The distinction between dictatorial and incentive policy interventions and its implication for IV estimation

Christian Belzil, J. Hansen

▶ **To cite this version:**

Christian Belzil, J. Hansen. The distinction between dictatorial and incentive policy interventions and its implication for IV estimation. 2010. hal-00463877

HAL Id: hal-00463877
https://hal.science/hal-00463877

Preprint submitted on 15 Mar 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ECOLE POLYTECHNIQUE

CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

THE DISTINCTION BETWEEN DICTATORIAL AND INCENTIVE POLICY INTERVENTIONS AND ITS IMPLICATION FOR IV ESTIMATION

Christian BELZIL
Jörgen HANSEN

March 2010

Cahier n° 2010-08

DEPARTEMENT D'ECONOMIE
Route de Saclay
91128 PALAISEAU CEDEX
(33) 1 69333033
http://www.enseignement.polytechnique.fr/economie/
mailto:chantal.poujouly@polytechnique.edu

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

March 14, 2010

*We would like to thank Steve Durlauf, Chris Ferrall, Bo Honore, Susumu Imai, Arnaud Maurel, James McKinnon and Chris Taber for comments and discussions. We thank seminar participants at Ecole Polytechnique, University of Wisconsin, Queen's University, McMaster, Concordia University, and the First French Econometrics Conference (Toulouse). The usual disclaimer applies.

Abstract

We investigate if, and under which conditions, the distinction between dictatorial and incentive-based policy interventions, affects the capacity of Instrument Variable (IV) methods to estimate the relevant treatment effect parameter of an outcome equation. The analysis is set in a non-trivial framework, in which the right-hand side variable of interest is affected by selectivity, and the error term is driven by a sequence of unobserved life-cycle endogenous choices. We show that, for a wide class of outcome equations, incentive-based policies may be designed so to generate a sufficient degree of post-intervention randomization (a lesser degree of selection on individual endowments among the sub-population affected). This helps the instrument to fulfill the orthogonality condition. However, for a same class of outcome equation, dictatorial policies that enforce minimum consumption cannot meet this condition. We illustrate these concepts within a calibrated dynamic life cycle model of human capital accumulation, and focus on the estimation of the returns to schooling using instruments generated from mandatory schooling reforms and education subsidies. We show how the nature of the skill accumulation process (substitutability vs complementarity) may play a fundamental role in interpreting IV estimates of the returns to schooling.

Key Words: *Returns to schooling, Instrumental Variable methods, Dynamic Discrete Choice, Dynamic Programming, Local Average Treatment Effects .*

JEL Classification: B4, C1, C3.

1 Introductory Remarks

Consider the following econometric problem. A researcher seeks to estimate an outcome equation using Instrumental Variable (IV) techniques. To do so, he/she maintains the hypothesis that the equation contains only one right-hand side endogenous variable, and an error term. However, the error term is generated by a sequence of unobservable (latent) choices, which are exercised after the realization of the endogenous variable. Suppose that the econometrician has to choose between one instrument generated by a dictatorial policy intervention, and another one generated by an incentive-based policy reform. The dictatorial policy enforces a minimum consumption/investment level of the endogenous right-hand side variable, while the incentive policy implements a subsidy conditional on achieving a certain consumption/investment level.[1]

This paper addresses the following questions. Is the nature of the intervention relevant in assessing the capacity for IV to recover the relevant treatment effect parameter? If so, which type of intervention is more likely to fulfill the identifying orthogonality condition? Is the underlying data generating process (whether the mis-specification is caused by a sequence of neglected static discrete choices, or a sequence of neglected dynamic discrete choices) relevant? If so, for which class of outcome equation structure is the distinction relevant?

Although the differences between dictatorial and an incentive-based policy interventions will be made more explicit later, at this stage it is sufficient to recognize that a subsidy may be characterized by three elements; 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). A typical dictatorial intervention does not offer as much design flexibility. We illustrate these concepts within a calibrated dynamic life cycle model of human capital accumulation, and focus on the estimation of the returns to schooling using two popular types of interventions; namely a mandatory schooling reform, and a set of education subsidies.

In the first portion of the paper, we present an economic analysis of the

[1]Throughout the paper, we restrict ourself to the case where the econometrician estimates only one treatment effect, and disregards the option of estimating treatment effects associated to residual choices.

effects of dictatorial and incentive-based policy interventions. As a second step, and using simulated data on wages, schooling and policy exposure, we implement an IV estimation strategy of the returns to schooling, and stress the relative performance of instruments generated by various education subsidies as opposed to compulsory schooling regulations.

The main contribution of this paper is methodological. To our knowledge, this is the first paper that considers the distinction between dictatorial and incentive-based interventions within a stochastic dynamic programming framework, and that translates the distinction into formal economic concepts.

Although econometricians have stressed the fact that different instruments may generate different population treatment effects, discussions about IV remain based on explicit identifying orthogonality conditions.[2] Our paper is also the first that examines the implications of the economic nature of a policy intervention, for IV estimation. To do so, we introduce the notion of "Post-Intervention Randomization" (PIR), and define it as the degree of selection on individual endowments among the sub-population affected). Throughout the paper, we stress its importance for IV estimation.

Until now, the literature on the economics of IV estimation has concentrated mostly on three issues. First, there is a substantial literature that debates the relevance of estimating Local Average Treatment effects.[3] Second, some recent papers have debated the distinction between internal and external validity, in relation with the concept of randomization.[4] Thirdly, many economists have concentrated on the role that economic models may play in interpreting orthogonality conditions.[5]

[2]In the treatment effect literature, it is customary to consider outcome equations that are multiplicative in the error term. In such a case, it is widely recognized that different experiments generate different Local Average Treatment Effects (LATE). The LATE parameter is defined in Imbens and Angrist (1994) in an IV context. Björklund and Moffitt (1987) and Heckman and Vytlacil (2005) introduced the notion of marginal treatment effect, which generalizes the LATE parameter. Finally, Belzil and Hansen (2007) estimate LATE parameters within a stochastic dynamic programming model.

[3]See Heckman, Urzua and Vytlacil (2005) and Imbens (2009) for opposite views about the relevance of the LATE parameter.

[4]The distinction between internal and external validity is discussed in Todd and Wolpin (2005). Deaton (2008) and Imbens (2009) present opposite views on the desirability of ex-ante randomization.

[5]See Rosenzweig and Wolpin (2000), and Keane (2007), who argue that orthogonality conditions usually hide a large number of implicit assumptions. Belzil (2007) discusses

The absence of discussion of the economic nature of various policy interventions is exemplified in the returns to schooling literature. Compulsory schooling reforms are systematically used to estimate returns to schooling, while incentive-based education policies are much less popular.[6] Interestingly, both types of policy interventions are treated symmetrically. 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 a same data generating process, and as a consequence, the econometrician practically never gets to choose between those two types of intervention. So, our analysis therefore sheds light on the vast empirical literature devoted to returns to schooling.

The main results are as follows. For a wide class of models (including static and some dynamic specifications of the outcome equation), and because of its relative flexibility in policy design, an incentive-based intervention may be capable of generating a small degree of selection on individual endowments among the sub-population affected. Dictatorial policies that enforce minimum consumption (unless set at unrealistically high levels approaching the maximum consumption level) cannot fulfill this condition. This feature translates into a much higher capacity of incentive-based policies to generate an instrument that fulfills the identifying orthogonality condition, and that enables the econometrician to estimate a particular treatment effect. In the returns to schooling example, our findings apply both (i) when schooling has no causal impact on post-schooling choices after conditioning on heterogeneity (a static outcome equation), as well as when (ii) schooling stimulates post-schooling skill accumulation (a dynamic skill complementarity model). However, when schooling counterbalances the effect of heterogeneity (a form of dynamic skill-substitutability model), nothing can be said.

Within the particular model structure that we investigate, IV estimates obtained from compulsory schooling regulations lie well outside the support of the distribution of the returns to schooling (in the negative orthant), and are significantly far from the population treatment effect. At the same time, a properly designed education subsidy paid conditional on reaching the in-

similar issues in the sole context of the returns to schooling literature.

[6]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).

termediate schooling level, translates into an IV estimate falling within an arbitrarily small distance of the relevant population treatment effect parameter. However, this is not true for all education subsidies. For instance, a lower-education subsidy generates an instrument that performs almost as badly as a compulsory schooling instrument.

The residual parts of the paper are organized as follows. In Section 2, we present some important definitions, as well as some background material. In Section 3, we lay-out the behavioral model. The 4th section is devoted to the calibration exercise. In the 5th section, we illustrate the fundamentals of the model by summarizing choices and outcomes of the control group. The economic analysis of each policy intervention is found in Section 6. In Section 7, we introduce the notion of post-intervention randomization, and discuss its implications for IV estimation in Section 8. In Section 9, we proceed with IV estimation and discuss the main results. Section 10 is devoted to an intuitive discussion. Finally, in the last section, we discuss the implications of our analysis for empirical work, and identify other areas of micro-econometrics where a similar analysis is likely to be relevant.

2 Background Material

The set of policy interventions used in various fields of economics is large, but it is informative to consider two distinct classes; those that are dictatorial in nature, and those that are based on incentive provisions.

A dictatorial intervention affects individual decisions by restricting the choice set. In this paper, we focus on policy interventions that impose a minimum consumption/investment level. 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 the 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 consuming a pre-determined level of a choice variable. In a sequential framework (when consumption takes time), 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 is typically available to all individuals. Before reaching the relevant consumption/investment level, 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.

At the outset, it should be clear that it is possible to assign all policy experiments to a single class of incentive interventions if, for instance, we allow the cost (subsidy) of choosing a particular option to approach plus (or minus) infinity. Also, nothing prevents the existence of hybrid policy interventions, which would be characterized by both some dictatorial aspects, and some incentive-based provisions.[7] Finally, it is also possible to imagine dictatorial policies that may be bypassed at some cost. However, we believe that the dichotomy between dictatorial and incentive-based policies is nevertheless highly relevant at an empirical level.

The use of IV methods is based on the assumption that the degree of (pre-intervention) randomization offered by a specific experiment is sufficient to uncover a parameter of interest.[8] In a context where the error term represents time invariant heterogeneity unobserved to the econometrician (but exogenous from the perspective of the agent), IV methods are particularly appealing, and the economic nature of the intervention has no effect on the accuracy of IV. This is true even if the nature of the intervention has an impact on the treatment effect parameter that may be estimated.

However, in an intertemporal (sequential) framework, where the error term of the outcome equation of interest may incorporate subsequent unobserved choices taking place beyond policy implementation, pre-intervention randomization is not sufficient to uncover quantities of interest. This may arise in a context where individuals continue to optimize on the basis of time

[7] In theory, natural experiments (such as disasters and the like) may also be represented as policy interventions that are (i) either affecting some key economic parameters, or (ii) restricting individual choice sets. As do policy interventions, natural experiments may also have hybrid components.

[8] This statement is, to some extent, contentious since many applied econometricians apply IV without appealing (explicitly) to orthogonality conditions. In such a case, they define the ratio of the correlations that define IV, as a population parameter of interest (treat the estimator as the estimand), and interpret it as a (weighted) policy effect.

invariant heterogeneity. It may also arise when subsequent choices are driven by heterogeneity, as well as past choices.

The first case implies that the error term of the outcome equation is a function of a sequence of static discrete choices. Individuals move across states regardless of past choices, and the correlation between current and subsequent choices is totally spurious. The second case delivers a life-cycle dynamic discrete choice framework, in which individual subsequent choices are correlated with previous choices, even after conditioning on tastes and skills.[9]

Because IV methods are typically implemented on observational data, it is unrealistic to rule out potential mis-specification of these sorts. This argument would be particularly true in the returns to schooling literature, since human capital investments, are by definition, unobservable.

