Economic Journal*
- [19] Heckman James (2010) “Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy” *Journal of Economic Literature* 48 (June 2010): 2, 356–398

- [20] Heckman, James, Lance Lochner and Chris Taber (1998) “Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings With Heterogeneous Agents,” *Review of Economic Dynamics*, January 1998.
- [21] Grenet, Julien (2010) “Is it Enough to Increase Compulsory Education to Raise Earnings? Evidence from French and British Compulsory Schooling Laws” CEPR Working Paper
- [22] Harmon Colm, and Ian Walker (1995) “Estimates of the Economic Return to Schooling for the United Kingdom” *American Economic Review*, vol 85, 1278-1296
- [23] Heckman, James (1997) ”Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations,” *Journal of Human Resources*, 32 (3), 441-62.
- [24] Heckman, James and Vytlacil, Edward (2005) “Structural Equations, Treatment Effects and Econometric Policy Evaluations” *Econometrica*. 73.
- [25] Heckman, James, Sergio Urzua and Vytlacil, Edward (2007) “Understanding Instrumental Variables in Models with Essential Heterogeneity”, *Review of Economics and Statistics*
- [26] Heckman, James, John Eric Humphries, Sergio Urzua and Gregory Veramendi (2011) “The Effects of Educational Choices on Labor Market, Health, and Social Outcomes” IZA Working Paper
- [27] Imbens, Guido and Angrist, Joshua (1994) “Identification and Estimation of Local Average Treatment Effects” *Econometrica*, 62, 467-76.
- [28] Imbens, Guido (2009) “Better Late than Nothing” , NBER Working Paper
- [29] Keane, M. (2010) “Structural vs. Atheoretical Approaches to Econometrics,” *Journal of Econometrics*, 156, 3-20.
- [30] Keane, Michael and Wolpin, Kenneth (1997) ”The Career Decisions of Young Men” *Journal of Political Economy*, 105, 473-522.

- [31] Lucas, Robert “On the Mechanics of Economic Development” *Journal of Monetary Economics* 22 (1988) 3-42
- [32] Pischke, Jörn-Steffen and T. von Wachter (2008) “Zero Returns to Compulsory Schooling in Germany: Evidence and Interpretation” *Review of Economics and Statistics*, vol 90 (3), 592-598
- [33] Rosenzweig Mark and K.Wolpin (2000) “Natural Natural Experiments in Economics” *Journal of Economic Literature*, December, 827-74.
- [34] Sunde, Uwe and Thomas Vischer (2011) “Human Capital and Growth: Specification Matters” IZA Discussion Paper 5991.
- [35] Vytlacil, Edward (2002) “Independence, Monotonicity, and Latent Index Models: An Equivalence Result”, *Econometrica*, 70(1): 331-341
- [36] Willis, R. and S. Rosen (1979) “Education and Self-Selection”, *Journal of Political Economy*, 87, S-7-S36.

Table A1
The Heterogeneity Distribution

| type | α^S | λ_s | $\lambda_s + \lambda_i^I$ | c_0^a | proportion |
|----------------|------------|-------------|---------------------------|---------|------------|
| 1 | -2.990 | 0.005 | 0.009 | 9.955 | 0.05 |
| 2 | -2.750 | 0.005 | 0.009 | 9.895 | 0.10 |
| 3 | -1.790 | 0.015 | 0.019 | 8.695 | 0.10 |
| 4 | -1.310 | 0.023 | 0.028 | 8.215 | 0.10 |
| 5 | -0.470 | 0.035 | 0.040 | 7.375 | 0.05 |
| 6 | 0.130 | 0.044 | 0.049 | 6.775 | 0.05 |
| 7 | 0.550 | 0.050 | 0.055 | 6.355 | 0.05 |
| 8 | 0.844 | 0.055 | 0.061 | 6.061 | 0.05 |
| 9 | 1.324 | 0.062 | 0.067 | 5.581 | 0.05 |
| 10 | 1.570 | 0.066 | 0.070 | 5.335 | 0.05 |
| 11 | 2.050 | 0.073 | 0.079 | 4.855 | 0.05 |
| 12 | 2.530 | 0.080 | 0.086 | 4.375 | 0.05 |
| 13 | 2.770 | 0.084 | 0.089 | 4.135 | 0.05 |
| 14 | 3.010 | 0.087 | 0.093 | 3.895 | 0.05 |
| 15 | 3.250 | 0.091 | 0.097 | 3.655 | 0.05 |
| 16 | 3.484 | 0.094 | 0.102 | 3.421 | 0.05 |
| 17 | 3.964 | 0.102 | 0.108 | 2.941 | 0.05 |
| Mean | 1.188 | 0.060 | | 5.717 | - |
| St Dev. | 2.062 | 0.031 | | 2.062 | - |

Note: The term, λ_i^I , which is equal to $\frac{\partial \{\lambda^e \cdot E_{it} + \lambda^a \cdot A_{it}\}}{\partial S_{it}}$, denotes the indirect effect of schooling on post-schooling earnings growth. It is a measure of non-separability between schooling and experience and is measured at $t = 35$.

Table A2
Life Cycle Choices in the Control Group:
Common Slope Model

Accumulated number of periods in each State

| year | In School | work (learning) | Work (Training) |
|------|-----------|---------------------|--------------------|
| 1 | 0.791 | 0.209 | 0.000 |
| 10 | 5.267 | 4.733 | 0.010 |
| 20 | 6.503 | 12.554 | 0.943 |
| 30 | 6.503 | 21.914 | 1.583 |
| 40 | 6.503 | 29.150 | 4.347 |
| 50 | 6.503 | 37.185 | 6.311 |

Table A3
OLS Regressions

| | Model |
|-------------------------|-------------------|
| intercept | 0.563 (0.012) |
| education | 0.113 (0.0002) |
| experience | 0.036 (0.001) |
| experience ² | -0.0002 ≈ 0 |
| R² | 0.52 |