Dynamic Skill Accumulation, Education Policies and the Return to Schooling*

Christian Belzil
CREST, CNRS, Ecole Polytechnique, Université de Paris-Saclay and IZA

Jorgen Hansen
Concordia University, CIRANO, CIREQ and IZA

Xingfei Liu
University of Alberta and IZA

January 30, 2017

*We would like to thank the co-editor and two anonymous referees for useful comments. We would also like to thank Orazio Attanasio, Lance Lochner, Salvador Navarro, Nirav Mehta, Jean-Marie Dufour, Xavier D'Hautefoeuille, and Arnaud Maurel. Financial support from Investissement d'Avenir (ANR-11-IDEX-0003/ Labex Ecodec/ANR-11-LABX-0047) (Christian Belzil) and Social Sciences and Humanities Research Council of Canada (grant 410-2011-1928) (Jorgen Hansen) is gratefully acknowledged. The usual disclaimer applies.

Abstract

Using a dynamic skill accumulation model of schooling and labor supply with learning-by-doing, we decompose early life-cycle wage growth of U.S. white males into four main sources: education, hours worked, cognitive skills (AFQT scores) and unobserved heterogeneity, and evaluate the effect of compulsory high school graduation and a reduction in the cost of college. About 60 percent of the differences in slopes of early life-cycle wage profiles are explained by heterogeneity while individual differences in hours worked and education explain the remaining part almost equally. We show how our model is a particularly useful tool to comprehend the distinctions between compulsory schooling and a reduction in the cost of higher education. Finally, because policy changes induce simultaneous movements in observed choices and average per-year effects, linear IV estimates generated by those policy changes are uninformative about the returns to education for those affected. This is especially true for compulsory schooling estimates as they exceed IV estimates generated by the reduction in the cost of higher education even if the latter policy affects individuals with much higher returns than than those affected by compulsory schooling.

Key Words: *Dynamic Skill Accumulation, Education Policies, Returns to Schooling, Learning-by-Doing, Life-cycle Labor-Supply, IV estimation*

JEL Classification: I2, J1, J3.

1 Introductory Remarks

We estimate an early life-cycle dynamic skill accumulation model of schooling and labor supply by simulated maximum likelihood techniques using a sample of white males from the 1979 National Longitudinal Survey of Youth. Our model separates the effect of education on entry wages from its effects on wage growth while allowing for endogenous labor supply decisions both at the extensive and intensive margins. The model has a single skill but incorporates a particularly rich heterogeneity specification as both the effects of education on entry wages and the returns to work experience (hours worked) depend on observed (AFQT scores) and unobserved heterogeneity.

Our model provides an ideal framework to quantify four potential explanations for the existence of steeper age-earnings profiles for the more educated (Heckman, Lochner and Todd, 2006). First, it allows for the more educated to have higher observed and unobserved ability endowments generating higher returns to work experience after conditioning on hours worked. Second, it allows the more educated to have a higher utility of working long hours. Third, after conditioning on ability endowments and tastes, education may raise the productivity of work experience. Finally, and again after conditioning on ability and taste, education affects the utility (or disutility) of working a high level of hours. The first two explanations constitute pure selection effects. The last two generate causal effects of education on earnings dynamics.

We use the model to investigate the economic impact of two policy interventions: a compulsory schooling policy and a reduction in the cost of attending higher education (college). Our compulsory schooling policy consists of a mandatory high school graduation regulation while the reduction in the cost of college that we implement is approximately equivalent to $5,000 per year (in 1997 dollars) and corresponds nearly to a 80 percent reduction in the total direct cost (net of institutional transfers) of attending a 4-year college over the early 1980's (Abel and Deitz, 2014).

Using the heterogeneity distribution of those affected (and unaffected) by each specific policy, we illustrate how the dual impact of education on returns to post-schooling skill investment and on hours worked can explain

the effects of those policies on education, accumulated hours of work and human capital by age 30.

Our last objective is to answer the following question: would linear IV applied to data generated by a dynamic skill accumulation model estimate the average effects of education for the sub-population affected by any specific policy? Although there is a substantial methodological and empirical literature on IV estimation of the return to schooling, this issue has never been investigated formally.[1] Most of the literature on returns to schooling offers interpretations that are based on linearly separable Mincer representations of the wage equation (or even ignore work experience) and therefore remains agnostic about the potential existence of dynamic effects of education on post-schooling skill formation. To answer this fundamental question, we use model simulations to generate IV estimates of the returns to schooling associated to each policy and compare them to the average effects of education on wages for the sub-populations affected.

The main findings can be summarized as follows. The average entry-wage effect of education is around 10 percent per year of schooling although 27 percent of white males have very low returns (approaching 0). After netting out the effect of education, increasing hours of work is a major source of skill formation and is virtually as important as education in the early phase of the life-cycle. Each year of work experience with 2,000 hours or more raises wages by more than 6 percent per year although there is a very high level of dispersion across individuals.

After conditioning on observed and unobserved heterogeneity, each additional year of schooling raises both the return to one year of work experience and the probability of working more than 2,000 hours. After compounding the effect of education on returns and on hours worked, a college graduate would have a 18 percent higher probability of working 2,000 hours or more and would experience a 1.2 percent higher average wage growth rate per year of experience than a high school graduate. About one third of the total wage returns to schooling measured by age 30 are explained differences in growth rates induced by schooling.

However, despite the existence of a strong dynamic impact of education on post-schooling skill formation, differences in hours worked are at least as important as education and most of cross-sectional differences in wage growth

[1] We review briefly the IV literature in Section 2 as well as in Section 5.

(about 63 percent) remain explained by unobserved heterogeneity which may partly be interpreted as non-cognitive skills.

We find that a compulsory high school graduation policy would affect a lower ability population but would have a slightly larger effect on average schooling attainments than would a $5,000 reduction in the cost of college. Compulsory schooling, unlike the reduction in the cost of college, would succeed in raising hours worked by age 30. The reduction in the cost of college would affect individuals who would be more likely to work a high level of hours and less likely to engage in home production ex-ante. For them, the reduction in potential experience slightly dominates the positive effect of education on the intensive margin and therefore induces a small reduction in total hours worked by age 30. However, each policy would raise accumulated human capital (wages) at age 30. Compulsory schooling would increase wages by 19 percent by raising both education and hours worked of individuals who have low returns. The reduction in the cost of college would increase human capital by 17 percent essentially by raising education of individuals who have high returns to schooling despite a small reduction in total hours worked.

Finally, we find linear IV estimates of the return to schooling to be rather uninformative. As normally expected, IV estimates that condition on work experience (either exogenous or endogenous) always exceed those that do not but they exceed the average effects of those affected (compliers) by a significant margin in 5 out of 6 cases considered. Compulsory schooling estimates are particularly uninterpretable as they are much closer to the average effects of those unaffected. They disclose an interesting paradox since they exceed IV estimates generated by a reduction in the cost of higher education even if the latter policy affects individual with higher returns than those affected by compulsory schooling.[2]

The disconnection between IV estimates and the average effects of education for compliers are explained by two distinct causes. First, the per-year effects of education are complicated functions of education, work experience and heterogeneity and are in general not orthogonal to the policy exposure indicator (the instrument). This is true even if work experience depends only on the instrument through education. This feature is at odds with the independence assumption commonly invoked in the linear correlated random

[2]In the paper, the terms "compliers" and "affected" are used interchangeably.

coefficient model of Imbens and Angrist (1994) and Heckman and Vytlacil (2005) and implies that the effect identified by IV is most likely a combination of a wage change caused by variations in schooling induced by a specific policy as well as change in returns (the average effects) induced by changes in education and experience. A second reason is that linear IV estimates admit an average per-year effect of education interpretation when compliers change their schooling by one year but are also more difficult to interpret when it is not the case. Indeed, about half of those affected by our counterfactual policies change their schooling level by two years or more.

The remainder of the paper is organized as follows. In the next section, we review the relevant literature. In Section 3, we describe the model and discuss various estimation issues. Model estimates are presented in Section 4. In Section 5, we analyze the effects of the two policies; compulsory schooling and a reduction in the cost of higher education. Section 6 is devoted to the analysis of IV estimates of the returns to schooling. The paper ends with concluding remarks.

2 The Literature

Our model bridges gaps between the structural literature on education choices and some recent papers that estimate (or calibrate) structural models of earnings dynamics but do not incorporate education. Along those lines, Adda, Dustmann, Meghir and Robin (2006) have estimated a dynamic model of job mobility using a sample of German youths who have attended professional education while Bagger Fontaine, Postel-Vinay and Robin (2013) estimated an equilibrium search model of the Danish labor market set within a sequential auction framework. Altonji, Smith and Vidangos (2013) estimated a reduced-form model of earnings dynamics that incorporates hours of work, unemployment and job transitions. In the macroeconomic literature, Hugget, Ventura and Yaron (2011) have calibrated a Ben-Porath model of the US labor market using the PSID. In all of those cases, education is either ignored or assumed to be exogenous and the focus is on labor market frictions. [3]

In the structural education literature, the few papers that have considered education and labor supply jointly have focused on the extensive margin. In

[3]The recent literature on earnings dynamics is surveyed in Magnac, Pistolesi and Roux (2013).

their seminal piece, Keane and Wolpin (1997) model education, occupation choices and household production but ignore hours worked. Sullivan (2010) integrates the education-occupational choice framework developed by Keane and Wolpin with some key features of job search theory. Bravo, Mukhopadhay and Todd (2010) also model labor supply at the extensive margin in their analysis of Chile's voucher system. Todd and Wolpin (2006) model fertility and parental decisions about children's time allocation (schooling, labor market participation) in rural Mexico. In all cases, the authors disregard labor supply at the intensive margin and therefore disregard learning-by-doing induced by modulating hours worked. None of those papers consider the dynamic effects of schooling on life-cycle income profiles nor do they consider heterogeneity in income or wage growth.