This raises a fundamental issue. If dictatorial and incentive interventions reserve a different role to comparative advantages (and random shocks) in determining the composition of the sub-population affected by it, the type of intervention that systematically affects a more selective sub-population, may also as a by-product, generate more differences between subsequent choices of those affected by the intervention and subsequent choices of those unaffected. If so, the intervention that affects a more selective group is less likely to generate the realization of the identifying orthogonality conditions. Given this premise, it is natural to investigate the relative performances of instruments generated by dictatorial policy interventions, and those generated by incentive-based provisions.

To do so, we cast our analysis within a life-cycle skill accumulation (potentially dynamic) model, and our analysis focuses on the implications of using either a dictatorial or an incentive-based policy intervention in order to obtain point estimates of the return to schooling.[10] Because educational

[9]The reader should note that we use the term "static" and "dynamic" to qualify the nature of the outcome equation. As is going to be clear later, the models that we calibrate are all based on the assumption that agents are "forward looking". However, the issues that we analyze in this paper would still be relevant even if agents were myopic.

[10]The choice of a dynamic skill accumulation model is only instrumental. Similar argumentation could be built on non-competitive labor market structures such as search/matching models (Adda, Dustman, Meghir and Robin, 2005), or dynamic incentive models of promotion and wages. However, the interest in dynamic skill accumulation models, which dates back to Ben-Porath (1967), has been revived by Keane and Wolpin (1997) and Heckman, Lochner and Taber (1998). It also plays a key role in the recent

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.

The choice of a life-cycle accumulation theoretical background is not innocuous. The return to schooling is one of the most frequently investigated parameters in modern economics, and the majority of estimates reported in the empirical literature are obtained in an IV framework. Our analysis is centered on an artificial population of forward looking heterogeneous agents making decisions between schooling, work, human capital accumulation, and household production activities. The behavioral model is calibrated to a set of well recognized stylized facts (or common conjectures) about life cycle human capital accumulation, and is presented in the next section.

3 The Behavioral Model

Our desire is to generate population moments that may characterize observational data on schooling, work, training, household production and earnings.[11] To do this, we select a certain number of population characteristics. Most of them are usually regarded as stylized facts.

In total, we selected the following 7 attributes:

1. Schooling should account, on average, for approximately one sixth of the total time horizon.

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).

literature on child development (Cunha, Heckman and Schennach, 2010).

[11]For instance, the artifical panel data that we generate could resemble the NLSY (one of the most popular data sets used in the structural literature on human capital).

4. Labor Market employment (either the sum of full-time work and work with training) must be the most common choice over the life cycle

5. Household production must be a relatively rare event

6. OLS regressions of simulated wages on education should produce a higher return than the average in the population.

7. There must be a positive correlation between schooling, and individual returns to schooling

The first characteristic would apply to most advanced countries (such as the US, Canada and Europe). The second characteristic is also observed in most countries, and arises in any finite horizon model (Ben-Porath, 1967). The third characteristic is an indirect implication of the declining incidence of productive investment.

The fourth and fifth are particularly relevant for a population of males.

The sixth may be somewhat more controversial. It is sometimes referred to as the classical "ability bias" hypothesis. It implies that the correlation between wages and schooling is an over-estimate of the true effect of schooling on productivity. This is not a stylized fact. Indeed, it is a pure conjecture. We choose a positive (as opposed to a negative) bias because of its intuitive aspect. It would obviously be possible to define the model structure differently, or to modify the dynamic structure, so to imply a negative OLS bias. However, this would have no implication for our analysis.[12]

Finally, the seventh condition ensures that there is positive selection. In other words, those who are endowed with more academic skills also tend to be (on average) more educated. This condition guarantees that the econometrician faces a non-trivial statistical problem. Without it, the outcome equation would be a classical random coefficient regression where the slopes (the individual returns to schooling) are orthogonal to the regressor (schooling).

[12]There may be other features of life cycle wages that may be occasionnally cited, but as pointed out recently in Heckman, Lochner and Todd (2006) and Belzil (2010, forthcoming), many widely accepted features of the standard Mincer wage equations are rejected when tested formally.

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 33 periods to allocate between the 4 mutually exclusive states. The states are the following; schooling (S), work with a low rate of skill accumulation (E), work with a high rate of skill accumulation (A), and Household Production (H). The corresponding variables (S_t, E_t, A_t, H_t) are used to measure the number of periods accumulated in each state. There is a maximum of 12 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). We assume that the utility of school changes with grade level and we consider 3 distinct levels; 1 to 4, 5 to 8, and 9 to 12. Throughout the analysis, we refer to the first level as Basic Schooling, the second level (5 to 8) as the intermediate level, and the third one as Higher Education.

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.

3.2 School

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

$$U_{it}^s = \alpha_i^s - \alpha_1^s \cdot I(S_t \geq 5)) - \alpha_2^s \cdot I(9 \leq S_t)) \ - \alpha_3^s \cdot (t - S_t) + \varepsilon_{it}^S \qquad (1)$$

where $I(.)$ is the indicator function. The parameters α_1^s,and α_2^s capture the higher direct costs of schooling faced by those who enrol in college. These parameters reflect tuition costs and the like. The parameter α_3^S captures the psychic cost of attending school for those who would have interrupted their education (the length of interruption is $t - S(t)$). 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 Household Production

The utility of household production, U_{it}^h, is given by the following expression

$$U_{it}^h = \alpha_i^h + \varepsilon_{it}^h \tag{2}$$

where α_i^h is individual specific utility of household activities and ε_{it}^h is a stochastic shock.

3.4 Employment and Training

The utility of work and learning, U_{it}^e, and the utility of work with training, U_{it}^a, are constructed as the difference between the wage rate and the monetary costs of occupying a specific state. Precisely, U_{it}^e and U_{it}^a, and their related costs, $C_{it}^e()$ and $C_{it}^a()$, are given by the following equations;

$$U_{it}^j = W_{it} - C_i^j(S_{it}, H_{it}, E_{it}, A_{it}) \text{ for } j = e, a \tag{3}$$

$$C_{it}^j() = c_{0i}^j + c_{1j} \cdot S_{it} + c_{2j} \cdot H_{it} + c_{3j} \cdot E_{it} + c_{4j} \cdot A_{it} \ + \varepsilon_{it}^j \text{ for } j = e, a \tag{4}$$

where c_{1j}, c_{2j}, c_{3j} and c_{4j} are parameters capturing the effect of accumulated schooling, home time, employment and training on the cost (or disutility) of work, or work and training. They illustrate the dynamics of skill accumulation (skills beget skills). The $\varepsilon_{it}^{j\prime}s$ are stochastic shocks.

The distinctions between static and dynamic models arise from alternative parametrization of equation (4). When both c_{1a} and c_{1e} equal 0, schooling choices have no impact on future utilities, and we obtain the static version of the model. When both c_{1a} and c_{1e} are negative (positive), we obtain the skill-complementarity (skill-substitutability) model.[13]

3.5 Market Productivity

The reward to human capital investment is embedded in the following wage equation

[13]We have defined complementarity (and therefore substitutability) in terms of the dynamic effect of schooling (one input) on the total cost of accumulating skills. Obviously, it would also be possible to model it in terms of total skills.

$$\log W_{it} = w_{it} = \alpha^w + \lambda_i \cdot S_{it} + \delta_i \cdot E_{it} + \theta_i \cdot A_{it} + \varepsilon_{it}^w \tag{5}$$

where W_{it} is the wage rate per unit of time, α^w is the intercept term, and $\lambda_i, \theta_i, \delta_i$ are individual specific returns to schooling, learning, and training.[14] Altogether the vector $\{\beta_i, \delta_i, \theta_i\}$ summarizes individual labor market skills. ε_{it}^w is the stochastic term affecting earnings.

3.6 The Bellman Equations

The choices are summarized in the binary indicators, d_{tk}. Precisely, $d_{tk} = 1$ when option k (w, h, w, a) is chosen. 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 k, there is a choice specific value function, $V_t^k(\Omega_t)$, equal to

$$V_t^k(\Omega_t) = U_t^k + \beta E \max\{V_{t+1}^1(\Omega_{t+1}), ..V_{t+1}^K(\Omega_{t+1}) \mid d_{kt} = 1\} \tag{6}$$

or, more compactly, as

$$V_t^k(\Omega_t) = U_t^k \ + \beta E V_{t+1}(\Omega_{t+1} \mid d_{kt} = 1) \tag{7}$$

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

3.7 The Distribution of Individual Heterogeneity and Random Shocks

- The heterogeneity vector is distributed as a multi-variate discrete distribution with 20 vectors of support points;[15]

$$\nu_r \sim \{\alpha_r^S, \alpha_r^H, \lambda_r, \delta_r, \theta_r, c_{0r}^a, \beta_r; p_r\} \text{ for } r = 1, 2, ..20 \tag{8}$$

  where p_r is the population proportion of type r.

[14] Although it would also be possible to allow the intercept term to be individual specific, the rich multiplicative heterogeneity structure makes it redundant.

[15] The heterogeneity structure is sufficiently rich that we do not even need to introduce an individual specific (or type specific) psychic cost of choosing employment (c_0^e).

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

3.8 Model Solution

As is relatively common in the literature, we solve the Bellman equations using simulated realizations of the random shocks. The Bellman equations need to be solved for each single type separately. Our solution method is exact in the sense that we do not use any approximation or interpolation methods.

4 Calibration of the Model

Because it would be tedious to describe all parameters separately, we present the general philosophy that underlies our choices. A set of parameters describing the heterogeneity components is found in appendix (Table A1). The correlations are found in Table A2.

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.

4.1 Outcome Equation and skills

The distribution of returns to schooling is centered at 0.06 (a value close to estimates reported in the structural literature). However, we allow for a high degree of dispersion. The support of the distribution of returns ranges between 0.00 and 0.12 (see Table A1). These numbers therefore reflect estimates reported in both the structural and the IV literature.[16]

The average returns to learning (0.01) and training (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

[16] A detailed comparison between structural and IV approach is found in Belzil (2007).

will lie between 1% and 2% per year (a well known stylized fact for the US). We treat the utility of school, the cost of on the job training and the return to education as driven by an academic skill, and enforce a perfect correlation between these components. However, to deviate from a trivial ability bias structure, we assume that the return to work experience may be driven by skills that may be non academic, and enforce a weak correlation between these two components, and the other academic heterogeneity components (the utility of school, the cost of on the job training and the return to education).[17]

4.2 Post Schooling Dynamics

Our version of the dynamic skill accumulation is in the spirit of Ben-Porath (1967).[18] The parameters $\{c_{1a}, c_{2a}, c_{3a}, c_{4a}\}$ and $\{c_{1e}, c_{2e}, c_{3e}, c_{4e}\}$ are capturing the effect of accumulated schooling, home time, employment and training on the cost (or disutility) of work/training (A) and work (E). The vectors are equal to {-0.5, 0.0, 0.0, 0.0) for state A, and {-0.3, 0.0, 0.0, 0.0) for state E. The parameter values for c_{1a} (-0.50) and c_{1e} (-0.30) imply that accumulated schooling reduces both the cost of investing in human capital and the cost of labor market work. The larger effect of education on the cost of training is a reflection of the academic nature of the work/training activity. The non-negative values for c_{3e} and c_{4e} allow us to introduce some dynamics in the decision to work. The null values for c_{2a} and c_{2e} imply the absence of skill depreciation.

4.3 Preference Heterogeneity and Discount Rates

To reflect preference heterogeneity, we allow discount rates to differ across individuals. They range between 0.00 and 0.10. The average discount rate

[17] In Belzil and Hansen (2002), the correlation is above 0.9. In Keane and Wolpin (1997), the correlation between the utility of attending school and white collar skills would also be very high.

