The Distinction between Dictatorial and Incentive Policy Interventions and its Implication for IV Estimation*

Christian Belzil
Ecole Polytechnique, ENSAE, and IZA

Jörgen Hansen
Concordia University, CIRANO, CIREQ and IZA

April 29, 2011

*We would like to thank Steve Durlauf, Chris Ferrall, Paul Gomme, Bo Honore, Nikolay Gospodinov, Susumu Imai, Winfried Koeniger, James McKinnon, Hessel Oosterbeek, Paul Devereux, Steve Pischke and, more particularly, Arnaud Maurel, Jean-Marc Robin and 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, Concordia University, the First French Econometrics Conference in Honour of Alain Montfort (Toulouse), the Malinvaud Seminar at CREST, and the Institute for Economic Analysis at Autonoma University in Barcelona. We have also received useful comments and suggestions from an Editor and four anonymous referees. The usual disclaimer applies. Finally, we thank Xingfei Liu for capable research assistance.

Abstract

We consider the estimation of the treatment effect of education on earnings by IV, while maintaining the hypothesis that the data are generated by a dynamic life-cycle skill accumulation model. We provide an answer to the following question: would the econometrician prefer an instrument generated by the implementation of compulsory schooling regulation (a form of "dictatorial policy intervention"), or one generated by an educations subsidy (an "incentive-based" policy reform). Because the identifying moment condition can be decomposed in two mutually exclusive ex-post conditions that are based on the complying status; namely *Post-Treatment Choice Orthogonality and Post-Treatment Choice Opposite Selection,* we can introduce both the identity of those affected (and unaffected) and the distinctive nature of a policy into the analysis. The policy design exercise requires two elements; to adjust the identity of those affected (the complier/non-complier heterogeneity gap) to the conflicting effects of schooling on post-schooling skill formation and on its opportunity cost, and to keep track of the statistical strength of the instrument. Because of its relative degree of flexibility, education subsidies are capable of generating a reliable instruments. However, compulsory schooling regulations, which basically annihilate comparative advantages for a non-random subset of the population, offer practically no design possibilities. Our calibrated model sheds light on the incidence of the low (and possibly negative) IV estimates that have recently been reported in the recently emerging literature on compulsory schooling.

Key Words: *Returns to schooling, Instrumental Variable methods, Dynamic Discrete Choice, Compulsory Schooling Reforms, Dynamic Programming, Treatment Effects .*

JEL Classification: B4, C1, C3.

1 Introductory Remarks

Consider an econometrician, who seeks to estimate the treatment effect of education on earnings by Instrumental Variable (IV) methods, while maintaining the hypothesis that the data are generated by a specific dynamic skill accumulation model. To place this exercise in a concrete framework, and in order to eliminate all ambiguity related to the choice of a treatment effect, we initially set the analysis in a classical model where the parameter capturing the effect of schooling is common across all individuals (a common slope model).

This paper addresses the following question. How would the econometrician design a policy intervention in order to generate an instrument that fulfills the identifying orthogonality condition, and that allows to estimate the parameter of interest accurately? The issue is therefore about finding the "optimal policy design" for a given behavioral model, as opposed to imposing behavioral assumptions (assuming orthogonality between the instrument and the error term), given a specific instrument.[1]

The distinctive characteristics of these potential policies constitute the primary objects of our analysis. In our setting, the econometrician has to choose between instruments generated by the implementation of possible compulsory schooling policies, which belong to the class of "dictatorial policy interventions", and others generated by various education subsidies paid conditional on achieving a certain schooling level, which belong to the class of "incentive-based" policy reforms.

In the dynamic skill accumulation model, the effect of schooling on wages is a structural policy invariant parameter that summarizes part of the skill formation technology. Within an IV estimation procedure, this parameter is identified by an orthogonality condition.

Towards the end of the paper, and in order to stress the empirical relevance of our analysis, we relax the common slope assumption and consider a model with heterogenous slopes (a model with comparative advantages in schooling).

At the outset, it should be clear that our approach is targeted toward

[1]The term "optimal" is used somewhat loosely. We use it to characterize interventions that generate instruments that are as close as possible to meet the identifying (orthogonality) conditions. In a more general setting, the cost of implementing a policy, or the degree of precision of the IV estimate could be an input in the analysis.

empirical studies that use reforms directly as instrument. This is always the case in the literature on schooling and earnings.[2] We however disregard other cases where reforms are used to create difference-in-difference estimation strategies.

To mimic what is done in empirical work, the econometrician considers schooling as the only right-hand side endogenous variable because he/she has access to a single instrument. Because the data generating process is a life-cycle skill accumulation model, the error term of the econometrician is generated by a sequence of unobservable (latent) choices, which are exercised after the realization of schooling.

The following questions naturally arise.

1. How can one re-express the identifying moment (orthogonality) restriction relating treatment and control, in order to introduce the distinctive characteristics (whether the intervention is dictatorial or incentive-based) of the policies into the analysis?

2. Is the nature of the intervention relevant in assessing the capacity for IV to recover the treatment effect parameter of interest?

3. Which type of intervention is more likely to fulfill the identifying orthogonality condition?

4. Can our analysis shed light on the discrepancy between very low (potentially negative) IV estimates reported in the recently emerging literature on compulsory schooling and those generated by other types of instruments?

Because of the nature of the problem we raise, a full set of answers cannot really be obtained only from "paper and pencil" analysis. To answer some of those (and other peripheral) questions, as well as to illustrate some key concepts, we construct a calibrated life cycle model of human capital accumulation (in the spirit of Ben-Porath, 1967). To remain as general as possible, we build the most "stripped-down" version possible of a modified Ben-Porath model. Some features of our model resemble the Heckman, Lochner and Taber (1999) model, although our stochastic specification is richer.

[2]Empiricists often refer to the treatment effect of schooling on earnings as the "return to schooling" even though the expression is misleading.

The model is used as the data generating process of the control group, and we implement IV estimation by merging data from the control group and a large number of different treatments groups, generated by specific policy interventions. This allows us to investigate the relative performance of instruments generated by education subsidies and by compulsory schooling regulations.

As the life-cycle skill accumulation model may be viewed as the "benchmark economic structure" for analyzing the relationship between schooling, earnings, and experience, it is natural to condition our policy design analysis on it. Although our model is relatively simple, it nevertheless has some nontrivial aspects that render IV estimation challenging. For instance, it implies a dynamic relationship between schooling and subsequent skill accumulation decisions, as well as a high (and persistent) degree of selectivity.

In other words, we put IV estimation under the econometric analog of a "stress test", and analyze how, and if, each type of policy can survive the test (given a model).

The main contributions of this paper are both methodological and empirical. At the methodological level, and to our knowledge, this is the first paper that relates IV estimation to a policy design exercise. It is also the first that proposes an economic interpretation of the identifying moment condition, as a restriction involving the differences in post-intervention behavior of those affected and those who are not, and that uses this restriction to introduce the economic nature of the intervention into the analysis. Basically, it is also the first paper that considers the distinction between dictatorial and incentive-based interventions within a life-cycle framework, and that translates the distinction into formal economic concepts.

As we argue below, our analysis has also strong implications for empirical work, and more specifically for the emerging literature on returns to schooling, which is based on changes in compulsory schooling regulations. Our paper is also the first that provides a clear economic explanation for the incidence of very low (possibly negative) IV estimates of the effect of schooling.

To some extent, our paper departs completely from the recent literature on IV estimation. As our objective is neither to advocate, nor to impede IV estimation, our approach is positive. Effectively, we show how one can use a model, to design a policy intervention that may fulfill the identifying assumptions.[3]

The main results are as follows. First, we show that the identifying moment condition can be decomposed into two mutually exclusive ex-post conditions that are based on complying status; namely *Post-Treatment Choice Orthogonality (with Respect to Complying Status) and Post-Treatment Opposite Choice (with Respect to Complying Status).* We also discuss the link between those notions, and a third notion of ex-post randomization; namely *Post-Treatment Identity Randomization (with Respect to Complying Status).* Re-writing the classical moment condition as a behavioral condition based on complying status, allows us to introduce both the identity of those affected (and unaffected) and the distinctive nature of a policy into the analysis.[4]

At a more practical level, we show that policy design essentially requires two elements; to adjust the identity of those affected (the complier/non-complier heterogeneity gap) to the conflicting effects of schooling on post-schooling skill formation and to its opportunity cost, and to keep track of the statistical strength of the instrument (the correlation between schooling and the treatment control indicator) .

Because of its relative degree of flexibility, incentive-based policies (such as education subsidies) are capable of generating an instrument that may be used to estimate a policy invariant treatment effect.