The contributions of our paper are not confined to the literature on earnings dynamics and education choices. Because education and labor supply (hours worked) act as competing inputs to skill production and because some individuals are more effective at producing skills in the market while others are relatively more productive in school, our model also contributes to the emerging literature on labor supply and human capital formation.[4]

Our paper also adds to the literature on ex-ante evaluation of education policies. Keane and Wolpin (1997) used their model to simulate welfare changes induced by a reduction in the cost of college and found it to be partly ineffective at reducing life-cycle inequality. Eckstein and Wolpin (1999) estimated a structural model of high-school attendance, employment (while in school) and academic performance but ignored post-high school skill accumulation. They evaluated the effects of policies that would limit employment while attending high school and report that such a policy would increase graduation by no more than 2 percentage points. As far as we know, no paper has ever provided a comparative analysis of compulsory schooling and higher education subsidies (or cost reduction) within an integrated framework.

Finally, our paper complements the voluminous literature on estimating returns to schooling. In the vast majority of the literature, applied econometricians use IV methods to estimate a single parameter that subsumes

[4]The importance of allowing for human capital formation within models of labor supply is argued in Imai and Keane (2004) and also discussed at length in Keane and Rogerson (2012).

all dimensions of the returns to schooling into a scalar. There exist many surveys of the IV literature on returns to schooling. Card (1999) surveys the earlier literature and stresses the local average treatment effect (LATE) interpretation. Cameron and Taber (2004) also surveys the literature and present some compelling arguments explaining why low IV estimates tend not to be reported in the earlier literature. Belzil (2007) surveys the structural literature and focuses on the discrepancies between IV and structural estimates.

Although there exist a large number of existing papers that have debated the advantages and disadvantages of IV estimation at a methodological level, all papers devoted to the IV estimation of returns to schooling have based their analysis on simple representations of wage equation in which schooling is the only endogenous variable and have for the most part disregarded potential endogeneity of work experience.[5] As it stands now, there exists no quantitative analysis of the performance of IV estimates within a framework allowing for a complex post-schooling skill accumulation process in which both heterogeneity and dynamic schooling effects may interplay.[6]

Finally, it should be noted that our model shares similarities with the one analyzed in a companion paper (Belzil, Hansen and Liu, 2016), but it also differs with respect to some key features. In the latter, the dynamic effects of education are introduced through interaction terms and the intensive margin dimension of the model is not as rich as it is in the current paper. More importantly, our companion paper focuses on the evolution of inequality and on various economic implications such as the effect of taxation on skill accumulation but ignores counterfactual education policies.

3 Model

We estimate a stochastic dynamic discrete choice model of education and labor supply with human capital accumulation over the early life-cycle. We model choices from age 16 until age 30. To incorporate decisions at extensive

[5]See Imbens and Angrist (1994), Angrist and Imbens (1995), Heckman and Vytlacil (2005) and Heckman, Vytlacil and Urzua (2006).

[6]Ge (2013) uses simulated data from a structural model to analyze OLS and IV bias arising when estimating returns to schooling. The underlying data generating process is however not as general as the one we estimate in this paper.

and intensive margins, we partition annual hours of work into three intervals: low intensity (l) corresponding to fewer than 2,000 hours per year, medium intensity (m) corresponding to 2,000-2,499 hours per year, and finally high intensity (h) corresponding to 2,500 hours per year or more. These interval are easily interpretable in terms of part-time, full-time and extra full-time employment. As an example, individuals working 50 weeks per year at 40 hours per week would fall in the medium category. An individual working persistently overtime hours or holding multiple jobs, and who would work 50 hours per week, would fall into the high category.

In addition to labor supply decisions, we model schooling (s) and a residual state (r) which is meant to capture the activity of those who did neither work nor attended school during the year.

Individuals maximize the expected value of lifetime utility. The state-specific utilities are defined below. The choices are summarized in the binary indicators, d_{tk}, where $d_{tk} = 1$ when option k (s, l, m, h, r) is chosen at date t. The variables corresponding to the capitalized letters $(S_t, L_t, M_t, H_t, R_t)$ are used to measure the number of periods accumulated in each state when entering date t.

3.1 Employment

To estimate the model, we first set the per-period utility of the residual state (state r) as our benchmark and normalize it to 0. To separate pecuniary human capital accumulation motives from other components such as leisure or distaste (stigma) for marginal attachment to the labor force, we assume that the utility of employment depends on log wages (denoted w_{it}), accumulated schooling and an additive heterogeneity term measuring individual specific differences in the valuation of work intensity.

The per-period utility equations are defined as follows:

$$U_{it}^h = \alpha_i^h + \delta_w^h \cdot w_{it} + \delta_s^h \cdot S_{it} + \delta_h^h \cdot H_{it} + \varepsilon_{it}^h \tag{1}$$

$$U_{it}^m = \alpha_i^m + \delta_w^m \cdot w_{it} + \delta_s^m \cdot S_{it} + \delta_m^m \cdot M_{it} + \varepsilon_{it}^m \tag{2}$$

$$U_{it}^l = \alpha_i^l + \delta_w^l \cdot w_{it} + \delta_s^l \cdot S_{it} + \delta_l^l \cdot L_{it} + \varepsilon_{it}^l \tag{3}$$

where δ_w^h, δ_w^m and δ_w^l measure the effect of wages on utilities. The equation describing the log wage function is presented below. The parameters δ_s^h, δ_s^m, and δ_s^l capture the dynamic effects of schooling and allow us to take into account that education may affect the disutility of work effort. The heterogeneity terms α_i^h,$\alpha_i^m$and α_i^l are essential to our analysis as working more hours may deprive individuals from leisure consumption.[7] In standard labor supply models, their pendant would usually be represented by a single parameter (assumed to be homogenous) capturing the marginal utility of leisure and determine labor supply adjustments at the intensive margin.[8] Our model is therefore flexible enough to allow some individuals to prefer high, low or medium intensity labor supply, for a given hourly wage. More details about the specification of the heterogeneity terms are found below. Finally, δ_h^h, δ_m^m and δ_l^l are parameters that allow to capture persistence in choices, and may be explained by the existence of market frictions or habit formation, while $\varepsilon_{it}^h, \varepsilon_{it}^m$ and ε_{it}^l are idiosyncratic random shocks described below.

3.2 Schooling

The utility of attending school (state s) for individual i at time t is denoted U_{it}^s, and is defined as

$$\begin{aligned} U_{it}^s &= \alpha_i^s + \delta_1^s \cdot I(S_{it} = 11) + \delta_2^s \cdot I(12 \leq S_{it} < 14) \; + \\ \delta_3^s \cdot I(14 &\leq S_{it} < 16) + \delta_4^s \cdot I(16 \leq S_{it}) + \delta_5^s \cdot I(d_{t-1,s} = 0) + \; \varepsilon_{it}^s \end{aligned} \tag{4}$$

where $I(.)$ is the indicator function. The parameters $\delta_1^s, \delta_2^s, \delta_3^s$ and δ_4^s capture the variation in the utility of attending school with grade level. These parameters are standard in the education literature (Keane and Wolpin, 1997). The parameter δ_5^s captures the psychic cost of re-entering school for those who are currently not enrolled. The term α_i^s represents individual heterogeneity in taste for schooling (academic ability). Finally, ε_{it}^s is a stochastic shock.

[7] An alternative interpretation is that those heterogeneity terms may capture differences in the cost of work effort or differences in the disutility (stigma) associated to a weak participation to the labor market.

[8] See Keane and Rogerson (2012).

3.3 The Learning-by-Doing Technology

The skill accumulation technology encompasses both an effect of education on the entry level of wages and a life-cycle effect captured by allowing the impact of labor supply on skill formation to depend on education. Individual differences in returns to work experience are therefore partly determined by individual specific time invariant ability as in the literature on heterogeneous income profiles and by schooling and labor supply decisions (hours worked).