[18] 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 general equilibrium version of the Ben-Porath model. More recently, Adda, Dustmann, Meghir and Robin (2005) have implemented a dynamic model of Apprenticeship and on-the-job training based on partial equilibrium search/matching arguments.

(0.05) is standard. This form of preference heterogeneity may also be reinterpreted as a way to approximate the effects of liquidity constraints. It is important to do so, because in the IV literature, the high IV estimates are often conjectured to arise because they reflect the LATE parameters of a subpopulation of individuals affected by liquidity constraints, or of a subpopulation of individuals who have high discount rates. As economics offers no guidance for the choice of a correlation between discount rates and individual skills, we started the calibration procedure by imposing quasi orthogonality between discount factors and other heterogeneity components, and adjusted the correlations in order to match population characteristics. In other words, and as opposed to the correlations between various market skills and the costs of training and schooling, we regarded the correlations between discount rates and other heterogeneity components as secondary. Indeed, they do not play a key role in our analysis.

4.4 Heterogeneity vs. Ex-Ante Risk

In order to calibrate the model, we must implicitly choose the relative importance of heterogeneity (cross sectional dispersion in skills) vs. ex-ante risk (the variance of the random shocks affecting the outcome equation). This is difficult. The structural literature on dynamic discrete choices always assumes that individual specific heterogeneity is known by the agent, and that random shocks are not. While the issue has only started recently to raise interest, it is too early to establish a consensus.[19] For this reason, we relied on estimates reported in Belzil and Hansen (2007), who estimated a correlated random coefficient wage regression, and set the standard deviations of all random shocks to 0.5.

5 The Control Groups

To generate each control group, we simulate 33 years of choices and wage outcomes for 50,000 individuals. In our analysis, an individual is defined as the conjunction of (i) a heterogeneity type and (ii) a specific history of random shocks. Precisely, we simulate 2500 different sets of random shocks

[19] See Cunha and Heckman (2005) for a discussion.

for each type (we have 20 types). For each model, we construct of a panel data set that records choices and wages as follows:

$$\begin{aligned}\{(d^m_{k1i}, w^m_{1i}), (d^m_{k2i}, w^m_{2i}), ...(d^m_{k33i,} w^m_{33i})\}\ i &= 1, 2, ...50,000 \\ k &= s, h, e, a \\ m &= ST, DSC\end{aligned}$$

where the index ST refers to the static model, and DSC refers to the dynamic skill-complementarity model.

We provide a summary of individual choices in Table 1A (dynamic skill-complementarity), and Table 1B (static). To do so, we computed the average number of accumulated periods in each state for each model for period 1 to period 10, as well as periods 15, 20, 25,30, and 33. The complete distribution of life-cycle schooling attainments is found in Table 1C**.** Overall, the frequencies display the desired features that we advocated in Section 4. For instance, schooling is chosen mostly in the first 10 years. This is true for both models. Home production is rarely chosen as it accounts for less than 10% of total time allocation. The average schooling attainment is equal to 5.6 years in the skill complementarity model and 4.0 years in the static model.

In total, individuals spend between 21 and 23 years in market production. Moreover, the incidence of training (high accumulation state) is higher in the skill complementarity model (5.4 years), than in the static model (4.2 years).

In order to illustrate the determinants of schooling, employment and training decisions in the control group, we regressed the total number of periods in each state on all relevant individual endowments. Because some of the endowments are perfectly correlated, we only include those that are linearly independent.[20] These regressions are summarized in Table 2A, and Table 2B. To facilitate comparison, each endowment has been standardized.

Aside from the fact that realized schooling are highly (positively) correlated with both individual returns to schooling and discount factors, it is important to note that, in the dynamic skill-complementarity model, around 80% of schooling decisions are explained by heterogeneity (individual skills and preferences). The fraction of total employment and training decisions explained by heterogeneity are also comparable (84% and 83% respectively)**.**

[20]For instance, the psychic costs of schooling, α^S, and the psychic cost of work/training, c^a_0, are not linearily independent.

The results obtained for the static version of the model indicate a practically identical role for heterogeneity and ex-ante risk.

Obviously, as individual tastes and skills are never fully observed, it is impossible to assert that this heterogeneity/risk decomposition is indeed realistic. Nevertheless, it appears to be consistent with results reported in Cuhna and Heckman (2005) and Belzil (2006).[21]

To illustrate the resulting relationship between wages, education, and experience (for instance, to check concavity of age earnings profiles), we performed OLS regressions of log wages on education, and potential experience (using a cubic polynomial) for each control group. The results are in Tables A3 and A4. The OLS estimates (above 0.10) are in all cases higher than the population average return. Further, the parameter estimates for the effect of experience indicate that age earnings profile are concave. The estimates imply an average growth rate of 1% per year of potential market experience.[22]

6 Characterizing the Policy Interventions

6.1 The Design

Our definition of a mandatory schooling intervention is standard. It increases average schooling in the population by setting a minimum age (period) for leaving school. To be realistic, this mandatory level is below (or at most equal to) the minimal number of periods needed to obtain the ex-ante average educational attainment in the population. In the paper we consider two different dictatorial interventions; a two-year and a four-year compulsory schooling policy.

Each education subsidy consists of offering a reward conditional on attaining a precise grade level. To illustrate the design flexibility of incentive

[21] Cuhna and Heckman (1986) report that between 50% and 80% of the econometrician error term is actually heterogeneity (known to the agent). Belzil (2006), using a completely different method, reaches a similar conclusion.

[22] Although we did not want to impose any specific relationship between education and age-earnings profiles, we noted that our model implies a positive effect for the interaction between education and experience. This would be the case, for instance, with data taken from the NLSY.

policies, we consider education subsidies implemented at different grade levels, and of unequal duration.

All policy interventions considered in the paper are described below.

- **Two-Year mandatory education:** The first two years are spent in school ($d_{s1i} = d_{s2i} = 1 \forall i$)
- **Four year mandatory education:** The first four years are spent in school ($d_{s1i} = d_{s2i} = d_{s3i} = d_{s4i} = 1 \forall i$)
- **Higher Education Subsidy:** A 2$ payment is awarded upon reaching grade level 9. This subsidy is paid for each period of school attendance until the maximum grade level (11) is reached.
- **Intermediate Education subsidy**: A 2$ payment is awarded upon reaching grade level 5. This subsidy is paid for each period of school attendance until the maximum grade level (11) is reached.
- **Basic Education subsidy:** A 2$ payment is awarded from grade 1 to grade 4 (inclusively). The subsidy is terminated after completion of grade 4.[23]

Altogether, this set of policy interventions are representative of the degree of flexibility (or lack thereof) inherent to each class of intervention. For instance, the basic education subsidy is naturally compared to the four year mandatory schooling regulation, because payment is made over the same period covered by the mandatory schooling. However, the monetary level of the subsidy can be manipulated so to avoid affecting the lower tail.

The Intermediate and higher education subsidies work differently. These subsidies are claimed upon reaching a relatively higher grade level, and are therefore not expected to affect the lower tail of the distribution. At a purely intuitive level, it is clear that the benefit of the intermediate and the higher education subsidies is largely damped by individual discounting. For instance, basic education subsidies should be much less affected by discounting, since the benefit is perceived already in period 1.

[23]For each case (intermediate and higher education), we actually changed the amount of the subsidy. The low, median and high subsidies were respectively equal to 1$, 2$, and 3$. However, to simplify the presentation, we report results obtained from the median subsidy.

Compulsory schooling regulations, on the other hand, have for only flexibility the possibility to increase the minimum number of periods in school. In presence of positive selection, and as the minimum number of periods is increased, the policy affects a population that is increasingly representative. However, for realistic compulsory levels, these interventions are bound to affect a set of individuals that comprise a large fraction of very low ability individuals.

In total, we generate 5 different treatment groups (2 compulsory schooling interventions, and 3 education subsidies) for each model:

$$\begin{aligned} \{(d_{k1i}^{m\Upsilon}, w_{1i}^{m\Upsilon}), (d_{k2i}^{m\Upsilon}, w_{2i}^{m\Upsilon}), ...(d_{k33i,}^{m\Upsilon} w_{33i}^{m\Upsilon})\}\ i &= 1, 2, ...50,000, \\ k &= s, h, e, a \\ m &= ST, DSC \\ \Upsilon &= 1, 2, 3, 4, 5 \end{aligned}$$

For each combination of intervention and model, we define the individual specific counterfactual reaction to policy intervention Υ, given model m, which we denote $R_{it}(\Upsilon; m)$, as

$$R_{it}(\Upsilon; m) = S_{it}\ (\Upsilon; m) - S_{it}\ (control; m) \tag{9}$$

Consistent with our definition of an individual, the $R_{it}(\Upsilon; m)'s$ are measured while holding fixed each individual specific random shock history. To avoid confusion, we will omit the time subscript in $R_{it}(\Upsilon)$, since we want to focus on total schooling attainments. So, in practice, we define $R_i(\Upsilon; m)$ as simply $S_{i,33}\ (\Upsilon; m) - S_{i,33}\ (control; m)$.[24]

With a vector of individual counterfactual reactions for each specific intervention, we can now investigate the properties of the conditional distribution of these reactions; namely

$$F_{R_i}(r_i(\Upsilon; m) \mid \nu_i) \equiv F_{R_i}(r_i(\Upsilon; m) \mid \lambda_i, \delta_i, \theta_i, \beta_i, \alpha_i^S, \alpha_i^H, c_{0i}^a)$$

So, as we move across policy interventions, it is also possible to assess the corresponding change in the degree of dependence between the $r_i(\Upsilon; m)$'s and the full set of individual characteristics.

[24]In practice, the parameter representing the psychic cost of returning to school ensures that accumulated schooling remains fixed from period 13 onward.

6.2 An Economic Analysis

The main effects of the policy interventions are summarized in Table 3. In order to relate the effects of the policy changes to the primitive objects of the model (individual endowments), we regressed the counterfactual individual reactions to policy intervention (the $R_i(I)'s$) on all relevant individual endowments (the full vector of skills and tastes). The regressions are summarized in Table 4A, and Table 4B. They allow us to determine which individual skill or which preference component is more important in explaining individual reactions to policy changes.

Because of its intrinsic counterfactual nature, computing the LATE requires to hold the vector of realized random shocks constant for each individual. For this reason, they are computed using the same realizations of the random shock vector that were used to generate the control group.[25] In other words, we evaluate the counterfactual effects of the new policy by computing individual decisions of the control group under the old regime, as well as under the new regime, while holding individual random shock histories constant. This guarantees that the Monotonicity condition will hold.

To present our economic analysis, it will be useful to select a specific set of measures, and to evaluate the differences as well as the commonalities of the subsidy and the mandatory schooling experiments. We focus our presentation on 3 aspects; the fraction of the population affected, the labor market skills of the population affected (the LATE of education, work/training and work/learning, and ex-ante education), and the effect of the discount rate on the individual reactions to policy interventions (because it often plays a key role in the interpretation of IV estimates). All these measures are either found in Table 3, or in Table 4A, or Table 4B.

6.2.1 The Density of the sub-populations affected

As normally expected, the summary statistics of Table 3 reveal that the intermediate level subsidy affects a higher fraction of the population than the subsidy paid conditionally on reaching grade level 9. This is true regardless of the model structure.

[25] Because the LATE parameters are computed from 50,000 realizations of the vector of random shocks, their sampling variability turn out to be very small. As a consequence, we treat it as a population parameters.

As an example, for the dynamic skill-complementarity model, the intermediate level subsidy affects 24%. When the subsidy is delayed until grade 9, the fraction of the population affected drops to 9.1%.