On the other hand, dictatorial policy interventions (such as compulsory schooling regulations) offer practically no design possibilities. Both the statistical strength of the instrument generated by the policy, and the identity of those affected, are bound to the distribution of ex-ante schooling choices. Within a non-trivial dynamic skill accumulation, and under realistic wage sampling procedures, the capacity to fulfill the IV identifying orthogonality condition is related to the nature of the policy intervention used to generate an instrument. In particular, when offered the possibility to design an optimal policy, the econometrician should never choose a dictatorial (compulsory schooling) policy.

Although stylized, our calibrated model of life-cycle skill accumulation may actually reconcile the well known discrepancy between IV estimates

[3]For opposite views regarding IV estimation strategies, the reader can refer to Keane (2007), Heckman, Urzua and Vytlacil (2005), Deaton (2008) and Imbens (2009).

[4]As will become clear later, the term "identity" refers to the values of the heterogeneity component vector (the individual specific primitives of the model) that characterize a set of individuals (for instance, those affected by the policy).

obtained from different policy interventions. We show precisely how the incidence of low (and negative) IV estimates, that have recently been reported in the literature, are tied to the existence of comparative advantages in schooling. Essentially, the intuition is that any intervention affecting the lower tail of the ability distribution, is bound to generate negative IV estimates as long as a non-trivial fraction of the lower tail distribution may achieve higher earnings growth in the market. The incidence of negative IV estimates does not even require negative effects of schooling on skill formation (skill destruction) for the lower tail of the ability distribution. In other words, compulsory schooling regulations basically annihilate comparative advantages for a non-random subset of the population.

The residual parts of the paper are organized as follows. In Section 2, we present some background material. In Section 3, we lay-out the behavioral model. In Section 4, we discuss issues related to the choice of the treatment effect that we target as estimand. The fifth section is devoted to the construction of the treatment and control groups. In the sixth one, we discuss the implicit restrictions on the post-schooling skill accumulation behavior of those affected and those who are not, that are implied by the IV identifying condition (between treatment and control). In Section 7, we implement IV estimation on a wide range of treatment groups emerging from the implementation of compulsory schooling and education subsidies. In Section 8, we extend our analysis to a slightly modified version of our calibrated model (allowing for comparative advantages in schooling), and show that it may easily reconcile the findings that are reported in the recent literature on compulsory schooling. The final section offers concluding remarks.

2 Background Material

As a first step, we provide a more precise discussion of the distinction between interventions that are dictatorial in nature, and those that are based on incentive provisions.

Subsequently, and in order to anchor our analysis in the empirical literature on earnings and schooling, we review the recent IV literature on compulsory schooling and argue that our analysis may shed light on results reported therein.

2.1 Dictatorial vs. Incentive-Based Interventions

A dictatorial policy intervention, as we define it, affects individual decisions by restricting the choice set. In this paper, we focus on policy interventions that impose a minimum school leaving age. In terms of an underlying economic model, the dictatorial intervention prevents some individuals (those who are affected) to act on the basis of both their comparative advantages and on realized idiosyncratic random shocks (ex-ante risk), over a period determined by the intervention itself.

An incentive policy intervention works differently. It offers a monetary incentive (or a disincentive) conditional on reaching a pre-determined level of schooling. In a sequential framework, the incentive policy intervention is characterized by at least three dimensions; the per-period subsidy, the starting period of the subsidy (the minimal consumption/investment level upon which payment is conditioned), and the duration (number of periods over which it is paid). The incentive policy intervention normally applies to all individuals, but like compulsory schooling, it usually affects a subpopulation. When the reward is set conditionally on reaching a high level of schooling, the effect of the intervention is perceived through an option value. The actual claim of the incentive payment therefore depends on individual skill heterogeneity, on the rate of time preference (or its distribution), on realized random shocks, and on the amount of the subsidy itself.

Because educational interventions (just like other natural or social experiments) are scarce, it is impossible to rely on observational data in order to comprehend the differences between IV estimates obtained from different policy interventions applied to a same population. Instead, we rely on a mixture of calibration and more standard IV econometric techniques. Although our results are illustrated within a relatively precise economic framework, we will show that some of our conclusions may easily be extended to other economic models.

2.2 The IV Literature on Compulsory Schooling

In the literature, compulsory schooling reforms are systematically used to estimate returns to schooling, while incentive-based education policies are much less popular.[5] However, their relative capacities to achieve identifying orthogonality conditions have neither been investigated, nor questioned. This is understandable. Because policy reforms (and natural experiments) are scarce, incentive and dictatorial policy interventions are never administered on the same data generating process, and as a consequence, the econometrician practically never gets to choose between those two types of intervention.

In the literature, it is common to rely on static interpretations of the wage schooling relationship, and the conventional wisdom was, until recently, that IV estimates of the returns to education are high (Card, 1999). However, a thorough review of the IV estimates obtained from recent studies using compulsory schooling reforms seem to point to returns that are very low.

Devereux and Hart (2010) analyze the increase in the compulsory schooling age from 14 to 15, which took place in the UK, in 1947. Interestingly, they analyze the same reform that Harmon and Walker (1995) analyzed in their seminal piece, and point out that their relatively high estimates are explained by lack of control for cohort effects. When corrected, the authors report IV estimates that are much lower, especially for females. Some of those estimates are actually close to 0.

In a more recent working paper, Chib and Jacobi (2011), using UK General Household Surveys, also analyze this policy change. Their approach is slightly different from Devereux and Hart (2010), as they use Bayesian fuzzy regression discontinuity methods. They nevertheless find evidence of very low IV estimates. Indeed, some of the estimates reported are actually negative (around -0.003).

Meghir and Palme (2005) find that a Swedish reform that raised compulsory schooling to 9 years (from 7-8 years), raised earnings by no more than 1%. 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 0, and even negative.

Similar findings are displayed 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. Similarly, Osterbeek and Webbing (2007), find negative estimates associated to a mandatory vocational program in the

[5] See Card (1999), for a survey of IV literature on the returns to schooling. For a survey of some incentive-based education policy interventions, see Nielsen, Sorensen and Taber (2009).

Netherlands.

Surprisingly, and despite the existence of strong similarities in the results reported in the recent literature, none of those papers aforementioned offer any economic explanation. As of now, the prevalence of low (or negative) IV estimates is at best regarded as an indication of the ineffectiveness of mandatory schooling regulations, and at worst, regarded as a puzzle.

This emerging consensus among empirical labor economists about very low IV estimates obtained from compulsory schooling, and coexisting with the conventional wisdom supporting high IV estimates obtained from other instruments, suggests the relevance of our analysis. Ideally, it would be interesting to compare IV estimates that use the same data.

As argued in the introduction, the scarcity of policy changes renders this comparison difficult. Indeed, among all those studies aforementioned, only one paper (Grenet, 2010) may be matched with another study that use the same data set (the French Labor Force Survey), but a different instrument. Maurin and Xenogiani (2007) use changes in the incentive to enter higher education in France (in the mid 1960's), and found IV estimates of the wage return to education around 13%. These estimates, also obtained from the French labor Force Survey, should be put in perspectives with the very low estimates reported by Grenet (2010).

This quasi-inexistence of studies using same data but different instruments justifies the need for a calibration exercise. Our conjecture is that the discrepancy between compulsory schooling and other IV estimates, and more specifically the negative IV estimates occurring in the literature, is a logical consequence of dictatorial education policies. As this will become clear later, our analysis will shed light on those empirical findings.

3 The Behavioral Model

Our model is in the spirit of the classical Ben-Porath model (1967), although it also has some fundamental differences.[6] In the Ben-Porath model, the cost

[6]Dynamic skill accumulations models are rarely estimated. Keane and Wolpin (1997) is the first known empirical model where individual skills (occupation dependent) are accumulated within a dynamic structure. Heckman, Lochner and Taber (1999) are the first to implement a calibrated general equilibrium version of the Ben-Porath model. More

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. Finally, the Ben-Porath model admits skill complementarity has the sole possible skill formation technology. This follows from the fact that skills are self-productive.[7]

Our desire is to formulate a model that avoids all those problems, and that has the lowest degree of complexity possible. For this reason, we keep the number of states as small as possible, and we limit ourselves to a one factor model for heterogeneity. Yet, the structure of the model must allow for a high degree of selectivity on the heterogeneity factor, so that policy design is not trivial.

Although the first part of the paper devoted to the policy design exercise does not strictly speaking require a model based on fully realistic parameters, one of our objectives, which is to provide explanations for the incidence of low IV estimates of compulsory schooling, leads us to construct a model that is as close as possible to reality.

To achieve this, our model is constructed as to fulfill the following conditions, which we view more or less as "stylized facts" for western countries.

1. Schooling should be located in the early phase of the life-cycle.

2. The incidence of the intensive human capital accumulation state (work with formal training) must be declining with age.

3. OLS regressions of simulated wages on accumulated experience (potential) should disclose a declining return (a concave wage profile).

recently, Adda, Dustmann, Meghir and Robin (2006) have implemented a dynamic model of Apprenticeship and on-the-job training based on partial equilibrium search/matching arguments.