The log wage function is given by the following equation

$$\begin{aligned} w_{it} &= \lambda_i + \lambda_i^s \cdot S_{it} + \lambda_i^l(S_{it}) \cdot L_{it}(S_{it}) \\ &\quad + \lambda_i^m(S_{it}) \cdot M_{it}(S_{it}) + \lambda_i^h(S_{it}) \cdot H_{it}(S_{it}) \quad + \varepsilon_{it}^w \end{aligned} \tag{5}$$

where the $\varepsilon_{it}^{w\prime}s$ represent stochastic wage shocks and where the dependence of L, M and H on schooling is meant to underscore the dependence of hours worked on accumulated years of schooling. In the descriptive (statistical) literature on earnings dynamics, it is common to argue that persistence in earnings may not only be explained by persistent unobserved heterogeneity but also by persistent wage shocks. In our model, utility shocks are i.i.d. but are indirectly persistent since they are affecting endogenous skill accumulation decisions which are persistent by nature. So, in order to take into account the possibility that skill prices may also be affected by persistent shocks, we also allow ε_{it}^w, to follow an autoregressive process. We limit ourselves to an AR(1) process. This implies that

$$\varepsilon_{it}^w = \rho \cdot \varepsilon_{it-1}^w + \upsilon_{it}^w \quad \textit{with} \quad 0 < \rho < 1$$

where υ_{it}^w is an i.i.d. normal random term with mean 0 and variance σ_υ^2 and where ρ is a parameter to be estimated.

Finally, and since we focus on the early stage of the life-cycle, we do not allow for concavity in age earnings profiles. All parameters capturing the returns to hours worked $(\lambda_i^l, \lambda_i^m, \lambda_i^h)$ depend on realized schooling so to capture the dynamic effects of education.

3.4 Heterogeneity

We allow all heterogenous components of the model (utilities and skill formation technology parameters) to depend on Armed Forces Qualification Tests

(AFQT) scores as well as an unobserved component orthogonal to AFQT scores.[9]

The individual-specific terms of the utility functions are parameterized as follows:

$$\alpha_i^k = \tilde{\alpha}_i^k + \alpha^{ka} \cdot AFQT_i \quad for \ k = s, l, m, h \tag{6}$$

where all $\alpha^{ka\prime}s$ are freely estimated and where the $\tilde{\alpha}_i^{k\prime}s$ represent unobserved heterogeneity orthogonal to AFQT scores.

The parameters of the log wage function are defined as follows:

$$\begin{aligned}
\lambda_i &= \tilde{\lambda}_i + \lambda^a \cdot AFQT_i \\
\lambda_i^s &= \exp\{\tilde{\lambda}_i^s + \lambda^{sa} \cdot AFQT_i\} \\
\lambda_i^h(t) &= \exp\{\tilde{\lambda}_i^h + \lambda^{ha} \cdot AFQT_i + \lambda^{hs} \cdot S_{it}\} \\
\lambda_i^m(t) &= \exp\{\tilde{\lambda}_i^m + \lambda^{ma} \cdot AFQT_i + \lambda^{ms} \cdot S_{it}\} \\
\lambda_i^l(t) &= \exp\{\tilde{\lambda}_i^l + \lambda^{la} \cdot AFQT_i + \lambda^{ls} \cdot S_{it}\}
\end{aligned}$$

where $\tilde{\lambda}_i, \tilde{\lambda}_i^s, \tilde{\lambda}_i^h, \tilde{\lambda}_i^m, \tilde{\lambda}_i^l$ represent unobserved heterogeneity orthogonal to AFQT scores while $\lambda^a, \lambda^{sa}, \lambda^{ha}, \lambda^{ma}, \lambda^{la}$ measure the contribution of cognitive skills to abilities. Assuming orthogonality between unobserved types and AFQT scores allows us to interpret differences across types as differences in non-cognitive skills. Another approach would have been to model type probabilities as a function of AFQT scores but doing so would have presented us to obtain clear estimates of the effects of cognitive ability on various heterogeneity components.

We assume that the unobserved heterogeneity distribution can be approximated by a multi-variate discrete distribution with four types. Each type q is endowed with the following vector of initial endowments (at age 16): $\{\tilde{\alpha}_q^s, \tilde{\alpha}_q^l, \tilde{\alpha}_q^m, \tilde{\alpha}_q^h, \tilde{\lambda}_q, \tilde{\lambda}_q^s, \tilde{\lambda}_q^h, \tilde{\lambda}_q^m, \tilde{\lambda}_q^l\}$. The type probabilities are expressed as follows:

$$\Pr(type = q) = p_q = \frac{\exp(\delta_q + \delta_{sq} \cdot S_{16,i})}{1 + \sum_{n=2}^{4} \exp(\delta_n + \delta_{sn} \cdot S_{16,i})}$$

[9] We also experimented with specifications that incorporate measures of non-cognitive skills such as the Rotter Locus of Control Scale and the Rosenberg Self-Esteem Scales,but found those to be insignificant once unobserved types were incorporated in the model.

where $S_{16,i}$ denotes initial schooling (grade level achievement at age 16). This means that AFQT scores and unobserved types are mutually orthogonal only after conditioning on grade completed by age 16.

3.5 Estimation

The elements of the vector of utility error terms $\{\varepsilon_{it}^s, \varepsilon_{it}^r, \varepsilon_{it}^h, \varepsilon_{it}^m, \varepsilon_{it}^l\}$ are assumed to be i.i.d. and to follow an extreme-value distribution. At each period t, the individual makes a decision based on the information set which includes the random shocks and accumulated periods in each state:

$$\Omega_t = \{\varepsilon_t^s, \varepsilon_t^h, \varepsilon_t^m, \varepsilon_t^l, \varepsilon_t^w, S_t, R_t, L_t, M_t, H_t\}$$

We model choices from age 16 onward over a total time horizon equal to 15 years (until age 30). For each possible choice, there is a specific value function, $V_t^k(\Omega_t)$, equal to

$$V_t^k(\Omega_t) = U_t^k + \beta EV_{t+1}(\Omega_{t+1} \mid d_{kt} = 1) \text{ for } k = s, r, l, m, h$$

where

$$\begin{aligned} EV_{t+1}(\Omega_{t+1} \quad &| \quad d_{kt} = 1) \\ &= \quad E \max\{V_{t+1}^s(.), V_{t+1}^r(.), V_{t+1}^l(.), V_{t+1}^m(.), V_{t+1}^h(.) \mid d_{kt} = 1\} \end{aligned}$$

where β is the discount factor.

Despite the extreme-value distribution assumption about the utility shocks, solving for the maximum lifetime utilities requires simulating the distribution of the wage shocks. To reduce computation burden, we follow Sauer (2015) and adopt a solution method that borrows from Geweke and Keane (1996, 2000) who have proposed to replace the future component of the value function by a flexible polynomial in state variables. Their approach is particularly well suited to frameworks where the econometrician has access to data on choices and outcomes. Geweke and Keane (1996) actually show from various numerical applications to artificial data that specifying the future component as a flexible polynomial has negligible effects of the estimated values of the parameters of the payoff functions, and that the mis-specified rule inferred from the data is itself very close the actual optimal rule.

As is done in Sauer (2015), we adjust the Geweke-Keane solution approach to incorporate more of the model structure in our estimation strategy.[10] At any period t, the future component of the intertemporal utility, $EV_{t+1}(\Omega_{t+1} \mid d_{kt} = 1, \Omega_t)$, is represented by the following expression:

$$EV_{t+1}(\Omega_{t+1} \mid d_{kt} = 1, \Omega_t) = E \max_k \{U^k_{t+1}(\Omega_{t+1}) + \mathcal{F}(\Omega_{t+2}(\Omega_{t+1}, d_{kt+1}))\} \quad (7)$$

where $\mathcal{F}(\Omega_{t+2}(.))$ is a flexible polynomial in state variables reflecting the relationship between $\Omega_{t+2}(.)$ and both Ω_{t+1} and d_{kt+1}. Our approach therefore differs from the approach suggested by Geweke and Keane (2000, 1996) in that the imbedded polynomial of the state space intervenes in $t+2$ as opposed to directly in $t+1$. This allows us to incorporate more of the model structure than in the original Geweke-Keane method. In a supplementary file, we provide more details about our estimation method and the form of the polynomial.

We estimate the model by simulated maximum likelihood techniques. For each individual i at date t, there is a vector of observed outcomes $O_{it} = \{d_{ist}, d_{irt}, d_{ilt}, d_{imt}, d_{iht}, w_{it}\}$. To estimate the model, we normalize τ to 1. The likelihood function for individual i is given by

$$L_i(.) = \sum_{q=1}^{4} \prod_{t=1}^{T} \Pr(O_{it} \mid type\ q) \cdot \Pr(type\ q) \quad (8)$$

The total likelihood is the product of each $L_i(.)$ over 1,199 individuals. Structural parameters are obtained by maximizing the logarithm of the likelihood function using Fortran routines.

4 Model Estimates

The model was estimated using a sample of white males taken from the 1979 youth cohort of the National Longitudinal Survey of Youth (NLSY). We restrict our sample to white males from the core random sample who were

[10] Compared to Sauer (2015), our model contains a smaller number of potential choices but it is estimated over a much longer period and also incorporates a richer heterogeneity distribution.

14 to 16 years old in 1979. More details regarding our sampling method are found in the supplementary file.

The model contains 84 parameters. As is often the case in complicated non-linear models, many parameters do not raise specific interest. For this reason, we simulate a large number of individual trajectories (119,900) and use simulated data to analyze the main properties of the model. Specifically, we simulate 100 trajectories for each individual (using estimated type probabilities) and end up with a total sample size equal to 119,900. This sample constitutes our control group which will be used later to evaluate counterfactual policies. Although this number may seem unduly high, we do so because the policies will be used to evaluate the capacity of IV estimates to target some population parameter and because IV estimates are known to be usually imprecise. A table containing all structural parameter estimates can be found in appendix.