As mentioned earlier, the lower education subsidy is of a different nature. It pays individuals to attend school during the first 4 years only, as opposed to the intermediate level subsidy. Despite all this, it affects more individuals than the intermediate subsidy (around 26% of the population is affected by the basic schooling subsidy). However, this may be easily explained by the fact that the basic school subsidy is less affected by discounting.

In comparison, 25.8% of the population is affected by the two-year mandatory schooling regulation. This fraction is very close to those observed for the intermediate level and the lower education subsidies. Not surprisingly, the four-year mandatory schooling affects a larger fraction of the population (34.6%) than the two-year compulsory schooling reform. This fraction should naturally be compared to the fraction of individuals affected by the lower education subsidy (26.2 %). Although very high in relation to expected earnings, the 2$ subsidy is not sufficiently high to affect all those who are affected by the 4-year compulsory regulation. As indicated earlier, this illustrates the difference in design flexibility between compulsory schooling and education subsidy.

Turning to the static version of the behavioral model, we observe a generally lower fraction of individuals affected by the subsidies, along with a higher fraction of people affected by the compulsory schooling regulation. These features are easily explained by the relatively lower average level of education generated by the static model, since the incentive to obtain education is reduced.

6.2.2 The skills of the sub-population affected

Again, we start with the dynamic skill-complementarity model structure. The LATE of education in the intermediate level subsidy is equal to 0.0717, and indicates that those affected tend to have relatively high skills, but are not too far from the population average. This tendency is also mirrored in the level of pre-intervention schooling, which is between 6 and 7 years, and therefore also not too far from the population average ex-ante level of

schooling (around 6 years).[26]

When the implementation of the subsidy is delayed until one is faced with the choice of entering grade level 9, the policy has the obvious implication of affecting an even more selective sub-population. The counterpart of this increase in selectivity is mirrored in the average returns to skills of those affected. The education Late parameters exceeds 0.08 (it is 0.0809). Not surprisingly, the higher education subsidy is more effective for those who would have reached an even higher level of schooling (when compared to the intermediate level subsidy). Precisely, the average pre-intervention level of schooling of those affected is around 8.4 years.

We now consider the opposite, and ask what happens when the subsidy is implemented for the first 4 years only (the basic schooling level). Intuitively, this policy should reach a larger proportion of low skill individuals than both the intermediate and the higher education subsidies. This is exemplified by the education LATE parameter (0.0369) and the ex-ante average education (2.7 years).

As logically expected from the mandatory schooling interventions, the average returns for those affected is much below the population average. Precisely, the local average treatment effect for the two-year regulation is equal to 0.0303, and it is equal to 0.0376 for the four-year policy. However, the most striking difference is at the level of the pre-intervention schooling, which is equal to 1.11 years of schooling in the two-year regulation, and 2.89 for the 4 year regulation. These numbers are much lower than those just reviewed for the subsidy interventions.[27] As noted earlier, this is an unavoidable feature of the mandatory schooling intervention.

The conclusions remain unchanged when the same policies are applied to the static model structure, even though it implies lower average schooling in the population, because the incentives to get educated are weaker. However, in relative terms, the distinction between education subsidies and the mandatory schooling interventions remain unchanged.

To summarize, mandatory schooling policies always affect a sub-populations that are endowed with low skill levels, as well as relatively low returns to

[26]Recall that the average level of schooling in the control group is between 5 and 6 years.

[27]The average is above 2.0 simply because our model allows individual to leave and return to school. So, it is possible for an individual who would have obtained more than 2 years ex-ante, to be affected by the dictatorial intervention. However, as indicated by the average (2.11), this type of behavior is relatively infrequent.

skill. Education subsidies, on the other hand, are flexible enough to affect both low-skill and high-skill individuals.

6.2.3 The Role of the Rate of Time Preference

To get a formal picture of the impact of the rate of time preference (the only fundamental preference parameter in our model), it is useful to examine the regressions of the counterfactual individual changes in schooling on individual endowments (Table 4A, Table 4B).

As earlier, we focus on the dynamic skill-complementarity model, and return to the static model later. For the intermediate education level subsidy, the simulations indicate that counterfactual schooling reactions decrease (increase) with the discount factor (rate). In order to explain it, it is useful to visualize two distinct effects that play at the same time. The effect of the education subsidy works through the future component of the utility of attending school. This effect, other things equal, must increase with the discount factor. In other words, those who value the future most, are also most likely to react to the subsidy experiment. However, when the subsidy is implemented early enough, it may also affect a pool of individuals who have lower (higher) discount factors (rates). So, in total, the correlation between individual reactions and discount rates is ambiguous. In our model, the negative effect of discount factors appears to be dominant, when the subsidy is implemented from grade 5 onward (second column, Table 4A) and when it is implemented in the first 4 years only (first column, Table 4A).

However, as we delay the education subsidy until grade level 9, the negative effect should gradually vanish. This is indeed what may be inferred upon examining the third column of Table 4A. Delaying the subsidy until higher education has changed the relationship between discount factors and individual reactions, as the regressions indicate that individual increases in schooling are increasing (decreasing) with the discount factor (rate). In other words, those who value the future most, are also most likely to react to the subsidy experiment.

Finally, and not surprisingly, counterfactual changes decrease (increase) with the discount factor (rate) within the compulsory schooling experiment. This is easily explained. Other things equal, those would are low educated ex-ante (and who are affected by mandatory schooling), also tend to have lower discount factor (higher discount rates). In other words, those who tend

to get very low levels of schooling ex-ante tend to discount the future heavily. This is true for both the two-year and the four-year compulsory reforms.

These results translate almost perfectly into the static model framework (Table 4B). That is, aside from the subsidy implemented at level 9, we observe that counterfactual changes decrease (increase) with the discount factor (rate).

7 Defining Post-Intervention Randomization

We now turn to the notion of PIR. Broadly speaking we define it as the degree of selection on individual endowments among the sub-population affected. Although it is one of the dimensions that help characterizing different types of policy interventions, it also has immediate impact for IV estimations. For this reason, we devote a separate section to it.

7.1 Definition

In the empirical IV literature, pre-intervention randomization is what practically defines "exogeneity" of a given policy intervention. In our context, Pre-Intervention Randomization guarantees that

$$F_{\nu_i}(. \mid Z_i = 1) = F_{\nu_i}(. \mid Z_i = 0) \tag{10}$$

In other words, the distribution of individual fixed endowments are invariant to policy exposure. For the sake of presentation, it is informative to distinguish between the distribution of heterogeneity endowments of those affected, $F^a_{\nu_i}(.)$, and those unaffected $F^u_{\nu_i}(.)$. More precisely,

$$F^a_{\nu_i}(.; \Upsilon, m) = F_{\nu_i}(. \mid r_i(I) > 0, \Upsilon; m)$$

$$F^u_{\nu_i}(.; \Upsilon, m) = F_{\nu_i}(. \mid r_i(I) = 0, \Upsilon; m)$$

For any policy intervention Υ (given a model m), we define (perfect) "PIR" as the realization of the following condition:

$$F_{\nu_i}(. \mid r_i(I) = 0, \Upsilon; m) = F_{\nu_i}(. \mid r_i(I) > 0, \Upsilon; m) \tag{11}$$

7.2 An illustration with a simple binary choice model

To illustrate the difficulty of achieving PIR, it is informative to consider a simplistic case in which schooling decision is binary ($S = 1$ or $S = 0$), and in which heterogeneity is a scalar. The CDF of θ is denoted $F(.)$ and its density $f(.)$.

Formally, we have

$$S_i = I(\theta_i > 0)$$

A dictatorial intervention Υ^d splits the region into two distinct sets; Θ^{ud} and Θ^{ad}, where Θ^{ud} denotes the set of values of θ of those unaffected by Υ^d, and Θ^{ad} denotes the set of values of θ of those affected. It follows that

$$\begin{aligned} \Theta^{ad} &= \{\theta \mid \theta \leq 0\} \\ \Theta^{ud} &= \{\theta \mid \theta > 0\} \end{aligned} \tag{12}$$

and that $\max(\theta) \mid \theta \in \Theta^{ud} \leq \min(\theta) \mid \theta \in \Theta^{ad}$. The two sets are disjoint:

$$\Theta^{ad} \cap \Theta^{ud} = \emptyset$$

and it is impossible to equate $F^u_\theta(.;\Upsilon^d)$ to $F^a_\theta(.;\Upsilon^d)$. Indeed, it is not even possible to equate the first moments since, by definition,

$$\begin{aligned} E\{\theta \mid \theta \in \Theta^{ad}\} &\equiv \int_{-\infty}^{0} \frac{f(\theta_i)\theta d\theta}{F(0)} \neq \\ E\{\theta \mid \theta \in \Theta^{ud}\} &\equiv \int_{0}^{\infty} \frac{f(\theta_i)\theta d\theta}{1-F(0)} \end{aligned} \tag{13}$$

Now, consider a subsidy policy Υ^s which increases the utility additively by an amount s, which we assume to be strictly positive. The decision criterion is simply

$$S_i = I(\theta_i + s > 0)$$

and the policy splits the support of θ into three distinct sets.

The set of those affected is given by the following;

$$\Theta^{as} = \{\theta \mid -s < \theta < 0\} \tag{14}$$

The set of those unaffected, Θ^{us}, is composed a two subsets. One subset is comprised of those who choose $S = 0$ in presence and in absence of the subsidy. This set, denoted Θ^{us0}, is defined as

$$\Theta^{us0} = \{\theta \mid \ \theta < -c\} \tag{15}$$

The other subset covers the individuals who exercise $S = 1$ both in presence and in absence of the subsidy. This subset, denoted Θ^{us1}, is made of those individuals who are endowed with high values of θ, and is defined as

$$\Theta^{us1} = \{\theta \mid \ \theta > 0\} \tag{16}$$

So, the total set of those unaffected is also disjoint, and it is simply

$$\Theta^{us} = \Theta^{us0} \cup \Theta^{us1} \tag{17}$$

As for the dictatorial experiment, Θ^{us} and Θ^{as} have different supports ($\Theta^{as} \cap \Theta^{us} = \emptyset$). However, the structure of those sets suggests that one could manipulate the subsidy so to equalize the first moments of $F^u_\theta(.;\Upsilon^s)$ and $F^a_\theta(.;\Upsilon^s)$. In practice, this entails finding a value $s' : 0 < s' < \infty$, such that

$$E\{\theta \mid \theta \in \Theta^{as}\} = E\{\theta \mid \theta \in \Theta^{us}\} \tag{18}$$

More precisely, we must find s' such that

$$\begin{aligned} E\{\theta \quad | \quad \theta \in \Theta^{as}\} &\equiv \int_{-s}^{0} \frac{f(\theta_i)\theta}{F(0) - F(-s)} d\theta = \\ E\{\theta \quad | \quad \theta \in \Theta^{us}\} &\equiv \frac{\int_{-\infty}^{-s} f(\theta_i)\theta d\theta + \int_{0}^{\infty} f(\theta_i)\theta d\theta}{F(-s) + (1 - F(0))} \end{aligned} \tag{19}$$

Indeed, under some restriction of the distribution of θ that guarantees existence of $\lim_{s' \to 0} E\{\theta \mid \theta \in \Theta^{as}\}$, it would be possible to prove existence.[28]

Obviously, the binary choice model we just considered is much simpler than the life-cycle skill accumulation model. First, in our model, schooling decisions are not binary. Second, those decisions are subject to the realizations of random shocks, so that the reservation values are stochastic. Thirdly, the heterogeneity terms are represented by a relatively high dimensional vector which may have effects of opposite signs on realized schooling. Finally, as schooling may take many values, the subsidy may be constructed as a full sequence of non-negative numbers.