[7] Heckman, Lochner and Taber (1999) resolve some of these problems by separating the schooling decision from the post-schooling accumulation process. They assume a Ben-Porath technology upon school termination. However, they do not consider stochastic specification as rich as the one we consider.

4. OLS regressions of simulated wages on education should produce a higher return than the average in the population (the classical "ability bias" hypothesis).[8]

5. There must be a positive correlation between schooling, and individual returns to schooling (individual specific slope coefficients)

As a starting point, we choose hourly wages as the benchmark utility. To choose the preference parameters, we relied mostly on the structural literature, in order to obtain a realistic range of the relevant parameters (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.

3.1 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. The states are the following; schooling (s), work with a low rate of skill accumulation (e), and work with a high rate of skill accumulation (a).

The choices are summarized in the binary indicators, d_{tk}, where $d_{tk} = 1$ when option k (s, w, 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 Full-time 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.

[8] One of the most convincing results in favor of the classical ability bias hypothesis, is that OLS estimates of the wage effect of schooling that are based on specification incorporating some measure of cognitive-skills, are always inferior to those that are obtained without any measure.

3.2 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 $I(.)$ 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.

3.3 The Dynamics of Skill Accumulation

We assume that activities generating skill formation entail some psychic cost.[9] 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.[10]

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, $C_{it}^a()$. Precisely, U_{it}^e, U_{it}^a, and $C_{it}^a()$, are given by the following equations;

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

[9]In 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).

[10]This utility reduction could be interpreted as the monetary equivalent of leisure reduction. Indeed, common sense suggests that in actual labor markets, individuals who experience important promotions/higher earnings growth, are also those who supply large numbers of hours of work (high effort). Their earnings growth achievements are therefore not at the expense of their current gross earnings.

$$U^a_{it} = W_{it} - C^a_i(S_{it}) \tag{3}$$

$$C^a_{it}() = c^a_{0i} + c_{1a} \cdot S_{it} + \varepsilon^a_{it} \tag{4}$$

and where c_{1a} captures the effect of accumulated schooling on the cost (or disutility) of work, or work and training. As in the classical Ben-Porath model, we only consider complementarity, and assume that c_{1a} is a negative parameter. The $\varepsilon^{j\prime}_{it}s$ are stochastic shocks.

3.4 The Wage Equation

The wage equation is given by the following expression:

$$\log W_{it} = w_{it} = \alpha + \lambda^s \cdot S_{it} + \lambda^e \cdot E_{it} + \lambda^a \cdot A_{it} + \varepsilon^w_{it} \tag{5}$$

where W_{it} is the wage rate per unit of time, α is the intercept term, λ^s is the effect of schooling on wages, λ^e is the effect of employment on wages, λ^a is the effect of training on wages, and ε^w_{it} is a random shock (described below) .

3.5 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, $V^k_t(\Omega_t)$, equal to

$$V^k_t(\Omega_t) = U^k_t + \beta E \max\{V^1_{t+1}(\Omega_{t+1}), ..V^K_{t+1}(\Omega_{t+1}) \mid d_{kt} = 1\}$$

or, more compactly, as

$$V^k_t(\Omega_t) = U^k_t \ + \beta E V_{t+1}(\Omega_{t+1} \mid d_{kt} = 1)$$

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

3.6 The Distribution of Individual Heterogeneity and Random Shocks

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

$$\upsilon_r = \{\alpha_r^S,\ c_{0r}^a; p_r\} \text{ for } r = 1, 2, ..R \tag{6}$$

where p_r is the population proportion of type r. In the second part of the paper, we also allow λ^s to be individual specific, so to allow comparative advantages in schooling.

We started to work with 20 types with equal density (0.05). In the end, we adjusted the number of types to 17 by equating three different pairs. So, we end up with three types (type 2, type 3, and type 4) that have density equal to 0.10. 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 $k = s, e, a$.

3.7 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 returns to schooling (λ^s) is set at 0.06 (a value close to estimates reported in the structural literature), while the return to employment (set to 0.01) 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

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).

3.8 Introducing Comparative Advantages in Schooling

For reasons that are discussed at length in the last section of the paper (Section 8), we also consider a version of our model in which the effect of education on wages is individual specific. This feature is important in order to explain the incidence of low IV estimates reported in the emerging literature on compulsory schooling policies.

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 wages is subject to cross sectional dispersion. As in Belzil and Hansen (2007), the distribution of λ^s ranges between 0.005 and 0.12. To keep a certain degree of symmetry across models, we set the population average effect to the same level as in the common slope model (0.06). The distribution of the other heterogeneity components remain unchanged.[11]

4 Defining the Treatment Effect

Before going further, we need to select the treatment effect that we wish to estimate. In our model, the effect of schooling is summarized by two time invariant parameters; λ^s and c_{1a}. One possibility is to target the effect of schooling of earnings, after conditioning on post-schooling skills, namely λ_s. Another approach is to measure a total (unconditional) effect. Intuitively, the unconditional effect should subsume the direct effect of λ^s and the indirect effect of c_{1a} on earnings. It should therefore exceed λ^s since schooling stimulates skills through its effect on training.

Because this issue has been raised in the statistical literature on dynamic treatment effects, we first discuss various approaches considered by statisticians.

[11]Both heterogeneity distributions, namely $\{\alpha_r^s,\ c_{0r}^a; p_r\}$ in the common slope model, and $\{\alpha_r^s, \lambda_r^s\ c_{0r}^a; p_r\}$ in the comparative advantage model, are presented in Table A1.

4.1 Conditional or Unconditional Treatment Effect?

In the statistical literature, the difficulty to estimate treatment effects within a non-trivial dynamic framework has already been recognized. As such, there is no "universally accepted" solution.

In biostatistics, Robins (1997), Gill and Robins (2001), and Murphy (2003) have focused on "sequential randomization" arguments, which are more or less equivalent to "Matching". These notions have also been recently imported in Econometrics (Lechner and Miquel, 2002). However, for most economists, the interest for this type of approach is limited, as the distinction between the econometrician's information set and the agent's information may be highly important in applied work (see Aabring and Heckman, 2005, for an exhaustive survey of the treatment effect literature).

In applied micro-econometrics, the difficulty to implement valid IV exclusion restrictions has been recognized in the training literature. For instance, Eberwein, Ham and Lalonde (1997) examine a multi-state duration model of training participation, and post-program-labor market transitions. Although they are not concerned with IV estimation, they discuss the lack of foundations of standard IV exclusion restrictions.

Aside from adopting a specific estimation method suited to a dynamic treatment effect model, statisticians sometime advocate to define the population treatment effect as unconditional with respect to intermediate choices and outcomes. This is largely explained by the fact that in many applications cited in the statistical literature on dynamic treatment effect, treatment is short and has no impact of time allocation.[12] However, this view may sometime be difficult to transport to a dynamic life-cycle skill accumulation context. For reasons that are explained below, and that are more or less related to convenience, we choose a conditional definition of the treatment effect.

4.2 Separating Causal Effect, Spurious Correlation, and Opportunity Costs

At this stage, it should be clear that moving from a conditional to an unconditional analysis is not a solution to IV estimation. To see this, it is sufficient to recognize that the correlation between post-schooling skills and schooling

[12] See Rosenbaum (1984).

is due to three different channels.

One channel is causal, and is explained by the skill formation technology. In our model, this arises because of the negativity of c_{1a}.

The two other channels are disconnected from the skill formation technology. One is explained by the fact that treatment (namely education) may be highly time consuming. Effectively, the opportunity cost of being in one state is the reduction in time spent in alternative states. Finally, a last source of correlation between schooling and post-schooling skills is explained by persistent unobserved heterogeneity.

In theory, it would be possible to incorporate the effect of education on training into the total effect of schooling, but this new parameter would introduce some inconveniences.

First, because of nonlinearity, the total effect is not policy invariant, and as is the case for any correlated random coefficient model, it would be instrument dependent. This would imply a departure from a common slope specification.[13]

A second reason has to do with the absence of closed-form expression for the total causal effect of schooling. When the effects of schooling and post-schooling skills are not separable, it is difficult to express the total effect as a single time invariant parameter.

Aside from the inconveniences aforementioned, the limit of the unconditional approach is also well illustrated by considering a version of the model in which there would be no causal effect of schooling on the cost of learning future skills ($c_{1a} = 0$). In such a case, there is no logical reason to conduct an unconditional analysis. Yet, the estimated IV parameter would still be affected by the co-movements between schooling and post-schooling skills, which comes from some undefined (and uninterpretable) combination of spurious correlation and opportunity cost.[14]

[13] In the treatment effect literature, it is customary to consider outcome equations that are multiplicative in the error term. In such a case, the Local Average Treatment Effects (LATE), defined in Imbens and Angrist (1994) plays a central part. In a seminal piece, Vytlacil (2002) has established the fact that the LATE approach to IV estimation (Imbens and Angrist, 1994) is equivalent to a non-parametric selection model. See Heckman, Urzua and Vytlacil (2005) and Imbens (2009) for opposite views about the relevance of the LATE parameter. Rosenzweig and Wolpin (2000), and Keane (2007) argue that orthogonality conditions usually hide a large number of implicit assumptions.