4.1 Model Fit

In Table 1, we report the predicted number of accumulated periods in each state from age 16 to age 30 and compare them to actual frequencies. The model's capacity to fit the data is quite clear. As is the case in the data, our model predicts that the average white male will have spent between four and five years in school between age 16 and age 30 and will end up with an average grade level attainment equal to 13.3 years. The average white male also spends more than nine years in the labor market over the same period. The 2,000-2,500 hour range is the most common employment choice (around four years) while young individuals spend on average three years working less than 2,000 hours and two years working more than 2,500 hours.

4.2 The Skill Formation Technology and Educational Selectivity

The effect of schooling on entry wages (λ_i^s), as well as the returns to each type of work experience ($\lambda_i^h, \lambda_i^m, \lambda_i^l$) evaluated at the average predicted education level (13.3 years) are summarized in Table 2A. The average entry-wage returns to schooling is equal to 10.2 percent per year of schooling but its high standard deviation, equal to six percent, illustrates an important level

of cross-sectional dispersion. This is exemplified by the fact that about 27 percent of the white male population (type two and type four) has an entry wage return to schooling practically equal to 0 percent.[11]

At the same time, and after netting out the effect of education, the average returns to work experience (around three percent per year when working less than 2,000 and around six percent per year when working more than 2,000 hours) indicate that increasing hours of work is a major source of skill formation. This stresses the importance of the intensive margin as a major source of learning by doing. Most of the gain in skill formation associated to hours worked is captured when moving from the low level (1,999 hours or less) to the medium level (between 2,000 and 2,500 hours). However, there is also a high degree of dispersion characterizing the returns to work experience. For instance, it is interesting to note that both type two and type four individuals are more effective at producing skills in the market than in school. This is especially true about type four individuals who are endowed with very high returns to medium and high hours work experience, lying between eight and 10 percent per year.

One way to illustrate the importance of selectivity is to examine differences between high school and college graduates. To do this, we measure returns to schooling on entry wages and returns to different types of work experience for both groups. Unlike returns to work experience found in Table 2A, those documented in Table 2B do not only reflect differences in abilities between high school and college graduates but also differences in schooling.

There is a one percent differential in entry wage returns to schooling between the two groups as the average effect is equal to 9.4 percent per year for high school graduates and 10.3 percent for college graduates. This apparently mild difference is explained by the fact that most of the differences in skill accumulation rates are found at the level of the returns to work experience. For instance, college graduates earn a one percent differential over high school graduates for each year of experience when working between 2,000 and 2,500 hours per year and a 1.5 percent premium when working high hours. This tendency is actually explained by the structural parameters measuring the causal effect of schooling on the growth rates ($\lambda^{hs}, \lambda^{ms}$ and λ^{ls}) which are all found to be strictly positive and which also increase with

[11] Belzil and Hansen (2007) also find a significant fraction of white males with very low returns to schooling within a framework where post-schooling accumulation is exogenous.

hours worked. It should also be noted that the very high standard deviations of the returns to high hours for college graduates (0.03) also implies that a substantial fraction of college graduates will experience very high returns to work experience.[12]

4.3 Labor Supply

In our model, education affects the slope of age-earnings profiles not only because it raises the productivity of work experience (as documented in Table 2B) but also because education has an impact on the utilities of working at each specific hours level. The latter effect may be crucial since we already noted that working more than 2,000 hours per year conveys an additional three percent growth rate premium per year of experience. It is therefore interesting to measure by how much the frequency of the high-payoff labor supply states is increased by schooling, after conditioning on heterogeneity.

To do so, we construct the simulated fractions of non-school years spent working between 2,000 and 2,500 hours and working more than 2,500. Formally, we compute the ratios $\frac{M_{30}}{L_{30}+M_{30}+H_{30}+R_{30}}$ and $\frac{H_{30}}{L_{30}+M_{30}+H_{30}+R_{30}}$ for each individual and regress it on education outcomes and on heterogeneity components. The ratios are not affected by the automatic reduction in potential experience induced by schooling. This allows us to capture the causal effect of education on the intensive margin without introducing unduly an "opportunity cost" effect.

The estimates are found in Table 2C. The marginal effect of education on the incidence of medium and high hours (0.028 and 0.017), are easily interpretable in the context of a comparison between high school and a college graduates. After conditioning on observed and unobserved heterogeneity, a college graduate would have a 11 percent higher probability to work between 2,000 and 2,500 hours per year and a seven percent higher probability of working more than 2500 hours than would a high school graduate. However, and as indicated by the change in $R^{2\prime}s$ of the regressions observed when education is excluded, more than 90 percent of the differences in hours worked are explained by heterogeneity.

[12] As noted in Murphy and Topel (2016), the complementarity between education and the labor supply intensive margin may magnify inequality between college and high school graduates.

4.4 The Causal Effect of Education and Labor Supply on Life-Cycle Wage Growth

Early life-cycle wage growth is a relatively complicated object that depends on both endogenous choices (education and labor supply) as well as individual specific technological parameters which depend themselves on observed and unobserved heterogeneity as well as endogenous investment decisions. To quantify the causal effect of both education and the labor supply intensive margin on age-earnings profiles, we use simulated outcomes to obtain a measure of the average growth rate (per year of experience) realized by each individual and decompose it into four components; the fraction of non-school years working high and medium hours ($\frac{M_{30}}{L_{30}+M_{30}+H_{30}+R_{30}}$ and $\frac{H_{30}}{L_{30}+M_{30}+H_{30}+R_{30}}$), education, AFQT scores and unobserved heterogeneity. To do this, we use standard regression methods. The dependent variable of the regression is defined as $\frac{w_{i30}-\lambda_i-\lambda_i^s \cdot S_{i30}}{L_{i30}+M_{i30}+H_{i30}}$. The regressions are summarized in Table 3.

The marginal effect of education on the average growth rate per year, equal to 0.003, point to the evidence that a fair share of the returns to schooling are captured beyond entrance in the market.[13] Illustrated in terms of the usual high school-college differential, this causal effect implies the existence of a 1.2 percent differential in realized growth rates in favor of college graduates. It is important to note that even after taking into account the effect of education on the intensive margin, differences in hours worked are practically more important than education. The effect of an increase in the frequency of the medium hours range, which is equal to 0.03, is three times larger than an increase in the high hour range frequency. This is consistent with the fact that working more than 2,500 hours is more productive than the medium range only for a subset of the population. After conditioning on schooling, AFQT and unobserved type, working systematically 2,000 or more would therefore generate a supplementary average growth rate between one and three percent per year. Finally, unobserved heterogeneity is a far more important determinant of annual wage growth than both schooling and the intensive margin as it accounts for about 63 percent of the explained part of

[13] For instance, the early career wage growth realized by U.S. white males documented in earlier papers such as Topel and Ward (1992) and Taber (2001) averages 10 percent per year around age 25. See Taber (2001) for a review of the earlier literature.

wage growth.

These results are in accord with those reported in Keane and Wolpin (1997), who find that unobserved types account for a large share of life-cycle earnings inequality, although their model relies on occupation-specific Mincer equations. They are also consistent with findings reported in Belzil, Hansen and Liu (2015) in which we find that unobserved heterogeneity is the dominant factor behind wage growth over the early phase of the life-cycle but that the importance of cognitive skills and education increases as individuals approach age 50. At the same time, wage growth remains mostly explained by stochastic shocks as indicated by the relatively low R squares.

4.5 OLS Estimates

The dichotomy between the entry-wage return to schooling and the dynamic effects of schooling realized over the life-cycle is a key feature of our model. It cannot be addressed within classical reduced-form or standard IV approaches. However, it is important to see if our model is also capable of generating OLS estimates similar to those often reported and in particular to those obtained on our sample of white males at age 30. To verify this, we estimated a Mincer regression using simulated wages and outcomes (at age 30) and examined the sensitivity of the OLS estimates of education to the removal of AFQT scores. Then we estimated the same specification on our sample of white males using wage data measured at age 30. The results, found in Table 4, are clearly coherent with patterns reported in the literature. First, the OLS estimate of the effect of education on simulated wages is equal to 0.12 when AFQT scores are omitted (column three). Second, and as is often noted in the empirical literature, the OLS estimate drops when AFQT scores are included. In the present case, the drop to 0.09 (column four) represents a 25 percent decrease in the OLS estimate. These results indicate clearly that our model is capable of generating features of the wage distribution similar to those reported in the applied literature.[14] It is also striking to note that the OLS estimates obtained from simulated data are practically equal to those obtained on actual data (columns 1 and 2).

Finally, another specificity of our model is the allowance for persistence in wage shocks, as we model the stochastic term affecting wages as an AR(1)

[14]See Cameron and Taber (2004) for a discussion.

process. Our estimate, which is found in the third section of the supplementary file (along with other structural parameters) is equal to 0.57 (with a standard error equal to 0.09) and indicates a low level of persistence in wage shocks. It therefore implies that wage persistence is mostly explained by heterogeneity.