So, in such a multi-dimensional/multi-period framework, and despite that analytical results may not be obtained, simple intuition suggests that a subsidy policy may be manipulated so to equate (at least) the first moments of

[28]For instance, assuming existence of $\lim_{s' \to 0} E\{\theta \mid \theta \in \Theta^{as}\}$, it would be sufficient to show that $\lim_{s' \to 0} E\{\theta \mid \theta \in \Theta^{as}\} < \lim_{s' \to 0} E\{\theta \mid \theta \in \Theta^{us}\}$ and that $\lim_{s' \to \infty} E\{\theta \mid \theta \in \Theta^{as}\} > \lim_{s' \to \infty} E\{\theta \mid \theta \in \Theta^{us}\}$. Then, using continuity these arguments with respect to s', it would follow that equality must be fullfilled at some s'.

$F^a_{\nu_i}(.)$ and $F^a_{\nu_i}(.)$. Indeed, in a case where individuals decisions are affected by random shocks, it is not even clear that perfect post-intervention realization may not be achieved.

To anticipate on Section 8, it relatively easy to see why the capacity to use a policy intervention that offers a high degree of PIR may be attractive. When a policy interventions creates seemingly identical sub-populations of individuals affected and unaffected, the estimation of a relevant treatment effect is much less likely to be confounded with subsequent unobserved choices.

7.3 Evaluating the Degree of Post-Intervention Randomization

We now return to our original model. The issue is to investigate which class of interventions tend to generate a more "randomized" sub-population of individuals affected. One direct way of verifying it, is to use the fraction of the total variance of individual reactions explained by skills and tastes for each model and intervention combination. Another possibility is to regress individual reactions on ex-ante schooling . A weak marginal effect of ex-ante schooling (a weak correlation) would indicate a low degree of selection based on-ex-ante schooling, which itself, would most likely signal a weak correlation between individual endowments and counterfactual reactions. These marginal effects, along with the R squares of Table 4A and Table 4B, are reported in column 1 of a summary table (Table 4C).

Upon examination of Table 4C, one notes that individual skills and preferences account for 10% of individual reactions to the intermediate education subsidy intervention, and for 6% for the higher education subsidy (within the dynamic skill-complementarity model). As expected, the implementation of a subsidy at basic schooling level should display more selective effects, since it may loose the opportunity to affect the upper tail of the distribution of abilities, but affects more strongly the lower tail. Precisely, individual skills and preferences account for 31% of individual reactions to the low education subsidy intervention.

When the static model is assumed to be the data generating process, the fractions are practically identical. In economic terms, these numbers indicate that randomness plays a large part in determining the counterfactual reactions to policy interventions. In practice, this implies that there

is a non-negligible degree of randomization that takes place after the policy intervention.

The regressions applied to the mandatory schooling interventions disclose a completely different picture. Individual skills and preferences account for 47% of individual reactions to the two-year mandatory schooling regulation, and as much as 60% for the four-years reform, within the dynamic skill-complementarity model. This fundamental difference between dictatorial and incentive-based interventions is in no way linked with a specific model. Virtually identical results are obtained with the static model structure (51% and 70%).

Finally, a quick examination of the marginal effects of education (ex-ante) on individual reactions found in the second column of Table 3C lead to a similar conclusion. Except for the subsidy set over the first 4 years, the marginal effects of the other subsidies are almost 10 times lower in absolute values than those computed for the mandatory schooling regulations. These numbers illustrate again the capacity of subsidy policies to generate a "randomized" set of individuals affected by it.

7.3.1 A Summary of the main differences between Education subsidies and Mandatory Schooling

To summarize, and with respect to all measures that we have chosen to characterize the different policy interventions, there is a fundamental asymmetry between incentive-based and dictatorial policy interventions. This is a result of the inherent difference in the degree of design flexibility between those two families of interventions.

Because incentive-based policies may be designed so to affect a particular set of individuals, they are therefore able to generate a relatively representative (randomized) group of individuals affected by it. Apart from those who belong to the extreme upper tail of the distribution (those who exercise the maximum consumption level ex-ante), incentive based policies may target virtually any group of individuals. Dictatorial policies that enforce minimum consumption, do not have this degree of flexibility. By definition, a minimum consumption regulation affects the bottom of the skill distribution.

In the particular context of education policies, and with reference to our model, this difference is striking. Both intermediate and higher education subsidies affect a set of individuals whose average skill levels are much closer

to population average than the compulsory schooling regulation. The relative flexibility of education subsidies is also exemplified in terms of the statistical strength (proportion of individuals affected) and in terms of the effect of the discount rates. Depending on the timing of the policy, different subsidies that are characterized by the same monetary amount, may affect of different sets of individuals. Finally, while a subsidy policy may affect individuals with higher discount rates as well as lower discount rates, a compulsory schooling affects individual with higher discount rates.

8 IV Estimation and Post-Intervention Randomization

We now turn to the implementation of classical IV estimates. This constitutes the second stage of our analysis.

8.1 The Implication of Post-Intervention Randomization for IV Estimation

In order to interpret IV in a dynamic setting, it is useful to re-express the log wage regression as follows:

$$w_{it} = \lambda_0 + \bar{\lambda} \cdot S_{it} + \varepsilon_{it}^* \tag{20}$$

where

$$\begin{aligned} \varepsilon_{it}^* &= \varphi_{it}(.) + \omega_{\lambda i} \cdot S_i + \varepsilon_{it}^W \\ \varphi_{it}(.) &= \delta_i \cdot E_{it} + \theta_i \cdot A_{it} \\ \omega_{\lambda i} &= \lambda_i - \bar{\lambda} \end{aligned}$$

The term $\bar{\lambda}$ is the population average return to schooling, , λ_0 is an intercept term, $\bar{\lambda}$ is the population average and ε_{it}^* is the composite error term which contains three distinct elements.

The crucial term is $\varphi(.)$. It collapses the effects of all post-schooling choices made until date t.[29] At any period t, it is defined as the sum of each in-

[29]The model allows individuals to leave school for work, and to return to school subsequently.

dividual post-schooling choices (the $A'_{it}s$ and $E'_{it}s$), which are pre-multiplied by the proper individual return to employment and training. More generally, $\varphi(.)$ depends on the entire heterogeneity distribution (not only the returns), and on accumulated schooling (in the case where schooling affects the cost of training). It also depends indirectly on the policy shock indicator (Z).

From now on, we remove the model subscript (m) so to simplify notation. In the static case, we have that

$$\varphi_{it}(\Upsilon) = \Phi^{\Upsilon}(\nu_i) \tag{21}$$

while in the dynamic model, we have

$$\varphi_{it}(\Upsilon) = \Phi^{\Upsilon}(S_{it}(\nu_i), \nu_i) \tag{22}$$

where $\nu_i = \lambda_i, \delta_i, \theta_i, \beta_i, \alpha_i^S, \alpha_i^H, c_{0i}^a$, and where $\Phi^{\Upsilon}(.)$ is a function that maps endowments and schooling (in the dynamic case), onto post-schooling skill accumulation.

The term $\varphi_{it}(.)$ is what renders IV estimation precarious. In a setting where the error term is now re-interpreted as a collection of post-schooling choices, the difference between treatment and control not only consists of schooling choices made by those affected by the policy, but also of post-schooling choices that are themselves correlated with schooling, as well as fixed endowments. So, given that some positive fraction of the individuals are affected by a policy (and increase their schooling level), identifying the treatment effect of education requires absence of differences in post-schooling behavior between those affected and those who are not.

Although PIR is not needed in order to estimate the treatment effect (LATE) of schooling, we expect it to be highly correlated with the fulfillment the identifying orthogonality conditions. We now examine the implications of the nature of the intervention for the potential realization of the identifying orthogonality conditions.

Let's re-consider the orthogonality conditions, and without loss of generality, re-express post-schooling choices as

$$\begin{aligned} \varphi_{it}(\Upsilon) &= \varphi_{it}(. \mid r_i(\Upsilon) = 0) \cdot I(r_i(\Upsilon) = 0) + \\ \varphi_{it}(. &\mid r_i(\Upsilon) > 0) \cdot I(r_i(\Upsilon) > 0)\} \end{aligned} \tag{23}$$

where $I(.)$ is the indicator function equal to1 when individual i belongs to the set identified in parenthesis (either $r_i(\Upsilon) = 0$, or $r_i(\Upsilon) > 0$). Before doing so, we define a binary indicator, x_i^{Υ}, that splits the population into two groups (those affected, and those who are not) as follows:

$$x_i^{\Upsilon} = I(r_i(\Upsilon) > 0) \tag{24}$$

In the static case, the realization of the IV identifying moment condition may be interpreted as follows:

$$Corr\{\varphi_{it}(\Upsilon), x_i^{\Upsilon}\} = Corr\{\varphi_{it}(\Upsilon), x_i^{\Upsilon}\} = 0 \tag{25}$$

or, as follows:

$$Corr\{\varphi_{it}(\Upsilon), x_i^{\Upsilon}\} = Corr\{\varphi_{it}(S_{it}(x_i^{\Upsilon}), x_i^{\Upsilon}), x_i^{\Upsilon}\} = 0 \tag{26}$$

in the dynamic case.

8.1.1 The Static Model

This is the simplest case. When past choices (schooling) have no impact on $\phi_{it}(.)$, after conditioning on heterogeneity, PIR will contribute to diminish the correlation between $\varphi_{it}(.)$ and the treatment/control indicator, simply because $\varphi_{it}(. \mid r_i(\Upsilon) = 0)$ will approach $\varphi_{it}(. \mid r_i(\Upsilon) > 0)$. So, in other words, as PIR becomes more important, $Corr\{\varphi_{it}(x_i^{\Upsilon}), x_i^{\Upsilon}\}$ vanishes. The intervention that offers more PIR is also more likely to result in the realization of the orthogonality condition.

8.1.2 The Dynamic Skill-Complementarity Model

To analyze the dynamic representation, we must first assume that the sign of correlation between S_{it} and x_i^{Υ} is the same as the correlation between φ_{it} and x_i^{Υ}(after conditioning on S_{it}). In practice, this implies that when a policy intervention splits the population into two groups (those affected and those who are not), the group that has the highest level of post-schooling accumulated skills (after removing the effect of schooling if the model is

dynamic) must also have higher schooling attainments. This assumption is basically in line with the popular positive "Ability Bias" hypothesis.[30]

When schooling increases the intensity of post-schooling accumulation, differences in endowments between those affected and those who are not, are reinforced by the effect of schooling on post-schooling accumulation. In such a case, there would be a clear benefit to using a policy intervention that offers PIR, and we should also expect incentive-based experiments to outperform dictatorial experiments. Indeed, the superiority of incentive-based interventions should be even more obvious in such a case.

8.1.3 Skill-Susbtitutability

While the static and the skill-complementarity specifications of the outcome equation represent a relatively wide class of models, it is important to note that if schooling decreases φ_{it} (and we retain the assumption that the signs of the effects of x_i^{Υ} on schooling and post-schooling accumulation are identical), nothing can be said about the desirability of PIR. In such a case, the relative performance of incentive-based and dictatorial experiments depends on the relative importance of the dynamic effect and selectivity. In a returns to schooling framework, it implies that schooling reduces post-schooling wage growth. This would be highly counterintuitive. For this reason, our presentation focuses almost entirely on the static and the skill complementarity models.

9 Implementing IV Estimation

We now proceed with the core of the second step of our analysis, namely the estimation of the returns to schooling using both the dictatorial and the incentive policy interventions as instruments.

[30] In the classical literature, the positive ability bias hypothesis (or OLS bias) states that, within an linear regression of wages on schooling, those who have more schooling, also tend to have a higher value of the error term (interpreted as ability). We therefore rule out the rather unrealistic case where a policy intervention would affect more (less) able individuals, and who in turn, would have less (more) schooling than those unaffected.