[14] In economic models where the treatment is short (when it corresponds to a negligible fraction of total time endowment), opportunity costs no longer play a major role. This

This last point raises one ultimate reason for not conducting an unconditional analysis, namely that one of our objectives is simply to show that it is neither necessary, nor sufficient to adjust the parameter of interest if the policy is designed accurately.

5 The Control and Treatment Groups

5.1 Control Group

To generate the control group, we simulate 50 years of choices and wage outcomes for 50,000 individuals. We basically 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. 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.

Descriptive statistics of the number of periods spent in each state is found in Table 1. A summary of the distribution of schooling attainments by level is found in Table 2. A more detailed summary is found in appendix (Table A2).

Overall, the frequencies display the desired features. The average schooling attainment is equal to 6.5 years in both models, and as normally expected, the incidence of training (high accumulation state) is smaller than regular employment. However, the schooling distribution displays a higher degree of skewness in the comparative advantages model. In appendix Table A3, we report OLS regressions using wage outcomes measured at age 35. The OLS estimate for the common slope model, equal to 0.09, is well above the population return (0.06). This also confirms that the model is consistent with the popular notion of positive Ability Bias.

may be the case, for instance, in a model where treatment takes the form of a short training course for the unemployed. If so, an unconditional approach is more likely to be reliable than if treatment is time consuming.

Table 1
Some Descriptive Statistics

	Model			
	Common Slope		**Comp. Advantages**	
	Mean	**St dev.**	**Mean**	**St. dev**
Schooling	6.5	5.3	6.5	5.8
Work	38.1	13.1	37.2	14.6
Training	5.4	8.6	6.3	9.3
Total # periods	50	-	50	-

Table 2
The Distribution of Schooling by Level
in the Control Groups

	Model	
	Common Slope	**Comp. Advantages**
Schooling	**Percentage**	**Percentage**
No schooling	9 %	21%
Level 1 (grade1-4)	39%	30%
Level 2 (grade 5-8)	6%	6%
Level 3 (grade 9-12)	26%	20%
Level 4 (grade 13-16)	20%	23%

5.2 Designing Policy Interventions

We now describe the policy interventions that are used to generate instruments.

5.2.1 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

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

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

5.2.2 Education Subsidy

Each education subsidy consists of offering a reward conditional on attaining a specific grade. We consider different timings of the subsidies, as well as variations in the amount of the subsidy. To do so, 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 per-period subsidies of 1, 2, and 3 dollars. To implement the subsidies, we simply need to adjust the relevant utility parameters (α_1^s, α_2^s, α_3^s or α_4^s)

5.2.3 Control Groups

In total, and for the common slope model, we generate 28 different treatment groups (16 compulsory schooling interventions, and 12 education subsidies). So, for each case, we compute the following quantities:

$$\begin{aligned} \{(d_{k1i}^{m\Upsilon}, w_{1i}^{m\Upsilon}), (d_{k2i}^{m\Upsilon}, w_{2i}^{m\Upsilon}), ...(d_{k50i,}^{m\Upsilon} w_{50i}^{m\Upsilon})\} \ i &= 1, 2, ...50, 000, \\ k &= s, e, a \\ \Upsilon &= 1, 2, ., 28, control \end{aligned}$$

5.2.4 Ex-Ante Randomization

For all policies, and by construction, pre-intervention randomization is fulfilled. That is the distribution of individual fixed endowments (the model primitives that are individual specific) is invariant to policy exposure:

$$H_{\nu_i}(. \mid Z_i = 1) = H_{\nu_i}(. \mid Z_i = 0) \tag{7}$$

where $Z_i = 1$ for treatment, $Z_i = 0$ for control.

6 Interpreting Moment Conditions

In this section, we re-interpret the identifying moment (orthogonality) condition, as an ex-post condition that restricts behavioral differences between those affected and those unaffected. There are at least two good reasons to do it.

First, most realistic policy interventions set in a context where individuals choose among a discrete set of options, can only affect a subset of the population. So, any econometrician implementing an IV approach is bound to have priors about the identity of those affected (and unaffected). It is therefore natural to express the identifying assumptions in terms of the complying status.

Second, in the applied treatment effect literature, the most favored interpretation of IV estimates (the LATE parameter) is itself explicitly based on the complying status.

As will become clear later, the orthogonality condition can be decomposed in two mutually exclusive conditions that restrict post-treatment choices. We also introduce a third ex-post notion of randomization, that is based on individual endowments, as opposed to choices, and which is neither necessary nor sufficient for identification. All three notions of randomization are based on complying status.

6.1 Moment Condition and Complying Status

As a starting point, assume that the econometrician wants to estimate λ^s by IV, using Z_i as an instrument, and that t is given (fixed). Re-write the wage equation as

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

where

$$\varphi_{it} = \lambda^e \cdot E_{it} + \lambda^a \cdot A_{it} + \varepsilon_{it}^w \tag{8}$$

The error term, φ_{it}, depends on individual time-invariant endowments (υ_i), on accumulated schooling, and on calendar time (age) t. Orthogonality between Z_i and the model primitives (υ_i), which we referred to as ex-ante randomization, is no longer sufficient to generate orthogonality of Z_i with respect to φ_{it}. Using standard vector notation, the IV estimator is defined as

$$\hat{\lambda}_{s,IV} = (Z'S)^{-1} Z'W \tag{9}$$

where W is a vector of log wages. At any date t, the IV bias is equal to $\frac{\sum Z_i \varphi_{it}}{\sum Z_i S_{it}}$.

In order to introduce the complying status in the analysis, we write the vector of error terms of the treatment and control groups as

$$\varphi_t = \{\varphi_t^T, \varphi_t^C\}$$

where the superscript T refers to treatment, and C refers to control. Denote the total number of observations in treatment and control by N^T and N^C, and the total number of observations by N ($N = N^T + N^C$).

Second, re-write φ_t^T, as

$$\varphi_t^T = \{\varphi_t^{Ta}, \varphi_t^{Tu}\}$$

where φ_t^{Ta} represents the sub-vector of those who are affected by the policy, and where φ_t^{Tu} is the sub-vector of those unaffected. Denote the number of individuals affected and unaffected by N^{TA} and N^{TU}, where $N^T = N^{TA} + N^{TU}$.

Because the treatment/control indicator is binary (0,1), the identifying moment condition may be written as

$$\begin{aligned} & E\frac{\sum Z_i \varphi_{it}}{N} \\ = \;& E\frac{\sum \varphi_{it}^T}{N^T} \\ = \;& E\;\frac{\sum \varphi_{it}^{Ta}}{N^{TA}} + E\;\frac{\sum \varphi_{it}^{Tu}}{N^{TU}} \\ = \;& 0 \end{aligned}$$

There are therefore two distinct ways to obtain the orthogonality condition.

Case 1: *Post-Treatment Choice Orthogonality (with Respect to Complying Status)*

It arises when those affected and those unaffected share the same average φ_i, at date t. This is the case when

$$E\frac{\sum \varphi_{it}^{Ta}}{N^{TA}} = E\frac{\sum \varphi_{it}^{Tu}}{N^{TU}} = 0 \tag{10}$$

Case 2: *Post-Treatment Opposite Choice (with Respect to Complying Status)*

A policy is characterized by post-intervention *Opposite Choice* when those affected and those unaffected have opposite values of φ_{it}, at date t. This is the case when

$$E\frac{\sum \varphi_{it}^{Ta}}{N^{TA}} = -E\frac{\sum \varphi_{it}^{Tu}}{N^{TU}} \tag{11}$$

Note that one of these two conditions must be met for the orthogonality condition to be fulfilled.[15] The re-expression of the moment condition, as a condition restricting complying and non-complying behavior, allows us to introduce the identity of those affected (and unaffected) into the analysis.

However, the fact that we condition on t hides a major complication. As one moves from period to period (as t changes), both φ_{it}^{Ta} and φ_{it}^{Tu} change,

[15] In empirical work, and to our knowledge, empiricists never argue in favor of orthogonality conditions using ideas along the lines of opposite choice.

even though the identity of those affected and unaffected does not. Because each individual leaves school with no post-schooling skills, and assuming that the less able leave school earlier than the more able (something inherent to the dynamic skill accumulation model), it is always possible to find a period s at which $\varphi_{is}^{Ta} \approx \varphi_{is}^{Tu}$ and for any policy intervention. This is true even though the two groups may have completely different identities.[16] In other words, the identifying moment condition cannot be analyzed independently from the wage sampling design, because those affected not only have different skill endowments than those unaffected, but they also start accumulating post-schooling skills at different dates.[17]

This is a key issue in understanding the differences between compulsory schooling instruments and incentive-based instruments. It basically means that there exist two different ways to optimize the performance of IV. One way is to affect behavior by changing policy parameters, and therefore create a specific heterogeneity gap between those affected and those who are not. This is the focus of our analysis.