4.6 The Reduced-Form Effects of Education

Until now, we have examined three different structural components of the returns to education. Those are the entry-wage effect, the effect of education on the productivity of work experience, and the effect of education on hours worked. All those effects are structurally interpretable. However, the optimal skill investment problem is essentially about time allocation. Education reduces potential experience and has therefore also an indirect effect on wages.

In anticipation of our analysis of IV estimation to be presented in Section 6, we now evaluate the reduced-form effect of education on wages. As will become clear later, those estimating returns to schooling using IV techniques are not capable of estimating the structural components of the skill accumulation technology but may target an average effect of education for a sub-population.

In a classical Mincer model, measuring the total effect would be trivial since the effects of education and experience are linearily separable. In our model, the opportunity cost of education is a complicated object that depends itself on the individual specific return to education, on the accumulated level education (since education affects the return to work experience) and on the individual specific components of the returns to various types of experience.

The total effect (at a given age) and its components are defined as:

$$\textit{Total Effect} = \text{Partial Effect} + \text{Experience Loss} \tag{9}$$

where each component are expressed as:

$$\text{Partial Effect} = \lambda_i^s + \frac{\partial \lambda_i^h(.)}{\partial S} \cdot H_{i,age} + \frac{\partial \lambda_i^m(.)}{\partial S} \cdot M_{i,age} + \frac{\partial \lambda_i^l(.)}{\partial S} \cdot L_{i,age} \tag{10}$$

$$\text{Displacement Effect} = \lambda_i^h(S_{it}) \cdot \frac{\partial H_{i,age}}{\partial S} + \lambda_i^m(S_{it}) \cdot \frac{\partial M_{i,age}}{\partial S} + \lambda_i^l(S_{it}) \cdot \frac{\partial L_{i,age}}{\partial S} \tag{11}$$

The partial effect measures the marginal effect of education on skill formation holding accumulated experience fixed. The total effect is the sum of the partial effect and the experience loss term, which itself measures the opportunity cost of education investment (the wage penalty of reducing work experience). Both the experience loss and the partial effects (and therefore the total effects) are individual specific quantities that depend non-linearly on heterogeneity and realized choices. This feature implies that our outcome equation cannot be reduced to the classical correlated random coefficient model which has been analyzed at length in Imbens and Angrist (1995) and Heckman and Vytlacil (2005).

The population averages of the partial effect, the wage cost of education and the total effects are reported in Table 5 along with bootstrapped standard errors. To illustrate how these effects change over the life-cycle, we measured them at age 25, 30, 35 and 40.

At age 30, the mean partial effect of education in the population is equal to 0.15 and therefore implies the realization of an average supplementary 5 percent return to schooling after labor market entrance (as the population average entry effect is around 0.10). This means that by age 30, about one third of the wage return to schooling of an average white male has been realized beyond entrance in the labor market. Not surprisingly, the total marginal effect associated to the early life-cycle is much lower. This is explained by the reduction in work experience induced by education. Our estimate of the wage displacement effect is equal to -0.037. In total the average marginal effect of education in the population is equal to 0.11 and is therefore 26 percent smaller than our estimate of the partial effect.

5 Analyzing Counterfactual Education Policies

We focus on two different types of interventions; a compulsory high school graduation policy and a reduction in the cost of college attendance. Both of them constitute policy interventions intensively used in the applied literature on returns to schooling. For instance, empirical labor economists often consider distance to college (or presence of a college within a county) as a measure of the cost of higher education and use it as an instrument to measure returns to schooling. As well, changes in compulsory schooling age that took place in most western countries over the second half of the 20th century have also been used as instrument to estimate returns to schooling.[15]

The interventions are defined as follows.

- *Compulsory High School Graduation*: This policy intervention dictates school attendance for the first x_i periods, where x_i is defined as the difference between 12 (the minimum required) and initial schooling attainment (recorded by age 16). Formally, we impose the following restriction:

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

  and assume that individuals start optimizing at date $t_i = x_i + 1$.

- *Reduction of the Cost of Higher Education (College)*: To generate a realistic value, we calibrate the utility change on the average full-time equivalent wages observed at age 20. To do this, we make use of the fact that the estimated preference parameters imply that the utility of working 2,000 hours is approximately equal to the utility of attending grade 13 to 16 and re-interpret the average utility of attending higher education in monetary terms. We then multiply the average wage per hour at age 20 by 2,000 hours to obtain a value equal to \$16,000 (in \$1997) and then assume the existence of a subsidy of \$5,000 on an annual basis (corresponding approximately to 35 percent of the full-time equivalent income).[16] The amount of the decrease corresponds

[15] See Card (1999) or Cameron and Taber (2004) who both review the IV literature on returns to schooling.

[16] For a comparison, in Keane and Wolpin (1997) the average full-time equivalent wage

approximately to a reduction corresponding to a 80 percent reduction in the total direct cost (net of institutional transfers) of attending a 4-year college over the early 1980's (Abel and Deitz, 2014).[17] To translate this into a change in the net utility of attending school we simply apply a 35 percent increase by manipulating the parameters of the utility of attending school.

5.1 The Identity of those Affected

Before measuring the impact of each policy on education and labor supply, we investigate the distribution of the counterfactual changes in schooling induced by each policy in Table 6A and the identity of those affected and unaffected in the upper section of Table 6B. In the modern IV terminology, those affected are referred to as "compliers" whereas those unaffected comprise both the "never takers" and the "always takers".[18] To do so, we compute the average values of some of the key structural parameters such as the wage intercept (λ_i), the utility of attending school (α_i^s) as well as the effect of schooling on entry wages (λ_i^s). To ease presentation, all parameters have been standardized. A negative (positive) entry in Table 6B therefore indicates that a particular group is below (above) population average.

Approximately 18 percent of the population is affected by compulsory schooling. However, a fair share of those affected increases their schooling level by more than one year. Among those reacting to the policy, 57 percent reacted by two or more years. In total, compulsory schooling raised educational attainments by 1.9 years.

As it may easily be inferred from Table 6B, compulsory schooling affects the bottom tail of the skill distribution (those who have lower taste for schooling and lower wage entry returns) and thereby generates a clear discrepancy between those affected and unaffected. This is particularly visible at the level of the utility of attending school, as the difference in average

rate is approximately equal to $13,000 (in 1988 dollars). They simulate a college subsidy of $2,000 dollars which has a much smaller impact on average schooling attainments than ours.

[17] Ehrenberg (2012) discusses the long run evolution of the cost of higher education in the US.

[18] Because the Monotonicity property applies to our model (by construction), the set of potential "defiers" is empty.

standardized utilities between those affected and unaffected is about 1.6 standard deviations (-1.29 for the affected and 0.28 for those unaffected). The difference in average entry-wage return to schooling between those affected and unaffected is also important as it is approximately equal to 1.1 standard error (-0.9 standard error for compliers and 0.2 standard error for those unaffected).

The experiment that reduces the cost of higher education works differently. First, it affects a slightly larger fraction of the population (about 22 percent) but generates an increase in schooling, equal to 1.6 years, which is smaller than compulsory schooling. About 38 percent of compliers reacted by two years or more. As was the case with mandatory schooling, a significant fraction of the compliers react by more than one year.

The reduction in the cost of college affects mainly two sets of individuals; those who were at the margin of entering higher education or those who would have participated in higher education (who would have obtained 13, 14 or 15 years of schooling ex-ante) but would not have graduated. The sub-population of those unaffected comprises two completely distinct groups; those individuals with very low returns and low taste for education as well as high ability individuals who would have graduated from college in absence of the policy. When averaged together, those two groups generate a sub-population of individuals whose average endowments are not as different from those affected. This is a key difference with compulsory schooling.

More precisely, those affected by the reduction in cost of college have a slightly higher return to schooling at entrance in the market but the difference is only 0.14 of a standard deviation (0.11 vs.-0.03). The difference is more pronounced at the level of the utility of attending school as the difference is about 0.64 standard deviation (0.50 vs. -0.14) but it is still much lower than its counterpart observed for compulsory schooling (which was equal to 1.6 standard deviations).

This result is not innocuous. Although the set of individuals affected by a change in the cost of education may contain some individuals with high returns and low utility of attending school (such as those who are endowed with a low level of consumption while in school because they face liquidity constraints), our results imply that it is dominated by individuals with relatively high utility of attending school. While our model does not incorporate explicit liquidity constraints, we expect that more able individuals who are prevented from attending college should be characterized by relatively high

market abilities and low utility of attending school, thereby explaining their decision not to attend college in absence of the counterfactual policy. Our results are therefore only partly consistent with the popular claim that individuals affected by a decrease in the cost of college attendance are individuals with high returns to schooling who face liquidity constraints. More precisely, our findings suggest that a decrease in the cost of college will primarily affect individuals with high returns to schooling but who are also endowed with high returns to market experience and who would not attend college unless its opportunity cost is lowered.[19]

5.2 The Effects of Education Policies on Education, Labor Supply and Human Capital

We now investigate how differences in tastes and abilities between those affected and unaffected by each specific policy translate into changes in education, labor supply (hours worked) and human capital (wages) by age 30.[20] The results are found in the lower portion of Table 6B. To compute the effect of each policy on total labor supply, we use the mid-points of the intervals and construct the following expression: $\sum_t \{d_{tl} \cdot 1000 + d_{tm} \cdot 2225 + d_{th} \cdot 3000\}$.