9.1 Creating a Cross-Sectional Data Set

In order to build the treatment groups, we proceed as we did for the control groups. Again, each type of individuals is duplicated in 250 different realizations of the random vector (for a total of 50,000 units). We simulate 33 years of choices and wage outcomes under the new policy, as described in the previous section. To mimic what is achieved in observational cross-section data, we first select one time period per-individual. The period is selected randomly between period 5 and period 33. We do so because most actual cross-section data sets contain wages that are realized over the entire life cycle, but do not include very young workers.[31] To do so, we use a uniform random number generator. We end up with 100,000 observations (50,000 in control and 50,000 in treatment).

In the second step, we use simulated choices, and construct a sub-population of individuals who are either in state e or state a at the actual period randomly selected. If this condition is satisfied for an individual, we record the wage as non-missing and include it in the cross-section used to implement IV.

9.2 The Dynamic Skill-Complementarity Model

For exposition purposes, we start by describing IV results applied to the Dynamic-Skill-Complementarity model. We do so for two main reasons. First, it is the model structure that should normally display the highest sensitivity to the degree of PIR. As a consequence, the distinction between IV estimates coming from dictatorial and incentive interventions should come out clearly.

Second, on a purely economic basis, the dynamic skill-complementarity model as the most interesting specification. Indeed, it is also likely to be the most realistic.

9.2.1 The Education Subsidies

The IV estimates obtained when the data generating process is the dynamic skill-complementarity model are reported in the upper portion of Table 5.

[31] In our model, period 5 would be naturally compared to the period at which individuals decide to enter higher education (say, around 18 years of age).

Evidently, the IV estimate associated to the intermediate level subsidy, which is equal to 0.0715, lies within a very narrow distance of the population LATE (0.0717). It is also relatively precise with a standard error equal to 0.01. There is an obvious way to explain the good performance of IV. The low correlation between Z and post-schooling choices, equal to 0.0019, practically imply orthogonality between policy exposure and post-schooling skill accumulation.

As normally expected, the higher education subsidy generates a weaker instrument as indicated by the relative imprecision of the IV estimate (the standard error is equal to 0.03), or by the correlation between schooling and policy exposure (equal to 0.0183). Although the IV point estimate, equal to 0.0846, is targeting a population LATE parameter equal to 0.0809, it would be imprudent to conclude that it is performing well. We do not analyze it further.

In an earlier section, we already pointed out the difference between the lower education subsidy and the subsidies that are set at the intermediate level. Indeed, the basic education subsidy may be compared to 4-year mandatory schooling policy (because it is set over the same period), which will be analyzed below. At this stage, it is sufficient to note that the IV estimate, equal to -0.0894 with a standard error equal to 0.02, fails spectacularly to estimate its estimand (0.0369). It lies outside the support of the distribution of returns. It is certainly less precise than the intermediate level subsidy but more precise than the higher education subsidy. As argued earlier, this may be simply explained by discounting.

As indicated earlier, a different way to visualize the realization of orthogonality conditions, is to investigate post-schooling accumulation within the treatment group. To provide a better explanation of the results, we took all observations of the treatment group, and regressed post-schooling accumulation ($\varphi_{it}(.)$ on a set of regressors that include (i) the variable that records if an individual has been affected (x_i^{Υ}), (ii) post-intervention schooling and its square, as well as (iii) age and its square. We focus on the estimate associated to the variable that records if an individual has been affected, since it tells us how important are differences in accumulation, after removing the effect of schooling, and more particularly, how this marginal effect may vary across policy interventions. The results of the regressions for the dynamic skill-complementarity model are in Table 6A.

The impact of PIR is obvious upon examination of the first three columns

of Table 6A, which are devoted to the education subsidies. One has to examine the parameter estimates obtained for the variable that records if an individual has been affected (x_i^{Υ}). These estimates (equal to -0.0005 for the intermediate level, and -0.0001 for higher education) are miniscule, and insignificant. They reveal simply the absence of differences in post-schooling behavior between those affected and those who are not. This is why, in this particular context, IV is capable of identifying the true causal effect of education on wages. However, the corresponding parameter obtained for the lower education subsidies, -0.0142, illustrates the opposite outcome. This estimate, which is in relative term quite enormous, simply illustrates that the lower education subsidy appears to affect a highly selective sub-population. The post-schooling accumulation behavior of this sub-population prevents the econometrician from inferring the proper return to schooling.

9.2.2 Mandatory Schooling

We now analyze the IV obtained from the mandatory schooling interventions set at a minimum of 2 years and 4 years. Not surprisingly, the 4-year mandatory schooling intervention generates a stronger instrument than the 2-year compulsory schooling regulation. However, and regardless of the statistical strength of the intervention, the instruments generated by mandatory schooling reforms are incapable of estimating the proper local average treatment effects.[32] As argued earlier, the accuracy of IV will depend on the capacity of the policy intervention to generate enough PIR. However, there is no guarantee that a stronger instrument generates a more dispersed sub-population. Other things equal, we already have shown that the mandatory schooling experiment implies a highly selective sub-population of individuals affected.

The consequences of this lack of randomization is now clear. The IV estimate, equal to -0.1198 (with a standard error equal to 0.02) lies outside the support of the distribution of returns, and moreover, is nowhere near its population counterpart (0.03). The negativity of the estimates is also easy to explain. Our model is calibrated according to a skill complementarity hypothesis. That is it assumes that skills beget skills, and as a consequence, a higher level of skills conveys higher post-schooling wage growth.

[32] As the reader may have noticed, in a context where individuals continue to exercise choices beyond policy intervention, and beyond school completion, access to a strong instrument is no longer automatically desirable.

The mandatory schooling experiment, which affects mostly a low skill population, large negative correlation between the control/treatment indicator and post schooling wage growth (-0.0295). This is 10 to 20 times higher than the corresponding correlations obtained in the subsidy interventions. These differences coexist even though the correlations between schooling and the policy shock indicator are relatively similar across policy interventions.[33] As a result, the IV estimate lies within a region that is far away from the population average (while the population LATE parameters are within some reasonable distance from the population average). Although increasing the minimum schooling requirement has slightly improved the performance of the IV estimate, which is now equal to -0.0846 for a population parameter equal to 0.0303, there is still evidence that IV is incapable of estimating the relevant treatment effect. This is well illustrated by the higher correlation (in absolute value) between the control/treatment indicator and post schooling wage growth (-0.0485).

As we did before, we now investigate post-schooling accumulation behavior within the treatment group, and compare the results with what was obtained with the education subsidies. The parameters associated to x_i^{Υ}, and obtained for the 2-year and 4-year compulsory schooling, are found in the last 2 columns of Table 6A. They are respectively equal to -0.0108 and -0.0162, and are 20 to 100 times larger than those found for the intermediate and higher education subsidies. They are, however, comparable, with the estimates obtained for the basic schooling subsidy. They indicate clearly that the set of individuals affected by the mandatory schooling reforms is highly selective, and displays post-schooling behavior significantly different from the set of those unaffected. It explains why IV confounds wage growth induced by schooling. with post-schooling skill accumulation.

[33]Obviously, our analysis has also implications for the economics of the weak/strong instrument paradigm. Within a dynamic skill accumulation model affected by multivariate heterogeneity, the correlation between the instrument and the structural elements of the composite error term also change when the correlation between schooling and the instrument varies. As a result, increasing (reducing) the power of an instrument does not guarantee a reduction (increase) in the asymptotic bias. However, we focus our discussion on the distinction between incentive and dictatorial interventions.

9.3 The Static Model

When the outcome equation is affected by subsequent choices, that are solely driven by persistent skills and tastes, we already established that the intervention generating more post intervention randomization would also perform the best, although the difference between compulsory and incentive-based interventions is likely to be less pronounced than what was observed for the dynamic skill-complementarity model. To implement it, we reconstruct four different treatment groups using a version of the model where the parameters $\{c_{1a}, c_{2a}, c_{3a}, c_{4a}\}$ and $\{c_{1e}, c_{2e}, c_{3e}, c_{4e}\}$ are set to 0.

9.3.1 The Education Subsidies

The estimates are found in the lower portion of Table 5. Again, the IV estimate associated to the intermediate education subsidy (0.0630) is quite accurate and quite precise (the standard error is 0.0069). It targets a LATE parameter equal to 0.0645.

As observed before, the higher education subsidy generates a weak correlation between schooling and the policy reform indicator (equal to 0.0095), and this translates into a high degree of imprecision. The IV estimate is equal to 0.0937, but he standard error, which is equal to 0.0829, is almost as high as the IV estimate itself. Again, the high accuracy of the point estimate is hardly interpretable giving the high degree of imprecision.

Again, we note the lower education subsidy generates a IV estimator which is more precise than the higher education subsidy. However, it does not recover the population treatment effect. The IV estimate, equal to 0.0183, misses the population parameter by 3 percentage points (0.0478). Its standard error, is equal to 0.0103.

9.3.2 Mandatory Schooling

The benefit of using an instrument that generates a higher level of PIR comes quite clearly again, upon examining the mandatory schooling IV estimates. Those estimates, equal to -0.0044 and -0.0466, are very precise and are both significantly outside any reasonable confidence interval, as they are targeting LATE parameters equal to 0.0340 and 0.0379. As for the dynamic skill complementarity model, we can compare the correlation between the

control/treatment indicator and post schooling wage growth (3rd column of Table 5), or compare the parameter estimates measuring the effect of x_i^{Υ} in the treatment group (Table 6B) .

In either case, the differences between the incentive and the dictatorial interventions come out clearly. The mandatory schooling instrument generate a stronger (negative) correlation between Z and $\varphi_{it}(.)$. The correlations, equal to -0.0231 and -0.0715, are more than a hundred times superior to the correlations obtained for the subsidies. Equally, the effects of x_i^{Υ} in the treatment groups measured for the compulsory schooling interventions, equal to -0.0051 and -0.0207, are much stronger in relative terms than those obtained from the intermediate and the higher education subsidies, which are equal to 0.00002 and -0.0002.[34]

9.4 Assuming Substitutability

As indicated earlier, in presence of substitutability differences in skills between those affected and those unaffected, may be counterbalanced by the effects of differences in schooling . As a consequence, it is impossible to say if PIR is beneficial. All simulations and estimations performed in the current version of the paper have also been carried under the skill substitutability assumption. To do so, we just replaced the parameters generating skill complementarity by their opposite. However, to save space and to simplify the presentation, we do not report them. Basically, the results indicated no marked difference in performance between incentive-based IV and compulsory schooling IV.

9.5 A Summary of the IV Results

To summarize, our results illustrate the benefit of accessing an instrument that generates a higher level of PIR, in a context where the error term of the outcome equation is generated by wide classes of post-schooling discrete choice models.

[34]For the sake of completeness, we have also investigated this possibility to control for endogeneity of post-schooling experience, which may be defined as the sum of employment and training as work experience, using age and its square as instruments. Because we have obtained results practically identical to those where only schooling is instrumented, and in order to minimize the number of redundant tables, we do not report those estimates.

Because of its relative degree of flexibility, incentive-based policies (such as education subsidies) may be capable of generating a sufficient degree of randomization. This is not true for any incentive policy, but when it is, this translates into a much higher capacity of incentive-based policies to generate an instrument that fulfills the identifying orthogonality condition. This is what enables the econometrician to estimate a particular treatment effect.

On the other hand, dictatorial policy interventions (such as compulsory schooling regulations) that set a minimum consumption/investment level cannot generate a reasonable degree of PIR, unless the minimum is set at an unrealistic level, or unless the behavioral model displays no selection. For this reason, a compulsory schooling regulation cannot realistically generate an instrument that fulfills the identifying orthogonality condition.