A second one is to choose a sampling period (or a distribution of sampling periods). For instance, suppose a given policy has been implemented and that the complier-non-complier heterogeneity gap has been identified, *Post-Treatment Choice Orthogonality* may be achieved by sampling wages in the period where the more able are about to over-take the less able in terms of post-schooling skills. However, if wages are sampled beyond this over-taking period, and if positive differences in time invariant ability already translate into positive differences in $\varphi' s$, it is no longer achievable. If orthogonality is met, it is through *Post-Treatment Opposite Choice* .

Because most cross-section data sets used in empirical work contain wages measured between age 25 and 65 (and often at an average age between 40 and 50), we disregard time sampling as an instrument of policy design. As will become clear later, our analysis will be conditioned on a sampling period that is representative of most cross-section data used by empirical labor economists. In practice, this implies avoiding sampling earnings in very early, or in very late phases over the life-cycle.

At this stage, there are three interesting implications that are worth ex-

[16] This period (age), denoting the time when the more able over-take the less able in terms of their post-schooling skills, is therefore close (but not identical) to the notion of over-taking age in Mincer's age-earnings profiles.

[17] Rosenzweig and Wolpin (2000) discuss related issues.

amined.

6.1.1 Post-Treatment Identity Randomization and Orthogonality Condition

First, we introduce a third notion of ex-post randomization. Denote the distribution of heterogeneity for those affected and unaffected by $H^A_{\nu_i}(.)$ and $H^U_{\nu_i}(.)$, where

$$H^A_{\nu_i}(.) = H_{\nu_i}(. \mid S_i(Z_i = 1) \neq S_i(Z_i = 0))$$

and where

$$H^U_{\nu_i}(.) = H_{\nu_i}(. \mid S_i(Z_i = 1) = S_i(Z_i = 0))$$

We say that *Post-Treatment Identity Randomization (with Respect to Complying Status)* is fulfilled when a policy intervention separates the population into two groups that have an identical heterogeneity distribution, that is when

$$H^A_{\nu_i}(.) = H^U_{\nu_i}(.) \tag{12}$$

As should be clear, *Post-Treatment Identity Randomization* is a notion that characterizes the set of individuals affected by a policy through model primitives, as opposed to choices.

The bottom line is that, at any period t, there is no guarantee that a policy that splits the population according to complying status into two identical groups, would indeed fulfill *Post-Treatment Choice Orthogonality* , let alone *Opposite Choice*. This follows from the fact that the time deficit due to schooling investment (the opportunity cost) may not exactly be compensated by the post-schooling accumulation stimulus induced by schooling. So, *Post-Treatment Identity Randomization* is neither necessary nor sufficient for the fulfillment of the identifying condition.

6.1.2 The Incidence of Negative vs. Positive IV Bias

These moment conditions also provide an intuitive interpretation for the incidence of high vs. low IV estimates, at least with respect to the population treatment effect. Namely,

$$E\frac{\sum \varphi_{it}^{Ta}}{N^{TA}} < -E\frac{\sum \varphi_{it}^{Tu}}{N^{TU}} \rightarrow E\sum Z_i\varphi_{it} < 0 \tag{13}$$

$$E\frac{\sum \varphi_{it}^{Ta}}{N^{TA}} > -E\frac{\sum \varphi_{it}^{Tu}}{N^{TU}} \rightarrow E\sum Z_i\varphi_{it} > 0 \tag{14}$$

Again, for realistic sampling periods, conditions (13) and (14) imply that, when estimating a common slope parameter, over-estimation (under-estimation) tends to arise when the set of individuals affected is increasingly located among the upper (lower) tail of the skill distribution. This will also have consequences for the distinction between compulsory schooling and incentive-based policy interventions.

As an example, a policy that affects the extreme lower tail of the factor distribution (a characteristic inherent to compulsory schooling), will exert negative bias if the average post-schooling skills of those affected (denoted $E\frac{\sum \varphi_{it}^{Ta}}{N^{TA}}$) are lower (more negative) than the opposite value of the post-schooling skills of those unaffected (denoted $-E\frac{\sum \varphi_{it}^{Tu}}{N^{TU}}$).

6.1.3 Weak or Strong Instrument?

One additional problem that arises when designing an intervention, is that the implementation of a policy that leads to a stronger instrument may not reduce IV bias. As one compares some initial policy intervention to a new policy characterized by a higher correlation between schooling and the instrument (as plim$\frac{Z'S(Z)}{N}$ increases), there is no guarantee that the new policy translates into a more accurate estimation of the treatment effect of interest, because plim$\frac{Z'\varphi(Z)}{N}$ may also change. This follows from the fact that φ_{it}, is made of unobserved individual choices and is generally not policy invariant.

This observation has also strong implications for policy design. As it is particularly unlikely that one can design a policy that meets orthogonality in a strict sense, it is highly relevant to keep track of the co-movements between the numerators and denominators of the asymptotic bias expression.

So, when designing policies in practice, the level of the correlation between the instrument and schooling may be viewed as practically as important as the correlation between the instrument and the error term. However, in a dynamic setting, the desirability of a stronger instrument is not automatic.

6.2 How to Design a Policy?

We are now in a position to give a heuristic approach to policy design. First, assume that a sampling period has been fixed to a realistic level (avoiding both very early and very late phases of the life-cycle). In the hypothetical case where the econometrician faces unlimited design possibilities, we already know that the best policies are those that either meet *Post-Treatment Choice Orthogonality*, or *Post-Treatment Opposite Choice.*

For the sake of the presentation, suppose that the Econometrician targets *Post-Treatment Choice Orthogonality.* The basic idea is to construct a policy for which the complier/non-complier heterogeneity gap (the difference in identity) is annihilated by the simultaneous (conflicting) effects of schooling on post-schooling skill formation and the opportunity cost of schooling.

As an example, if the model structure (parameters) implies that at date t, one can find two distinct groups for which the loss in potential experience is fully compensated by the reduction in the cost of accumulating skills, then orthogonality would be reached through *Identity Randomization*. This means that the compliers and non-compliers have same identity. However, if the experience loss and the stimulus do not compensate each other, it is necessary to calibrate the policy parameters so to create a heterogeneity gap between compliers and non-compliers.

For realistic dates of sampling, a policy that entails a huge identity gap (like mandatory schooling), is unlikely to fulfill *Post-Treatment Choice Orthogonality*, since the more able are likely to dominate the less able, in terms of post-schooling skills. It may, however, fulfill orthogonality through *Post-Treatment Opposite Choice* , if the time period is chosen far enough beyond the over-taking period, and if the skill gap is sufficiently large.

As mentioned earlier, it may be difficult to design a policy that meets orthogonality in a strict sense. When those "First Best" policies (those that fulfill *Choice Orthogonality* or *Opposite Choice)* do not belong to the choice set, the process is not really different, except that it requires an arbitrage between both the numerator and the denominator of the IV bias. In such cases, it is important to remember that choosing the best policy from a feasible set does not necessarily imply choosing the strongest one (statistically).

Finally, and in line with unconditional analyses advocated by statisticians, it is also possible to calibrate alternative policies so to incorporate the effect of the stimulus of schooling into the estimand. However, and as mentioned in Section 4, this total effect has no closed-form expression, and our approach is better illustrated by targeting the effect of schooling, conditional on post-schooling skills.

To illustrate the policy design exercise, we now turn to the empirical implementation of a wide range of IV estimates. To be realistic, we will illustrate the design exercise as a choice between a relatively limited number of possibilities.

7 IV Estimation

We now proceed with IV implementation. First, and 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). For each estimate, we compute a standard error using 100 bootstrap replications.

As is the case in the literature, our estimates are drawn from very large samples, and tend to have very small standard errors. However, it may be misleading to judge their validity based solely on whether or not the confidence interval covers the population parameter (0.06). Instead, for expositional purposes, we impose a tolerance level of half a percentage point (0.005), and consider, as reliable, those estimates between 0.055 and 0.065. As will become clear later, this has no consequence on the relative performance of compulsory schooling vs. education subsidies.

We present the policy design exercise as an iterative procedure. We first construct a set of feasible policies, and analyze how moving from one to another, improves or deteriorates the performance of IV.

As stated earlier we computed 12 education subsidies and 16 mandatory schooling policies. We implement subsidies of \$1, \$2 and \$3, at all four grade levels, as defined in 5.2.2. For illustrative purposes, we investigated minimum schooling age from 1 to 16 (even if setting a minimum age at a high level is rather unrealistic).