Because compulsory schooling affects lower skill individuals who are more likely to work low hours, and because schooling also affects the utility of employment, its impact on total hours worked is sizeable. Despite the inherent reduction in potential experience caused by education, compulsory schooling raises total hours worked by 11.3 percent. This is because the potential experience loss induced by spending more years in school is compensated by a reduction in home time, or by an increase in the likelihood of working long hours.

As seen earlier, the reduction in the cost of higher education tends to affect more able individuals. Those individuals are ex-ante more likely to work

[19] A different approach to modeling barriers to education would be to model labor supply while in school as in Keane and Wolpin (2001) or to allow for intermittent employment periods (delayed college entrance) devoted to higher education financing (Johnson, 2015). Keane and Wolpin conclude that liquidity constraints are reflected mostly in differences in labor supply while in school. Johnson (who uses the 1997 cohort of the NLSY) concludes that borrowing constraints have a minor impact on college enrollments.

[20] To compute total labor supply, we use the mid-points of the intervals and obtain the following expression: $\sum_t \{d_{tl} \cdot 1000 + d_{tm} \cdot 2225 + d_{th} \cdot 3000\}$.

2,000 hours or more, and are also less likely to involve in home production when compared to those affected by compulsory schooling. For these reasons, it has a much smaller impact on total hours worked. Indeed, we find a small negative impact, equal to -3.4%, which indicates that the reduction in potential experience dominates the positive effect of education on the intensive margin.

Finally, despite their divergent effect on hours worked, both policies would translate into a relatively similar increase in human capital (wages) by age 30. Compulsory schooling would raise human capital (wages) by about 19 percent. It would do so by raising both education and hours worked of individuals who have low returns to human capital investment. The reduction in the cost of college would only be slightly less effective as it would raise wages by 17 percent. This would be achieved essentially by raising education of individuals who have relatively high returns to schooling, despite a small reduction in total hours worked.

6 What Would IV Estimate?

We now ask the following question: would linear IV applied to data generated by a dynamic skill accumulation model estimate the average effects of education for the sub-population affected by any specific policy? There is a vast empirical literature devoted to IV estimation of returns to schooling but there is no quantitative analysis of the performance of IV estimates within a framework that merges some attributes of the treatment effect literature on schooling with a earnings dynamics model. At the same time, empiricists systematically apply IV methods to cross-sectional data on earnings and schooling, which are themselves most likely characterized by dynamic schooling effects such as those documented in Section 4. For this reason, it is fundamental to investigate what does IV deliver in such a context.

6.1 Methodological Controversies and Empirical Literature

Although the literature starts in the early 1990's, there remains an impressive level of controversy surrounding the interpretation of the numerous estimates reported in the empirical literature. This is exemplified in a volumi-

nous methodological literature that evaluates the relevance of IV estimation strategies. Although our objective is not to present a detailed survey of the literature, we now sketch its evolution and show how our approach complements this vast literature.

One major source of controversy is concerned with the relevance of IV as an estimation strategy. It is now widely recognized that in the presence of multiplicative heterogeneity in the outcome equation, IV fails to deliver a structurally interpretable (policy invariant) parameter and must be interpreted as an instrument dependent quantity. In a seminal piece, Imbens and Angrist (1994) have discussed conditions under which IV may still converge to an interpretable quantity and introduced the notion of Local Average Treatment Effect (LATE). Heckman and Vytlacil (2005) and Heckman,Vytlacil and Urzua (2006) have pointed out the importance of the first-stage specification when estimating the LATE parameter.[21] Other criticisms have focussed on economic interpretations of the independence assumption (Rosenzweig and Wolpin, 2000, and Keane, 2010).

At an empirical level, Card (1999) surveys a wide range of papers published mostly over the 1990's that essentially used education policy changes to estimate the effects of schooling on wages and earnings. Most of the papers use policy changes that affected the cost of college or compulsory schooling age. In all the papers surveyed in Card (1999), authors instrument out schooling using an indicator that records exposure to a specific policy reform. In some of them, the authors also condition on labor market experience thereby ignoring potential endogeneity issues caused by labor supply decisions at the extensive or intensive margins or by any other forms of post-schooling human capital investment. In others, experience is allowed to be endogenous and is instrumented-out using age.[22] In virtually all cases where labor market is considered, authors use potential experience. Finally, an alternative strategy is simply to ignore work experience and consider a wage equation specification in which education is the sole regressor.

Most of the estimates obtained for the US and reported in Card (1999) are between 0.10 and 0.15. In most cases, reported IV estimates exceed their OLS counterparts. One common interpretation is that many policies aimed at stimulating education may potentially affect high ability individuals who

[21] Heckman (2010) surveys the Treatment Effect literature.

[22] This approach is achieved in Cameron and Taber (2004).

would not attend higher education ex-ante but are induced to do so when faced with a new policy environment. This interpretation, based on the concept of Local Average Treatment Effects, hinges on the validity of the IV orthogonality (independence) and the monotonicity assumptions.

6.1.1 IV Estimates obtained from Simulated Data

Before analyzing IV estimates obtained from simulated data, we must choose the population parameter estimand to which IV should naturally be compared to. In the popular linear correlated random coefficient specification of the wage equation, education is the only endogenous variable and the slope parameter is assumed to be orthogonal to the policy change indicator. The independence condition can be used to generate a clear population parameter. This is not the case here. The orthogonality between policy exposure and the heterogeneity distribution is not sufficient to deliver an easily interpretable analytical expression since changes in wages induced by a given policy reform cannot be solely attributed to variations in education.[23] To see this, it is sufficient to note that the effects of education (equation 10 and equation 11), which are the pendant of the individual specific slopes in the linear correlated random coefficient model, depend directly on schooling and accumulated experience. Note that this is the case even if our model implies that the dependence of accumulated experience on the instrument is solely explained by schooling.

Although it is certainly too demanding to expect IV to estimate a structurally interpretable quantity, it is however natural to compare it with the average effects of education on wages for those affected by each specific policy. In cases where work experience is introduced in the wage equation, it is natural to compare the IV estimate to the average partial effect of education (equation 10). This estimand should compound both the effect that education has on entry wages as well as the component of post-schooling wage growth that is caused by education given experience. In cases where experience is ignored, it makes more sense to compare it to the total effect of education. The latter should also incorporate the indirect effect of education on experience reduction (equation 11) and therefore be smaller than the par-

[23] Obviously, those interested in estimating a model where schooling has dynamic effects may decide to rule out IV as an estimation method and focus on more general method of moment estimation techniques.

tial effect. In either case, the resulting estimand is a complicated non-linear function of the heterogeneity components and of the actual level of education and experience accumulated at the age at which they are measured.

Our approach allows us to answer two subsidiary questions. First, for a given model and along with a specific education policy, are IV estimates at least closer to the average effects of education of compliers than non-compliers? Second, do IV estimates obtained from the policy generating the highest average effects of education for compliers exceed IV estimates obtained from the other policy?

To proceed, we use each counterfactual policy to generate a treatment group which can be appended to our control group. More precisely, we simulate 119,900 individual trajectories under the compulsory high school graduation policy and 119,900 trajectories under the higher education cost reduction. To allow for IV strategies that take into account the potential endogeneity of work experience, we must allow for a sufficiently high degree of variation in age at which wages are sampled. At the same time, choosing a relatively wide age bracket also allows us to obtain a sample more representative of the cross sections used in the IV literature than the one obtained if we limited ourselves to age 30. To achieve our goal, we select randomly one wage per individual between age 25 and age 40. Experience is defined as it is usually in IV studies that make use of it; namely as the total number of periods spent in the labor market, thereby ignoring differences in hours worked. Extrapolating until age 40 is not likely to be a major drawback since concavity of the age earnings profiles does not set in before the late 40's or early 50's.

In total, we obtain a cross-section of 239,800 individuals for each policy. This represents an ideal IV setting as the treatment and control heterogeneity distributions are identical by construction.[24] For each policy change, we compute a set of three IV estimates reflecting the different approaches mentioned earlier.

In the first one, we ignore work experience. It is therefore implicitly introduced in the error term of the wage regression. This corresponds to the most popular specification of the wage equation found in the empirical

[24]In practice, those implementing IV estimation using a before-after comparison are subject to the curse of sample selection as individual wages are observed only when working and because labor force composition may be itself affected by the policy.

literature. In the second approach, we treat work experience as exogenous. Finally, in the 3rd approach, we instrument out experience with age. To obtain the average effects of education, we use the definitions introduced in Section 4 in conjunction with individual counterfactual reactions to each policy, and measure the relevant derivative at a randomly assigned age. It should therefore be clear that our population estimands depend not only on the identity of those affected, but also on the age structure of the cross-section generated by our selection mechanism.

Table 7 discloses different IV estimates that exhaust all three approaches with respect to the treatment of experience. We first review the compulsory schooling IV estimates. To start with, the estimands are so precisely estimated that their standard errors were practically equal to 0 up to 6 decimals. For this reason and in order to clarify the table, they are not reported in Table 7. As normally expected, the total and partial returns of those affected by compulsory schooling (0.068 and 0.094 respectively) are much below those observed for the unaffected (0.139 and 0.177). This is largely explained by the fact that compliers have a very low entry wage returns, as documented in Table 6B.