10 Some Interpretation

Obviously, we have chosen to interpret differences in IV from the perspective of an econometrician who seeks to estimate a treatment effect. Our results may also be interpreted from a policy perspective. It is the negative estimates, associated to the compulsory schooling reform, that are particular interesting to analyze. When policy designers implement mandatory schooling, individual decisions are annihilated for a subset of the population. So, those individuals who are endowed with low returns to academic skills (those for whom the return to schooling is close to 0), are forced to delay their post-schooling skill accumulation. At any date t, those individuals are facing employment opportunities that take into account both their higher schooling level, and their lower work experience (the sum of A_{it} and E_{it}). Consequently, for those endowed with relatively high returns to work experience (in particular E_{it}), the policy may entail a reduction in lifetime earnings. This is possible, even though the returns to skill accumulation are strictly positive for every individual.

As such, these negative estimates are explained by the negative correlation between wages, and policy intervention (what is usually called indirect least-squares). For instance, suppose one uses total experience as the endogenous variable (which may be regarded just as the mirror image of the schooling variable in a simple binary choice framework), and applies standard IV. Then, the IV estimate becomes positive, since a change is compulsory schooling

reduces employment for a subset of the population. In our model, there are as many IV estimates, as there are endogenous states. Obviously, some of these may be positive, while others may be negative. This multiplicity of signs (let alone the nominal values of the different IV estimates) prevails even if there is a single policy reform. Indeed, at a purely intuitive level, the occurrence of a negative IV estimate, should normally raise suspicions about the validity of the usual economic interpretation of IV, as some treatment effect parameter.

11 Implications for Empirical Work and Concluding Remarks

At this stage, two questions naturally arise. First, what are the implications of our results for the literature on the returns to schooling, or for empirical work in general? Second, is the distinction between dictatorial and interventions relevant in other contexts?

To answer the first question, it is informative to refer to the IV literature on the returns to schooling. In this literature, it is common to rely on static interpretations of the wage schooling relationship, and potential mis-specifications are practically always ignored. In the earlier literature surveyed in Card (1999), the conventional wisdom is that IV estimates of the returns to education are high. Interestingly, the nature of the policy reform is never discussed, but the majority of IV estimates reported appear to be obtained from various forms of mandatory schooling regulations. Indeed, a thorough review of the IV estimates obtained from various compulsory schooling reforms (including those published more recently) would reveal an important dispersion in reported estimates. For instance, different empirical analyses of a same compulsory schooling reform, with the same data, may lead to differences of the order of 10 percentage points.[35] It is important to understand that our claim is not that IV estimates of mandatory schooling reform need to be as low as they are in our example. The point estimates reported in the paper have no direct interest per-se. Our objective is solely to

[35] As an example, Deveraux and Hart (2008) report some IV estimates of a British compulsory school reform that are close to 0, and mention other studies that report estimates above 0.10 for the same policy intervention.

investigate the relative performance of IV strategies that use different policy interventions. For instance, different assumptions about the dynamic structure (a skill substitutability assumption, along with a different correlation structure) may well lead to estimates above the population LATE. Nevertheless, and as discussed in Cameron and Taber (2004), low estimates of the returns to schooling tended not to get reported for a long time, but are gradually being presented in the empirical IV literature. So, as IV estimates that are close to 0 are now attracting more attention, it is not even clear that negative IV estimates, such as those that are presented in the paper, are unrealistic.

Our conjecture is that applied econometricians would be reluctant to report IV specifications that result in negative estimates of the wage return to schooling. Notwithstanding this, our conclusion is that IV estimates of the returns to schooling reported in the literature, and which have been obtained from mandatory schooling reforms, are probably those that should be more subject to scrutiny. The reason is simple. Dictatorial policy interventions are ill-equipped to break the curse of selectivity in the sub-population affected. Dictatorial policies always directly (or indirectly) target a more selective sub-population.

With respect to the second question, although we cannot bring a formal analysis, our 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. 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.[36]

We can also find illustrations outside the educational context. In any

[36]See Imbens and van der Klaauw (1995), for an analysis of Dutch conscription.

model where completed fertility is a key determinant of some individual choice (labor supply) or some outcome (earnings), the distinction between dictatorial and incentive-based interventions may be relevant. As is the case for schooling, fertility has been the subject of various policies aimed at stimulating (or reducing) it. While many countries have adopted incentive provisions to raise family size (Sweden, France, Canada, to mention a few), China has adopted a more restrictive policy[37]. Empirical economists may be tempted to use some of these policy reforms in the context of IV estimation. However, our conjecture is that the issues we raised in this paper would also be relevant in the fertility case, as well as many other cases.

References

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

[2] 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

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

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

[5] 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

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

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

[37] In China, the policy intervention dictated a maximum number of children per-family.

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

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

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

[11] 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

[12] 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

[13] Paul Devereux and Robert A. Hart (2008) "Forced to Be Rich? Returns to Compulsory Schooling in Britain" IZA Discussion Paper 3305

[14] 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.

[15] Heckman, James (1997) "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources, 32 (3), 441-62.*

[16] Heckman, James and Vytlacil, Edward (2005) "Structural Equations, Treatment Effects and Econometric Policy Evaluations" *Econometrica. 73.*

[17] Heckman, James, Sergio Urzua and Vytlacil, Edward (2007) "Understanding Instrumental Variables in Models with Essential Heterogeneity", *Review of Economics and Statistics*

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

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

[20] 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.

[21] Keane, Michael (2006) "Structural vs. Atheoretic Approaches to Econometrics, *Journal of Econometrics*

[22] Keane, Michael and Wolpin, Kenneth (1997) "The Career Decisions of Young Men" *Journal of Political Economy, 105, 473-522.*

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

[24] 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

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

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

Table 1A
Life Cycle Choices in the Control Group:
Dynamic Skill Complementarity Model

Accumulated number of periods in each State

year	In School	Employment (learning by doing)	Employment (Training)	Home Production
1	0.789	0.146	0.001	0.063
2	1.526	0.289	0.021	0.164
3	2.224	0.449	0.057	0.270
4	2.872	0.633	0.118	0.377
5	3.444	0.849	0.218	0.489
6	3.946	1.096	0.354	0.605
7	4.384	1.358	0.533	0.724
8	4.758	1.632	0.767	0.843
9	5.044	1.911	1.083	0.962
10	5.260	2.192	1.466	1.082
15	5.587	3.750	3.982	1.682
20	5.613	6.739	5.371	2.276
25	5.614	11.125	5.398	2.864
30	5.614	15.530	5.398	3.458
33	5.614	18.173	5.398	3.815

Table 1B
Life Cycle Choices in the Control Group:
Static Model

Accumulated number of periods in each State

year	In School	Employment (learning by doing)	Employment (Training)	Home Production
1	0.586	0.271	0.006	0.136
2	1.123	0.526	0.046	0.305
3	1.622	0.791	0.105	0.482
4	2.080	1.067	0.192	0.661
5	2.464	1.347	0.330	0.860
6	2.788	1.650	0.491	1.072
7	3.063	1.959	0.682	1.296
8	3.298	2.279	0.901	1.521
9	3.488	2.607	1.159	1.747
10	3.643	2.940	1.443	1.974
15	3.950	4.772	1.761	3.113
20	3.993	7.547	4.192	4.268
25	3.995	11.334	4.218	5.453
30	3.995	15.101	4.218	6.686
33	3.995	17.338	4.218	7.448

Table 1C
The Distribution of Schooling in the Control Groups

	Model	
	Dynamic/Skill-comp.	**Static**
Years of Schooling	**Percentage**	**Percentage**
0	15.19	32.71
1	4.39	7.52
2	3.42	3.18
3	3.91	2.29
4	16.44	18.25
5	5.74	4.16
6	5.52	3.75
7	6.63	4.21
8	21.1	11.56
9	5.9	3.26
10-more	11.7	9.11

Table 2A
The determinants of Schooling, Employment and Training in the Control Group:
Dynamic Skill Complementarity Model

Dependent variables	**(1) Schooling (S)**	**(2) Employment (E)**	**(3) Training (A)**
	Estimate (st. error)	**Estimate (st. error)**	**Estimate (st. error)**
Heterogeneity Components			
intercept	6.3637	-0.5075	-7.1614
	(0.0063)	(0.0807)	(0.0590)
academic ability	0.4178	0.2873	0.7126
	(0.0122)	(0.0236)	(0.0159)
return to training/educ.	2.3503	-1.0235	0.0036
	(0.0122)	(0.0337)	(0.0226)
return to Emp. (learning)	0.1759	2.5546	-0.8639
	(0.0064)	(0.0130)	(0.0087)
home time	0.0325	0.2443	-0.4294
	(0.0063)	(0.0213)	(0.0087)
Discount factor	1.4943	-0.0747	–0.3372
	(0.0063)	(0.0213)	(0.0143)
Schooling	-	-0.1358	0.8952
		(0.0106)	(0.0071)
age		0.4736	0.7387
		(0.0078)	(0.0052)
age^2		0.0058	-0.0167
		(0.0002)	(0.0001)
R square	0.80	0.84	0.83
# of observations	50,000	50,000	50,000

Note: The regression is performed on 50000 individuals observations on completed schooling. All individual endowments are standardized.

Table 2B
The determinants of Schooling, Employment and Training in the Control Group: Static Model

Dependent variables	**(1) Schooling (S)**	**(2) Employment (E)**	**(3) Training (A)**
	Estimate (st. error)	**Estimate (st. error)**	**Estimate (st. error)**
Heterogeneity Components			
intercept	4.7947	-0.0729	-5.1195
	(0.0059)	(0.0632)	(0.0384)
academic ability	0.2571	1.1823	0.3491
	(0.0117)	(0.0242)	(0.0147)
return to training/educ.	3.4312	-0.2964	1.4642
	(0.0116)	(0.0412)	(0.0250)
return to learning	0.0059	3.2320	-0.3598
	(0.0061)	(0.0126)	(0.0076)
home time	-0.0128	0.0581	-0.4246
	(0.0060)	(0.0125)	(0.0076)
Discount factor	0.9144	0.6060	-0.4515
	(0.0061)	(0.0154)	(0.0093)
Schooling	-		-
age	-	0.6021	0.6663
		(0.0064)	(0.0039)
age^2	-	0.0006	-0.0145
		(0.0002)	(0.0001)
R square	0.86	0.85	0.81
# of observations	50,000	50,000	50,000

Note: The regression is performed on 50000 individuals observations on completed schooling. All individual endowments are standardized.