To ease presentation, we report estimates obtained for six different combinations of monetary amount and grade levels, and for 6 different mandatory schooling policies (the first 6 periods). More precisely, we report one subsidy of \$2 conditional on attending Level 1 (grade 1 to grade 4), one subsidy of \$1 conditional on attending Level 2 (grade 5 to grade 8), two different subsidies (\$1 and \$3) for Level 3 (grade 9 to grade 12) , and two subsidies (\$1 and \$3) for Level 4 (grade 13 to grade16).

Because we disregard issues surrounding the choice of a sampling period, and given our total time horizon of 50 years, sampling at age 35 happens to be more or less equivalent to the average age at which cross-sectional IV are measured in observational data.[18] So, all IV estimates reported are obtained for 100,000 realized wages, at age 35.

In Table 3, we summarize the main features of each policy intervention. Those include the fraction of population affected, and the corresponding correlation between schooling and the instrument (Z), the identity of the individuals affected and unaffected in terms of the average psychic cost of training, and their ex-ante level of schooling.

The IV estimates are found in Table 4. To illustrate the link between the moment conditions expressed in terms of the treatment/control, and those expressed in terms of the complying status, we also report the correlation between φ_i and Z, and the mean values of both $\varphi_t^{TA}$and φ_t^{TU} (after suitable normalization).

7.1 The Education Subsidies

To analyze the results, it is informative to consider a specific policy, and analyze what is happening when we move to a different one. To do so, we start with the education subsidy that grants a payment of \$1 conditional on attending Level 2. This intervention, labeled "Level 2-\$1" in Table 3 and Table 4, affects 16% of the population, and leads to a correlation between schooling and the instrument equal to 0.07.

In terms of the average factor (the psychic costs of training) of those affected (5.9) and unaffected (5.3), there is evidence that those affected are on average less able than those unaffected. However, there is a relatively big difference in terms of schooling (ex-ante) between those affected (4.2) and those who are not (6.9).

[18]In the typical IV literature on schooling and earnings, the average age is often between 40 and 45. We have also computed IV at a potentially different period (age) for each individual (using a uniform distribution), so to mimic standard cross-section data sets, but found practically identical results. So, for illustrative purpouses, it is clearer to sample wages at same period (age) for each individual.

The IV estimate for the Level 2-\$1 subsidy is 0.0503 (Table 4). Based on the criteria aforementioned, it is judged unacceptable. The issue is now to design an alternative policy that achieves a more reliable estimate. Following conventional wisdom found in the applied literature, one solution would be to design a "stronger" policy (one that increases the correlation between schooling and Z).

For instance, let's now consider a displacement of the subsidy to lower grade levels (grade 1-4), and let's double the payment. By setting it to a lower grade level, it circumvents part of the effect of discounting. By raising the amount, it could also induce more lower skill individuals to stay in school longer.

This intervention, labeled "Level 1-\$2" in Table 2, affects 48% of the population, and leads to a correlation between schooling and the instrument equal to 0.11. Therefore, it generates a stronger instrument, in a statistical sense. However, as is evident upon examining the IV estimates and φ^{Ta}, φ^{Tu}, and $Corr(Z, \varphi)$ in Table 3, moving to this policy deteriorates the performance of IV, as the estimate is now equal to 0.0465. To see why, it is sufficient to examine the identity of those affected, as well as their skills accumulated beyond schooling. Despite the large increase in the denominator of the asymptotic bias, the even larger identity gap in training costs between those affected (7.2) and those who are not (4.3), translates into an even larger difference between φ^{Ta} (equal to -0.0367) and φ^{Tu} (0.0214). As a consequence, the correlation between Z and φ (-0.0686) is 2 times larger than it was with the Level 2-\$1 subsidy (-0.0366). Put differently, the complier-non-complier heterogeneity gap created by the Level 1-\$2 policy is too large, at least to fulfill *Choice Orthogonality.*

A thorough examination of all IV estimates based on education subsidies indicate that, among all those cases considered, the subsidies implemented between grade level 9 and 12 are performing the best. For instance, with the Level 3-\$1 subsidy, we obtain an accurate estimate of the population treatment effect, as the estimate is 0.0609. With an amount of \$3, the estimate (0.0554) is also quite reasonable, given our tolerance level. When compared to the Level 2-\$1 subsidy, the improvement is explained by a parallel increase in $Corr(S, Z)$, and a decrease in $Corr(Z, \varphi)$. Indeed, the Level 3-\$1 dollar intervention approaches the orthogonality condition, as $Corr(Z, \varphi) = 0.0098$. This policy, unlike the preceding one, is therefore able to create a reasonable identity gap between those affected and those who are not. Note however that neither the Level 3-\$1 policy, nor the Level 3-\$3 policy, fulfill the orthogonality condition in a strict sense.

Finally, just like implementing a subsidy at the lowest grade levels did not work, the implementation of a subsidy at grade level 13-16 does not generate a good instrument either. First, because the benefit of higher education is damped by discounting, both Level 4 subsidies tend to lead to weaker instrument, when compared to subsidies of equal amount, but implemented at lower grade levels. For instance, only 14% of the population reacts to the payment of a \$1 subsidy to attend the highest level of education (Level 4), while 21% do so if the subsidy is implemented at Level 3 instead.

These upper education subsidies have another interesting particularity. Both the \$1 and \$3 subsidies affect mostly high ability individuals, as indicated by the average factor and the ex-ante education of Table 2, and as a consequence, lead to large positive values (of the order of 0.08) of the differential between $\varphi^{Ta} - \varphi^{Tu}$. As was already observed when analyzing equation (14), this is the reason why those IV estimates (0.0756 and 0.0738) exceed the population treatment effect.

7.2 Compulsory Schooling

Obviously, designing compulsory schooling policies is a more trivial task, as choosing the required number of years (school leaving age) is the only degree of freedom. To be realistic, we analyze compulsory schooling regulations up to 6 years.

As is evident upon consulting Table 2, both the fraction of people affected and the average ability of those affected is an increasing function of the minimum school leaving age.

At any realistic grade level (say 1 to 3 years), mandatory schooling policies can be labeled as inherently selective policies. For instance, even a policy that would force individuals to stay in school for a minimum of 3 years, which would affect 37% of the population, would translate into a complier-non-complier heterogeneity gap (7.97 vs. 4.39) that would be higher than for any of the subsidies considered in the analysis.

This is a clear consequence of the lack of flexibility that characterizes any dictatorial policy that raises minimum consumption/investment. To see this, it is particularly informative to compare the mandatory schooling policies with the Level 1 subsidy. Both of those classes of interventions target low grade levels, but only education subsidies have the capacity to avoid extreme differences between those affected and those who are not.

The poor performance of mandatory schooling IV is striking. The estimates range between 0.02 (1 year policy) to 0.0488 (6 years). Obviously, in this particular model, raising the minimum school-leaving age by only 1 year, would generate a very weak instrument, and would therefore generate a degree of imprecision that is too high. However, from 2 years onward, the degree of precision is substantially increased. As the minimum number of years is increased, the degree of non-orthogonality between treatment and control deteriorates, but the improvement of IV accuracy, is largely due to the increase in the correlation between schooling and Z.

These results are easily explained. Realistically, an instrument generated by compulsory schooling can only achieve orthogonality through condition (11). However, as opposite selection is likely to arise in presence of a statistically weak instrument (because it is associated with extreme tails of the skill distribution), it is much harder to achieve than orthogonality with respect to complying status. In practice, the econometrician has no design freedom.

One other interesting aspect of mandatory schooling IV estimates, is that they are bound to underestimate the treatment effect, at least for realistic minimum school leaving age. This is simply explained by the fact that mandatory schooling regulations are characterized by very small values for φ_{it}^{Tu} (as seen in equation 15).

7.3 Comparing Compulsory Schooling with Education Subsidies

Our results provide a good illustration of the difficulties inherent to finding the "optimal" design. Given the model, and for the specific wage sampling period, none of our incentive-based policies fulfill *Choice Orthogonality* or *Opposite Choice.* Wages are sampled too late for *Choice Orthogonality,* and too early for *Opposite Choice.* In the end, the "best performing" instrument among the various policies which we implemented, is found by matching the best combination possible for the numerator and the denominator of the IV bias. The result was satisfactory for some of the education subsidies, but not for compulsory schooling.

To summarize, there is a fundamental asymmetry between dictatorial and incentive-based regulations. Because subsidies can be set at different grade levels, and because their monetary values can be adjusted freely, policies may be designed so to reduce the asymptotic IV bias at a sufficiently low level. Flexibility is not a characteristic of dictatorial policy interventions. Compulsory schooling offers practically no design possibilities. Both the strength of the instrument generated by the policy, and the identity of those affected, are bound to the distribution of ex-ante schooling choices. Furthermore, the minimum targeted by most reforms, is rarely raised by more than one year at the time.