The first compulsory schooling IV estimate, equal to 0.136, has been obtained while ignoring work experience and is naturally compared with the total average effects of those affected (equal to 0.068) since the IV estimator will naturally incorporate the negative impact of education on experience accumulation. The estimate is about 10 standard errors above its corresponding average effect and any confidence interval set at a reasonable level would obviously fail to cover it.

This raises another question. Is the IV estimate at least closer to the average effect of those affected than unaffected? The answer is no as the total return of those unaffected is equal to 0.139. The first compulsory high school IV estimate therefore fails to capture the identity of those affected by its own instrument.

We now ask if the difference between the IV estimate and the average effects of education for those affected is due to the choice not to condition on work experience. The answer is clearly no. The second IV estimate, which assumes that experience is exogenous, is equal to 0.175. The third one is obtained after instrumenting out experience with age and is equal to 0.165. Because both of them are obtained in a framework where experience is introduced explicitly in the wage equation, they are both naturally compared with

the average partial effect of those affected. As noted earlier, this quantity is itself equal to 0.094. As for the first IV estimate, both of those estimates are well above their natural estimand as they lie between 11 and 12 standard errors above it. As was the case for the first IV estimate, the second and third estimates are also actually much closer to the partial effect measured for those unaffected and which is equal to 0.177.

IV estimates generated by compulsory schooling therefore appear to be uninformative of the average effects of education. In all three cases, they overestimate their natural population estimand and fail to capture to identity of those affected since they are closer to the effects of education of those who are not affected.

We now turn to IV estimates generated by a change in the cost of college. As expected, the total and partial returns of those affected by a change in the cost of higher education are higher than their compulsory schooling counterparts as they are equal to 0.127 and 0.160 respectively. As already noticed in Table 6B, the reduction in cost generates much smaller differences between compliers and non-compliers in terms of the most important structural parameters. This translates into small differences between total and partial returns of non-compliers, which are equal to 0.126 and 0.163 respectively, and those of compliers.

The IV estimates generated by a reduction in the cost of college are relatively closer to the average effects of schooling of compliers. The first IV, which ignores work experience, and which is equal to 0.1293, exceeds its natural estimand (equal to 0.1270) by three standard errors. Because this experiment generates compliers and non-compliers that share practically common average returns, it is irrelevant to ask if it is closer to the effects of those affected than unaffected. The IV estimate obtained when experience is assumed to be exogenous is equal to 0.1741 and is approximately two standard deviations away from its relevant partial effect, equal to 0.1603. Finally, the one obtained when experience is endogenous is equal to 0.1613 and is the closest to the average partial effect of compliers as it is approximately 1.5 standard deviation above it. A confidence interval would therefore cover the average effect of schooling of compliers in only one of those three cases with the reduction in the cost of college.

There are two main points to be retained from the IV estimates generated by our counterfactual experiments. First, compulsory schooling estimates are totally uninformative about average marginal effects of compliers as they lie

between 10 and 12 standard errors away from their corresponding average effects. Second, IV estimates also disclose an interesting paradox in that compulsory schooling estimates are higher than the education subsidy estimates even though the average effects of those who comply with changes in compulsory schooling are only about half the average effects of those affected by a cost reduction. This provides supplementary evidence against the capacity of compulsory schooling IV estimates to capture the identity of those affected by the policy.

Although our outcome equation is much richer than the prototypal model analyzed in Imbens and Angrist (1994) and Heckman and Vytlacil (2005), it is nevertheless possible to provide intuitive arguments for the apparent disconnection between IV estimates and the average effects of education for compliers. There are two distinct causes. First, linear IV estimates are easy to interpret when compliers change their schooling by one year but necessarily when some react by more (or less) than one year.[25] As this is documented in Table 6A, a large share of compliers are actually reacting by two years or more. Compulsory schooling induced 47 percent of compliers to increase their schooling by two years or more while 38 percent of those who reacted to the introduction of a higher education subsidy did so.

A second reason already noted in sub-section 4.4 is that the effects of education are complicated functions of education, work experience (hours worked) and heterogeneity and are in general not orthogonal to the policy exposure indicator. This is true even if endogenous work experience depends only on the instrument through education. This feature is at odds with the independence assumption commonly invoked in the linear correlated random coefficient model of Imbens and Angrist (1994) and Heckman and Vytlacil (2005) and implies that the effect identified by IV is most likely a combination of a wage change caused by variations in schooling induced by a specific policy as well as a change in returns (the average effects) also induced by the same variations in education and experience. Those changes cannot be separated by linear IV.

[25]The analysis of two-stage-least squares in the presence of a discrete (ordered) endogenous treatment variable is analyzed in Angrist and Imbens (1995). They show that under certain conditions, two-stage-least squares estimates may be interpreted as a weighted average of per-unit causal effects. However, when some individuals react by more than one year, IV is more difficult to interpret because individuals affected may not all receive equal weight.

Although, it is not possible to separate precisely the relative responsibilities of each specific cause, it is clear that the structure of our model is not naturally amenable to standard linear IV techniques because policy changes induce simultaneous movements in observed choices and average returns. As a result, the usual IV interpretation tying obtained estimates to average effects for a sub-population of individuals affected by a specific instrument cannot be transported to a dynamic skill accumulation model in which schooling has non-trivial effects beyond labor market entrance. For this reason, a formal statistical discussion of the performance of IV within this specific context is beyond the scope of the paper and it is not possible to say precisely what would linear IV estimate.

To summarize, we find IV estimates of the return to schooling to be uninformative. In five out of six cases considered, they exceed the average effects of compliers by a significant margin. Compulsory schooling estimates are particularly uninterpretable as they are much closer to the average effects of those unaffected and because they exceed IV estimates generated by a reduction in the cost of higher education even if the latter policy affects individual with higher returns than those affected by compulsory schooling.

7 Conclusion

In this paper, we have estimated an early-life cycle model of education, labor supply and earnings. The model identifies four separate reasons that contribute to the existence of steeper age earnings profiles for the more educated. To our knowledge, it is the first to separate the effects of education on entry wages from its causal effect on the returns to work experience. It is also the first to quantify the importance of learning-by-doing induced by the labor supply intensive margin and to evaluate how much of the differences in age-earnings profiles between college and high school graduates are due to a selection effect.

Our model has proven to be a particularly useful tool to comprehend the distinctions between two policy interventions often used in the applied literature on returns to schooling; compulsory schooling and a reduction in the cost of higher education. Our estimates indicate that compulsory schooling would affect the bottom tail of the skill distribution but would be effective at raising human capital because the dynamic effects of education

on hours worked would compensate for the reduction in potential experience. Policies reducing the cost of higher education would also be effective at raising human capital but for different reasons. The returns to schooling of those affected by a cost reduction would be sufficiently high to compensate for the experience loss generated by an increase in college attendance.

Our model has allowed us to answer the following question; would linear IV applied to data generated by a dynamic skill accumulation model estimate the average effects of education for the sub-population affected by any specific policy? The answer is clearly no. Compulsory schooling estimates of the return to schooling are particularly uninformative about the reduced-form effects of education for the sub-population affected as they systematically tend to exceed it by 10 standard errors or more. They are practically uninterpretable as they even fail to capture the identity of those affected.

In light of the sustained interest in income inequality disclosed by both micro and macro economists, and as panel data on schooling and earnings of more recent cohorts become increasingly available, it would be interesting to use our model to reconcile recent changes in the U.S. wage distribution with observed patterns in college attendance, college completion and hours worked.

References

[1] Abel, J.R. and R. Deitz (2014) "Do the Benefits of College Still Outweigh the Costs? *Current Issues in Economics and Finance, NY Federal Reserve Board,* 20 (3)

[2] Adda, J., Dustmann, C., Meghir C. and J.M. Robin (2006) "Career Progression and Formal versus On-the-Job Training", IZA Discussion paper 2260.

[3] Altonji, J., Smith, A. and I. Vidangos (2013) "Modeling Earnings Dynamics" *Econometrica*, 81 (4), 1395-1454.

[4] Angrist, J. and G. W. Imbens (1995) "Two-Stage Least Squares Estimation of Average Causal Effects in Models With Variable Treatment Intensity" *Journal of the American Statistical Association*, 90, 431-442.

[5] Bagger, J., Fontaine, F., Postel-Vinay, F. and J.M. Robin (2014) "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics" *American Economic Review*, 104 (6), 1551-1596.

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

[7] Belzil, C. (2007) "The Return to Schooling in Structural Dynamic Models: A Survey of the Literature" *The European Economic Review,* 51 (5), 1059-1105.

[8] Belzil, C. and J. Hansen (2013) "Characterizing the IV Identifying Moment Condition as Sample Selection Conditions" IZA WP 6339.

[9] Belzil, C. and J. Hansen (2007) "A Structural Analysis of the Correlated Random Coefficient Wage Regression Model" *Journal of Econometrics*, 140 (2), 827-848.