Table 3
Characteristics of the Sub-population Affected by all Policy Interventions

Model/Intervention	% affected	Returns		Education
		educ.	learning	(ex-ante)
Dynamic Skill-Comp				
lower educ. subsidy	26.2	0.0369	0.0075	3.79
intermed. educ. subsidy	24.4	0.0717	0.0086	7.93
Higher educ. subsidy	9.1	0.0809	0.0127	8.38
2-year mandatory	25.8	0.0303	0.0045	1.11
4-year mandatory	34.6	0.0376	0.0061	2.89
Static				
lower educ. subsidy	23.0	0.0478	0.0189	2.74
intermed. educ. subsidy	21.4	0.0787	0.0108	5.78
Higher educ. subsidy	5.8	0.0925	0.0112	8.04
2-year mandatory	45.9	0.0340	0.0170	1.65
4-year mandatory	52.2	0.0379	0.0189	2.07

Note: The Education variable measures ex-ante schooling (pre-intervention schooling)

Table 4A
The Effects of Skills and Preferences on Individual Counterfactual Reactions
Data Generating Process: Dynamic Model

	Educ. Subsidy			Mandatory schooling	
	Lower	Intermed.	higher	2 years	4 years
Dependent Variable	$R_i(.)$	$R_i(.)$	$R_i(.)$	$R_i(.)$	$R_i(.)$
	(1)	(2)	(3)	(4)	(5)
Skills/Preferences					
intercept	0.4985	0.3473	0.1075	0.5346	0.9373
	(0.0094)	(0.0011)	(0.0016)	(0.0023)	(0.0042)
academic ability	-0.2494	0.2156	-0.0480	-0.2557	-0.4192
	(0.0054)	(0.0020)	(0.0030)	(0.0045)	(0.0082)
return to educ.	-0.4016	-0.0515	0.1125	-0.3280	-0.5255
	(0 .0094)	(0.0020)	(0.0030)	(0.0045)	(0.0081)
return to Emp.	-0.1794	-0.0859	0.0130	-0.2706	-0.3664
	(0 .0051)	(0.0011)	(0.0016)	(0.0024)	(0.0043)
home time	-0.0423	-0.0891	-0.0082	-0.0147	-0.0037
	(0.0051)	(0.0011)	(0.0016)	(0.0023)	(0.0042)
Discount factor	-0.3313	-0.0539	0.0432	-0.2911	-0.6072
	(0 .0052)	(0.0011)	(0.0016)	(0.0023)	(0.0043)
R square	0.3201	0.0715	0.0612	0.4725	0.6038

Note: The dependent variable ($R_i(.)$) is measured as the difference between the individual optimal schooling under a specific policy intervention and the individual optimal schooling in absence of any intervention (control group). The regressions are computed using 50,000 counterfactual reactions. All individual endowments are standardized.

Table 4B
The Effects of Skills and Preferences on Individual Counterfactual Reactions
Data Generating Process: Static Model

	Educ. Subsidy			Mandatory Schooling	
	Lower	Intermed.	higher	2 years	4 years
Dependent Variable	$R_i(.)$	$R_i(.)$	$R_i(.)$	$R_i(.)$	$R_i(.)$
	(1)	(2)	(3)	(4)	(5)
Endowments					
intercept	0.4716	0.3843	0.0600	0.8980	1.6363
	(0.0065)	(0.0269)	(0.0012)	(0.0034)	(0.0047)
academic ability	-0.0046	0.2688	-0.0515	-0.2085	-0.4643
	(0.0115)	(0.0072)	(0.0023)	(0.0066)	(0.0091)
return to educ.	-0.3216	-0.0457	0.1052	-0.5617	-1.1171
	(0.0114)	(0.0071)	(0.0033)	(0.0065)	(0.0091)
return to Emp.	0.3205	0.0074	-0.0016	-0.0317	0.0098
	(0.0066)	(0.0038)	(0.0012)	(0.0034)	(0.0047)
home time	0.1243	-0.1848	0.0105	0.1036	0.2071
	(0.0066)	(0.0037)	(0.0012)	(0.0034)	(0.0047)
Discount factor	-0.0831	-0.0031	0.0092	-0.1875	-0.4652
	(0.0122)	(0.0037)	(0.0012)	(0.0033)	(0.0047)
R square	0.20	0.11	0.06	0.51	0.70

Table 4C
Summary Measures of the degree of Post-Intervention Randomization
by Intervention and Model

		Regression	
		Individual reactions on skills (R^2)	Individual reactions on ex-ante education (marg. effect of educ)
Interventions	**Model**		
Subsidy: Basic			
	Dynamic	0.32	-0.0768*
	Static	0.20	-0.0529*
Subsidy: Intermediate			
	Dynamic	0.07	0.0183*
	Static	0.11	0.0283*
Subsidy: Higher			
	Dynamic	0.06	0.0223*
	Static	0.06	0.0194*
Two-year mandatory			
	Dynamic	0.47	-0.0871*
	Static	0.51	-0.1072*
Four-year mandatory			
	Dynamic	0.6038	-0.1019*
	Static		-0.1096*

Note: All marginal effects with a "$*$" are significant at the 0.1% level.

Table 5
IV Estimates of the Returns to Schooling

Model/Intervention	$\hat{\lambda}_{iV}$ ($st.error$)	**LATE** Educ	**Correlations** ($Z_i,\ \varphi_{it}(.)$	($Z_i,\ S$)
Dynamic				
lower educ subsidy	-0.0894 (0.0188)	0.0369	-0.0338	0.0512
intermed. educ. subsidy	0.0715 (0.0099)	0.0717	0.0019	0.0641
Higher educ. subsidy	0.0846 (0.0333)	0.0809	-0.0008	0.0183
2-year mandatory	-0.1198 (0.0242)	0.0303	-0.0295	0.0375
4-year mandatory	-0.0846 (0.0094)	0.0376	-0.0485	0.0906
Static				
lower educ. subsidy	0.0183 (0.0103)	0.0478	-0.0184	0.0645
intermed. educ. subsidy	0.0630 (0.0069)	0.0645	-0.0048	0.0645
Higher educ. subsidy	0.0937 (0.0829)	0.0925	-0.0004	0.0095
2-year mandatory	0.0044 (0.0074)	0.0340	-0.0231	0.0996
4-year mandatory	-0.0466 (0.0046)	0.0379	-0.0715	0.1854

Table 6A
Analysis of post-schooling Skill Accumulation of the Treatment Group
Data Generating Process: Dynamic Model

	Educ Subsidy			Mandatory Schooling	
	Lower	Intermed.	higher	2 years	4 years
intercept	-0.0565	-0.0634	-0.0325	-0.0264	-0.0418
	(0.0050)	(0.0042)	(0.0055)	(0.0043)	(0.0048)
x_i^{Υ}	-0.0142	-0.0005	-0.0001	-0.0108	-0.0162
	(0.0019)	(0.001)	(0.0014)	(0.0014)	(0.0014)
Schooling	0.0054	0.0194	0.0231	0.0006	-0.0079
	(0.0011)	(0.0009)	(0.0009)	(0.0012)	(0.0012)
Schooling^2	0.0005	-0.0008	-0.0011	0.0008	0.0014
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
age	0.0066	0.0073	0.0070	0.0063	0.0057
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
age^2	-0.00007	-0.0001	-0.0001	-0.0001	-0.0001
	(0.00001)	(0.0000)	(0.00000)	(0.0000)	(0.0000)
R^2	0.05	0.05	0.05	0.05	0.05

Note: The variable x_i^{Υ} is equal to 1 when individual i is affected by the policy ($r_i(\Upsilon) > 0$) and 0 if not ($r_i(\Upsilon) = 0$)

Table 6B
Analysis of post-schooling Skill Accumulation of the Treatment Group
Data Generating Process: Static Model

	Educ Subsidy			Mandatory Schooling	
	Lower	Intermed.	higher	2 years	4 years
intercept	-0.0001	0.0170	0.0194	0.0089	0.0059
	(0.0035)	(0.0036)	(0.0036)	(0.0035)	(0.0033)
x_i^{Υ}	-0.0053	0.00002	-0.0002	-0.0091	-0.0207
	(0.0014)	(0.0014)	(0.0014)	(0.0013)	(0.0013)
Schooling	-	-	-	-	-
Schooling^2	-	-	-	-	-
age	0.0104	0.0074	0.0070	0.0087	0.0099
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
age^2	-0.0001	-0.00005	-0.00004	-0.0001	-0.0001
	(0.00001)	(0.00001)	(0.00001)	(0.00001)	(0.00001)
R^2	0.05	0.06	0.06	0.06	0.05

Note: The variable x_i^{Υ} is equal to 1 when individual i is affected by the policy ($r_i(\Upsilon) > 0$) and 0 if not ($r_i(\Upsilon) = 0$)

Appendix
Table A1
The Heterogeneity Distribution

	α^S	α^H	α^W	λ	θ	δ	c_0^a	$\frac{1}{1+r}$
type								
1	-0.1575	2.59		0.010	0.0000	0.0050	5.4851	0.97
2	0.0425	3.01		0.029	0.0246	0.0145	5.3771	0.99
3	0.6425	2.56		0.081	0.0000	0.0405	5.0531	0.91
4	0.4425	2.66		0.051	0.0066	0.0255	5.1611	0.97
5	-0.2575	2.81		0.027	0.000	0.0135	5.5391	0.95
6	0.3425	2.71		0.039	0.0366	0.0195	5.2150	0.92
7	-0.0575	2.91		0.033	0.0126	0.0165	5.4310	0.90
8	-0.1575	2.85		0.025	0.0056	0.0125	5.4851	0.96
9	0.1425	2.87		0.045	0.0306	0.0225	5.3231	0.93
10	1.1925	2.76		0.057	0.0186	0.0285	4.7561	0.98
11	-0.3575	2.91		0.020	0.0000	0.0100	5.5931	0.91
12	0.3425	3.01		0.043	0.0016	0.0215	5.2150	0.99
13	0.2425	2.61		0.075	0.0196	0.0375	5.2691	0.96
14	0.2425	2.61		0.069	0.0136	0.0345	5.2691	0.99
15	0.9425	2.85		0.087	0.0166	0.0435	4.8910	0.95
16	0.7425	2.87		0.093	0.0076	0.0465	4.9991	0.94
17	1.5425	2.81		0.079	0.0006	0.0395	4.5671	0.90
18	1.7425	2.76		0.118	0.0076	0.0590	4.4591	0.92
19	1.2425	2.63		0.112	0.0136	0.0560	4.7291	0.98
20	1.1425	3.21		0.107	0.0106	0.0535	4.7831	0.95
Mean	0.50	2.80		0.06	0.01	0.03	5.13	0.95
St Dev.	0.62	0.17		0.03	0.01	0.02	0.34	0.03

Note: Each type has a population proportion equal to 0.05.

Table A2
The correlation between the heterogeneity component

	cost of schooling α^S	value home α^H	wage intercept α^W	return schooling β	retutn to experience θ	return to training δ	cost of training c_0^a	Discount Factor $\frac{1}{1+r}$
α^S	1.00	-0.009		0.857	0.111	0.857	-1.00	-0.106
α^H		1.00		-0.043	0.087	-0.043	0.009	-0.046
β				1.00	0.110	1.00	-0.857	-0.076
θ					1.00	0.110	-0.111	0.112
δ						1.00	-0.857	-0.076
c_0^a							1.00	0.106
$\frac{1}{1+r}$								1.00

Table A3
OLS regressions on Simulated Wages in the Control Group: Dynamic Skill Complementarity Model

	Estimate (st. errors)	**Estimate (st. errors)**	**Estimate (st. errors)**	**Estimate (st. errors)**
education	0.1213 (.0007)	0.1333 (0.0008)	0.0964 (0.0018)	0.1299 (0.0008)
experience	-	0.0167 (0.0003)	0.0050 (0.0006)	0.0448 (0012)
educ*experience			0.0020 (0.0001)	
experience^2	-	-	-	-0.0008 (0.0001)
R^2	0.33	0.36	0.37	0.37

Note: The regressions are computed on a cross section of wages generated from simulated individual choices. The cross section contains 50,000 observations (20 types multiplied by 2500 different realizations of the full vector of random shocks).Experience is defined as the sum of the number of years in employment/learning by doing, and employment/training.

Table A4
OLS regressions on Simulated Wages in the Control Group: Static Model

	Estimate (st. errors)	**Estimate (st. errors)**	**Estimate (st. errors)**	**Estimate (st. errors)**
education	0.1232 (0.0007)	0.1355 (0.0007)	0.1045 (0.0018)	0.1343 (0.0007)
experience		0.0155 (0.0003)	0.0078 (0.0005)	0.0420 (0.0012)
educ*experience			0.0017 (0.0001)	-
experience^2				-0.0007 (0.00003)
R^2	0.3698	0.4022	0.4066	0.4079