Going back to the original questions stated in the Introduction, we can now offer relatively clear answers. Within a non-trivial dynamic skill accumulation, the capacity to fulfill the IV identifying orthogonality condition is related to the nature of the policy intervention used to generate an instrument. Precisely, when offered the possibility to design an optimal policy, and given a realistic sampling procedure, the econometrician should never choose a dictatorial (compulsory schooling) policy.

Table 3
Characteristics of various Policy Interventions
Common Slope Model

	% affected	$Corr(S,Z)$	Ave. training cost affected	Ave. training cost unaffected	Educ. Ex-ante affected	Educ. Ex-ante unaffected
Subsidies						
Level 1-\$2	48%	0.11	7.2	4.3	1.7	10.9
Level 2/\$1	16%	0.07	5.9	5.3	4.2	6.9
Level 3/\$1	21%	0.05	4.3	6.1	9.6	5.7
Level 3/\$3	42%	0.17	4.6	6.5	8.3	5.2
Level 4/\$1	14%	0.02	3.7	6.0	12.6	5.5
Level 4/\$3	37%	0.09	3.8	6.8	12.3	3.13
mandatory						
1 year	9%	0.01	9.62	5.41	0.00	7.16
2 years	25%	0.03	8.72	4.73	0.62	8.42
3 years	37%	0.07	7.97	4.39	1.08	9.69
4 years	49%	0.14	7.42	4.07	1.55	11.27
5 years	52%	0.22	7.3	4.0	1.6	11.7
6 years	53%	0.24	7.3	4.0	1.7	11.7

Note: Without loss of generality, the psychic cost of training is used to illustrate the identity of those affected and unaffected.

Table 4
IV Estimates for Various Policy Interventions
Common Slope Model

	IV (st.error)	φ^{Ta}	φ^{Tu}_{it}	$Corr(Z, \varphi)$
Subsidies				
Level 1/$2	0.0465 (0.0016)	-0.0367	0.0214	-0.0686
Level 2/$1	0.0503 (0.0012)	-0.0389	0.0126	-0.0366
Level 3/$1	0.0609 (0.0029)	-0.0426	0.0114	0.0098
Level 3/$3	0.0554 (0.0009)	-0.0582	0.0357	-0.0336
Level 4/$1	0.0756 (0.0068)	0.0849	-0.0113	0.0208
Level 4/$3	0.0738 (0.0016)	0.0632	-0.0240	0.0803
mandatory				
1 year	0.0200 (0.0238)	-0.0124	0.0008	-0.0190
2 years	0.0419 (0.0053)	-0.0213	0.0028	-0.0317
3 years	0.0462 (0.0024)	-0.0297	0.0095	-0.0510
4 years	0.0479 (0.0013)	-0.0388	0.0216	-0.0804
5 years	0.0486 (0.0009)	-0.0490	0.0289	-0.1150
6 years	0.0488 (0.0007)	-0.0535	0.0308	-0.1298

Note: Standard errors are computed using 100 bootstrap replications.

8 Reconciling our Results with the Literature on Compulsory Schooling

The objective is now to show that our analysis may shed light on the incidence of low (and even negative) IV estimates reported in the recent literature on compulsory schooling.

As noted earlier, within a dynamic skill accumulation model, mandatory schooling policies are naturally associated to negative IV bias. However, the existence of a negative IV bias is not sufficient in itself to generate negative IV estimates. To reconcile low IV estimates with a dynamic skill accumulation model, we basically need to introduce comparative advantages in the model.

Consistent with what is reported in the recent structural literature that allows for comparative advantages in schooling (Carneiro, Hansen and Heckman, 2003, and Belzil and Hansen, 2007), we assume that a non-trivial mass of individuals is endowed with treatment effects of education that are inferior to their return to post-schooling activities.[19]

To achieve this, we introduce heterogenous effects of schooling in the analysis, while keeping the average at the same level of the common slope model (0.06), and raise the effect of employment on earnings, λ^e, to 0.02 . In a single factor model, low skill individuals practically never choose training. In line with Belzil and Hansen (2007), we use a distribution of λ^s that ranges from 0.005 to 0.11, and that averages to the same value used in the common slope model, namely 0.06.

The distribution may be found in Table A1 (in appendix). To be conservative, we manipulate the heterogeneity distribution so that the mass of individuals (types) that are endowed with a value of λ^s below or equal to 0.02 (the effect of employment on earnings) is equal to 15%. We thereby introduce a very mild degree of comparative advantages.

In Table 5, we report IV estimates obtained from the same compulsory schooling regulations considered in the common slope framework. Because the wage equation is in line with the popular heterogenous treatment effect literature, the natural treatment effect to consider is the average effect of

[19] In Carneiro, Hansen and Heckman (2003), a certain proportion (close to 10%) of the population experience ex-post returns to schooling that are negative. In Belzil and Hansen (2007), more that 20% of the population is found to have values of the ex-ante treatment effect of schooling that are inferior to their early career post-schooling earnings growth.

schooling for those affected. For this reason, it is reported in Table 5, even if the prime objective of this section is simply to investigate the capacity of our model to generate very low IV estimates.

As noted earlier, the fraction of the population affected by mandatory schooling is a monotonic function of the minimum schooling level. It ranges from 21% (1 year) to 56% (8 years).

More importantly, and as suggested by the recent literature, IV estimates for compulsory schooling regulations are close to 0, and may even be negative for policies that correspond to a mild increase in minimum school leaving age (up to 3 years). The estimates are equal to -0.01 (1 year), -0.0033 (2 years), -0.0023 (3 years), and remain low even at a 8 years mandatory schooling regulation (0.0143). Note that the negative IV estimates (for the first 3 years) arise even if the average treatment effect of schooling for the compliers are positive (0.01 for 1 year, 0.02 for 2 years, and 0.03 for 3 years). There was indeed no need to introduce negative values for λ^s in the model in order to generate negative IV estimates.[20]

Finally, it should be reminded that the incidence of negative IV estimates has been obtained with a very mild degree of comparative advantages. Had we assumed a non-degenerate distribution of effects of employment on wages (for instance by introducing a second factor), it would be possible to generate even more pronounced negative IV estimates. This would be possible, even though returns to skill accumulation are strictly positive for every individual. Although stylized, our model of life-cycle skill accumulation is perfectly capable of explaining the discrepancy in reported IV estimates.

Returning to our objective, which is to provide an intuitive interpretation for the incidence of negative IV estimates reported in the compulsory schooling literature, we can now illustrate how the presence of comparative advantages may explain the negativity of those IV estimates. To do so, we shall now refer to the fundamental economic differences between dictatorial and incentive-based policy interventions, which were mentioned in Section 2.

When policy designers implement mandatory schooling, individual decisions are annihilated for a very selective subset of the population. So, at any given point in time, the amount of post-schooling skills accumulated by those

[20] As noted in a common slope framework, mandatory schooling IV are still unable to estimate the relevant treatment effect parameter. The gap between the IV estimate and the population treatment effects remains between 0.02 and 0.03. Expressed as a percentage of the population treatment effect, these estimates entail an enormous bias.

affected by the policy is reduced. This automatically creates a negative correlation between post-schooling skills and treatment/control status. Those individuals who are endowed with low returns to academic skills (those for whom the treatment effect of schooling is smaller that the treatment effect of work experience), are forced to delay post-schooling skill accumulation. So, in this model complexion, not only do compulsory schooling reduce post-schooling skill accumulation (as in a common slope model), they may also reduce wages since for those affected, the treatment effect of employment exceeds the treatment effect of schooling. This arises in our model since the effect of schooling on the incidence of training is sufficiently low.

Consequently, for those endowed with relatively high returns to work experience, the policy may entail a reduction in lifetime earnings. This was not the case in the common slope model, even if imposing minimum schooling age always reduce net utilities. This is a logical implication of compulsory schooling regulations, which basically annihilate comparative advantages for a non-random subset of the population.

Table 5
IV Estimates for Compulsory Schooling:
Comparative Advantages

	Fraction Affected	$Corr(S, Z)$	**Population Treat. Effect**	**IV(st.error)**	$Corr(Z, \varphi)$
1 year	21%	0.0183	0.0109	-0.0100 (0.0181)	-0.0262
2 years	32%	0.0477	0.0218	-0.0033 (0.0067)	-0.0684
3 years	41%	0.0805	0.0300	-0.0023 (0.0036)	-0.1263
4 years	50%	0.1474	0.0362	0.0094 (0.0021)	-0.2068
5 years	52%	0.2185	0.0366	0.0131 (0.0014)	-0.2972
6 years	53%	0.2602	0.0367	0.0122 (0.0012)	-0.3466
7 years	54%	0.3140	0.0374	0.0127 (0.0010)	-0.4044
8 years	56%	0.3806	0.0386	0.0143 (0.0008)	-0.4668

Note: The population treatment effect is the average value of λ^s for those affected by compulsory schooling. This parameter is usually called the Local Average Treatment Effect (Imbens and Angrist, 1994).

9 Concluding Remarks

Because of its relative degree of flexibility, education subsidies are capable of generating an instrument that may be used to estimate a policy invariant treatment effect. On the other hand, compulsory schooling regulations offer practically no design possibilities. Both the statistical strength of the instrument generated by the policy, and the identity of those affected, are tightly bound to the distribution of ex-ante schooling choices. This is a logical implication of compulsory schooling regulations, which basically annihilate comparative advantages for a non-random subset of the population.

Going back to the original questions stated in the Introduction, when offered the possibility to design an optimal policy, and for realistic sampling designs, the econometrician should never choose a dictatorial (compulsory schooling) policy.

Although stylized, our calibrated model of life-cycle skill accumulation may actually reconcile the well known discrepancy between IV estimates obtained from different policy interventions. It sheds light on the incidence of the low (and possibly negative) IV estimates that have recently been reported in the literature, and provide a straightforward economic explanation for it.

Our analysis raises the following question. Given that policy interventions are used as instruments in other circumstances, can our results be extrapolated to other areas? While we cannot bring a fully rigorous answer, we conjecture that the answer is yes.

Within the education literature, a increasingly large number of economists are investigating the effects of raising parents' education, on children's education. While this sort of question may be answered within a dynamic (intergenerational) skill accumulation, it is also customary to estimate this effect using IV techniques, and to rely on mandatory schooling reforms. Clearly, the impact of raising education of a minority of low skill individuals through some compulsory regulation may differ substantially from the impact of raising the average level of education of a more representative population.

The literature on estimating the returns to work experience is another area where our analysis may be relevant. Indeed, a literature on estimating the effect of labor market experience loss, using various types of dictatorial policy interventions, is also currently re-emerging[21] Incentive policies (such as tax reform) stimulating the decision to work and dictatorial policies that reduce work experience for a subset of the population (such as military/civil service reforms) are likely to give rise to the same issues discussed in this paper.

References

[1] Abbring, Jaap and Heckman, James (2007) "Econometric Evaluation of Social Programs, Part 3: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium

[21] See Imbens and van der Klaauw (1995), for an analysis of Dutch conscription, and Chib and Jacobi (2011) for a few citations of recent (and on-going) work devoted to it.

Policy Evaluation", *Handbook of Econometrics, Volume 6B.* Elsevier B.V.

[2] Adda, Jerôme, Dustmann, Christian, Meghir Costas and Robin, Jean-Marc (2006) "Career progression and formal versus On-the-Job Training", IZA Discussion paper 2260.

[3] Bellman, Richard (1957) "*Dynamic Programming*" Princeton, New-Jersey, Princeton University Press.

[4] 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

[5] Belzil, Christian (2010, in press) "Testing the Specification of the Mincer Wage Equation" *Annales d'Economie et de Statistiques*, vol 91-92

[6] Belzil, Christian and Hansen, Jörgen (2002a) "Unobserved Ability and the Return to Schooling" *Econometrica, 70, 575-91.*

[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] Ben Porath, Yoram (1967): "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy*, 75(4), pp. 352-365.

[9] Cameron, S. V. and Christopher Taber (2004) "Estimation Of Educational Borrowing Constraints Using Returns To Schooling" *Journal of Political Economy*, 112, 132-182.

[10] Card, David (1999) "The Causal Effect of Education on Earnings" Handbook of Labor Economics, edited by David Card and Orley Ashenfelter, North-Holland Publishers.

[11] 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.

[12] Chib, Siddhartha and Liana Jacobi “Returns to Compulsory Schooling in Britain: Evidence from a Bayesian Fuzzy Regression Discontinuity Analysis”, IZA Discussion Paper 5564.

[13] Cunha, F., J. Heckman and Salvador Navarro (2005) ”Separating Uncertainty from Heterogeneity in Life Cycle Earnings” NBER Working Paper 11024.

[14] Cunha, F., J. Heckman and Schennach, S. (2010) “Estimating the Technology of Cognitive and Noncognitive Skill Formation ” IZA Working Paper 4702.

[15] Chari, V.V., Kehoe, Patrick, and Ellen R McGrattan (2007) “Are Structural VAR’s with Long Run Restrictions” Useful in Developping Business Cycle Theory?, Federal Reserve Bank of Minneapolis, Report 364

[16] 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

[17] Eberwein, C, Ham, J.C. and Lalonde, R.J. (1997) “The Impact of Being Offered and Receiving Classroom Training on the Employment Histories of Disadvantaged Women: Evidence from Experimental Data

[18] Devereux, Paul 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, 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.

[20] Gill, R.D. and Robins, J.M. (2001) “Causal Inference for Complex Longitudinal Data: The Continuous Case”. *The Annals of Statistics 29 (6),* 1785-1811 (December)

[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] Imbens, Guido and Angrist, Joshua (1994) "Identification and Estimation of Local Average Treatment Effects" *Econometrica, 62, 467-76.*

[27] Imbens, Guido (2009) "Better Late than Nothing" , NBER Working Paper

[28] Imbens, Guido and van der Klaauw (1995) "Evaluating the Cost of Constription in The Netherlands", *Journal of Business and Economic Statistics,* 13(2), 207-215.

[29] Keane, Michael (2010) "Structural vs. Atheoretic 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] Lechner, M. and Miquel R. (2002) "Identification of Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions". Discussion paper, University of St-Gallen, Department of Economics

[32] Magnac, Thierry and Thesmar, David (2001) "Identifying Dynamic Discrete Decision Processes" *Econometrica, 70, 801-16.*

[33] Maurin, Eric, and Theodora Xenogiani (2007) "Demand for Education and Labor Market Outcomes: Lessons from the Abolition of Compulsory Conscription in France." *Journal of Human Resources 42(4)*: 795–819.

[34] Meghir, Kostas and Marten Palme (2005) "Educational Reform, Ability, and Parental Background" *American Economic Review, vol 95 (1), 414-424.*

[35] Murphy, S.A. (2003) "Optimal Dynamic Treatment Regimes" *Journal of the Royal Statistical Society, Series B65 (2).* 331-366

[36] Nielsen, Helena Skyt, Torben Sørensen and Christopher Taber "Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform", IZA Discussion Paper 3785

[37] Oosterbeek, Hessel and Webbink D. (2007) "Wage Effects of an Extra Year of Basic Vocational Education" *Economics of Education Review, 26, 408-419*

[38] 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*

[39] Rosenbaum, (1984) "The consequences of Adjustment for a Concomitant Variable that has been affected by the Treatment", *Journal of the Royal Statistical Society*, Series A 147 (5), 656-666.

[40] Rosenzweig Mark and K.Wolpin (2000) "Natural Natural Experiments in Economics" *Journal of Economic Literature*, December, 827-74.

[41] Stokey, N., Lucas, R.E. (with Ed Prescott) (1989). Recursive Methods in Economic Dynamics. Harvard University Press. Cambridge, Massachusetts.

[42] Vytlacil, Edward (2002) "Independence, Monotonicity, and Latent Index Models: An Equivalence Result", *Econometrica, 70(1): 331-341*

Table A1
The Heterogeneity Distribution

	α^S	λ_s	c_0^a	proportion
type				
1	-2.990	0.001	9.895	0.05
2	-2.750	0.001	9.955	0.10
3	-1.790	0.015	8.695	0.10
4	-1.310	0.023	8.215	0.10
5	-0.470	0.035	7.375	0.05
6	0.130	0.044	6.775	0.05
7	0.550	0.050	6.355	0.05
8	0.844	0.055	6.061	0.05
9	1.324	0.062	5.581	0.05
10	1.570	0.066	5.335	0.05
11	2.050	0.073	4.855	0.05
12	2.530	0.080	4.375	0.05
13	2.770	0.084	4.135	0.05
14	3.010	0.087	3.895	0.05
15	3.250	0.091	3.655	0.05
16	3.484	0.094	3.421	0.05
17	3.964	0.102	2.941	0.05
Mean	1.188	0.060	5.717	-
St Dev.	2.062	0.031	2.062	-

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

Accumulated number of periods in each State

	Common Slope Model			Comparative Advantages Model		
	In School	Work (learning)	Work (Training)	In School	work (learning)	Work (Training)
year						
1	0.906	0.094	0.00	0.791	0.209	0.000
10	5.475	4.525	0.010	5.267	4.733	0.010
20	6.494	12.615	0.891	6.503	12.554	0.943
30	6.494	22.006	1.500	6.503	21.914	1.583
40	6.494	29.564	3.492	6.503	29.150	4.347
50	6.494	38.119	5.387	6.503	37.185	6.311

Table A3
OLS Regressions

	Model	
	Common Slope	Comparative Advantages
intercept	0.410	0.563
	(0.013)	(0.012)
education	0.088	0.113
	(0.0003)	(0.0002)
experience	0.037	0.036
	(0.001)	(0.001)
experience^2	-0.0003	-0.0002
	≈ 0	≈ 0
R^2	0.50	0.52