[10] Belzil, C. and J. Hansen (2002) "Unobserved Ability and the Return to Schooling" *Econometrica,* 70 (2), 575-591.

[11] Belzil, C., Hansen, J. and X. Liu (2016) "Estimation of a Life-Cycle Skill Accumulation Model with Comparative Advantages and its Economic Implications", in progress.

[12] Bravo, D., Mukhopadhyay, S. and P. Todd (2010) "Effects of School Reform on Education and Labor Market Performance: Evidence from Chile's Universal Voucher System", *Quantitative Economics 1, 47-95*

[13] Cameron, S. V. and C. Taber (2004) "Estimation of Educational Borrowing Constraints Using Returns To Schooling" *Journal of Political Economy*, 112 (1), 132-182.

[14] Card, D. (1999) "The Causal Effect of Education on Earnings " Handbook of Labor Economics, Volume 3, Edited by O. Ashenfelter and D. Card

[15] Eckstein, Z. and K. Wolpin (1999) "Youth Employment and Academic Performance in High School" *Econometrica,* 67 (6), 1295-1339.

[16] Eckstein, Z. and K. Wolpin (1989) "Dynamic Labour Force Participation of Married Women and Endogenous Wage Growth" *Review of Economic Studies*, 56 (3), 375-390.

[17] Ehrenberg, R.G. (2012) "American Higher Education in Transition" *Journal of Economic Perspectives, 26 (1), 193-216*

[18] Ge, S. (2013) "Estimating the Return to Schooling: Implications from a Dynamic Discrete Choice Model, *Labour Economics* 20: 92-105 (2013) .

[19] Geweke, J. and M. Keane (2000) "Bayesian Inference for Dynamic Discrete Choice Models without the Need for Dynamic Programming" in *Simulation-Based Inference in Econometrics: Methods and Applications* ed. by R.S. Mariano, T. Schuermann and M. Weeks. Cambridge: Cambridge University Press, 100-131.

[20] Gicheva, D. (2013) "Working Long Hours and Early Career Outcomes in the High-End Labor Market" *Journal of Labor Economics* 31 (4), 785-824.

[21] Guvenen, F. (2007) "Learning your Earning: Are Labor Income Socks Really Very Persistent?" *American Economic Review,* 97 (3), 687-712.

[22] Heckman, J. (2010) "Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy" *Journal of Economic Literature,* 48 (2), 356–398.

[23] Heckman, J., Lochner, L. and C. Taber (1998) "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings With Heterogeneous Agents" *Review of Economic Dynamics*, 1 (1), 1-58.

[24] Heckman, J., Lochner, L. and P. Todd (2006) Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond, *Handbook of the Economics of Education, Volume 1.*

[25] Heckman, J. and E. Vytlacil (2005) "Structural Equations, Treatment Effects and Econometric Policy Evaluations" *Econometrica* 73 (3), 669-738.

[26] Heckman, J., Stixrud, J. and S. Urzua (2006) "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior" *Journal of Labor Economics*, 24 (3), 411-482.

[27] Heckman, J., Urzua, S. and E. Vytlacil (2007) "Understanding Instrumental Variables in Models with Essential Heterogeneity" *Review of Economics and Statistics*, 88 (3), 389-432.

[28] Huggett, M., Ventura, G. and A. Yaron (2011) "Sources of Lifetime Inequality" *American Economic Review*, 101 (7), 2923-2954.

[29] Imai, S. and M. Keane (2004) "Intertemporal Labor Supply and Human Capital Accumulation" *International Economic Review,* 45 (2), 601-642.

[30] Imbens, G. and J. Angrist (1994) "Identification and Estimation of Local Average Treatment Effects" *Econometrica,* 62 (4), 467-76.

[31] Johnson, M. (2014) "Borrowing Constraints, College Enrollment, and Delayed Entry" *Journal of Labor Economics,*Volume 31, (4) pp. 669–725

[32] Keane, M. (2010) "Structural vs. atheoretic approaches to Econometrics" *Journal of Econometrics* 156 (1), Pages 3–20

[33] Keane M. and R. Rogerson (2012) "Reconciling Micro and Macro Labor Supply Elasticities: A Structural perspective" *Journal of Economic Literature,* 50 (2), 464-476.

[34] Keane, M. and K. Wolpin (1997) "The Career Decisions of Young Men" *Journal of Political Economy,* 105 (3), 473-522.

[35] Magnac, T., Pistolesi, N. and S. Roux (2013) "Post Schooling Human Capital Investments and the Life Cycle Variance of Earnings" IZA Discussion Paper 7407, Bonn Germany.

[36] Murphy, K.M. and R. Topel (2016) "Human Capital Investment, Inequality and Economic Growth", *NBER Working Paper 21841.*

[37] Rosenzweig, M. and K. Wolpin (2000) "Natural Natural Experiments in Economics" *Journal of Economic Literature*, 38 (December), 827-874.

[38] Rust, J. (1987) "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher" *Econometrica*, 55 (5), 999–1033.

[39] Sauer, R. (2015) "Does it Pay for Women to Volunteer?" IZA Discussion Paper 6784, forthcoming in *International Economic Review*

[40] Sullivan, P. (2010) "A Dynamic Analysis of Educational Attainment, Occupational Choices, and Job Search" *International Economic Review, 51* (1), 289-317.

[41] Taber, C. (2001) "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?" *Review of Economic Studies*, 68 (3), 665—691.

[42] Todd, P. and K. Wolpin (2006) "Assessing the Impact of a School Subsidy Program in Mexico: Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility" *American Economic Review,* 96 (5), 1384-1417.

[43] Topel, R. and M.P. Ward (1992) "Job Mobility and the Careers of Young Men", *Quarterly Journal of Economics*, Vol. 107 (May 1992): 441-79

Table 1

Model fit: Accumulated Choices by Age 30

Age	**School**		**Low Hours**		**Medium Hours**		**High Hours**		**Home**	
	Model	Data	Model	Data	Model	Data	Model	Data	Model	Data
16	9.31	9.39	0.04	0.02	0.02	0.00	0.00	0.00	0.01	0.00
17	10.22	10.31	0.09	0.07	0.04	0.01	0.01	0.01	0.03	0.02
18	11.01	11.15	0.19	0.17	0.09	0.05	0.03	0.01	0.06	0.04
19	11.60	11.80	0.38	0.36	0.22	0.13	0.08	0.05	0.11	0.09
20	11.99	12.22	0.63	0.63	0.42	0.30	0.18	0.11	0.17	0.16
21	12.27	12.55	0.89	0.90	0.67	0.52	0.30	0.20	0.26	0.24
22	12.53	12.84	1.15	1.17	0.95	0.78	0.42	0.33	0.34	0.30
23	12.75	13.05	1.39	1,42	1.27	1.10	0.57	0.50	0.41	0.36
24	12.92	13.18	1.65	1.67	1.60	1.47	0.73	0.69	0.48	0.41
25	13.07	13.28	1.88	1.88	1.95	1.90	0.93	0.92	0.55	0.43
26	13.16	13.35	2.13	2.11	2.35	2.34	1.16	1.16	0.60	0.45
27	13.22	13.41	2.36	2.30	2.74	2.80	1.40	1.42	0.66	0.48
28	13.28	13.46	2.56	2.47	3.16	3.25	1.68	1.71	0.72	0.53
29	13.31	13.51	2.73	2.66	3.61	3.70	1.97	1.99	0.77	0.55
30	13.33	13.55	2.89	2.83	4.09	4.15	2.27	2.29	0.81	0.59

Note: The low hours class corresponds to 1-1,999 hours per year, the medium hours class corresponds to 2,000-2,499 hours per year, and the high hours class corresponds 2,500 hours per year or more.

Table 2A

The Skill Formation Technology: Illustrating Hetergoneity

	Mean (std. dev.)	Type 1	Type 2	Type 3	Type 4
Return to Schooling					
Entry wage	0.102 (0.061)	0.133	0.0001	0.145	0.001
Returns to Experience					
Low hours	0.035 (0.042)	0.008	0.007	0.095	0.00001
Medium hours	0.059 (0.016)	0.056	0.038	0.063	0.078
High hours	0.043 (0.032)	0.018	0.033	0.053	0.102

Note: The average returns to experience are computed at the average predicted level of schooling in the population (13.33). The type probabilities are 0.41 (Type 1), 0.12 (Type 2), 0.33 (Type 3) and 0.14 (Type 4).

Table 2B

The Skill Formation Technology: Illustrating Selectivity

	High School Graduates and Below Mean (std. dev.)	**College Graduates and Above Mean (std. dev.)**
Return to Schooling		
Entry wage	0.094 (0.067)	0.103 (0.055)
Returns to Experience		
Low hours	0.032 (0.036)	0.031 (0.044)
Medium hours	0.057 (0.017)	0.068 (0.015)
High hours	0.035 (0.015)	0.050 (0.029)

Note: The returns are evaluated at predicted schooling levels.

Table 2C

The Effect of Education on Labor Supply at the Intensive Margin

	High Hours	**High Hours**	**Medium Hours**	**Medium Hours**
Schooling	0.028 (0.001)	-	0.017 (0.001)	-
AFQT	-0.015 (0.001)	0.018 (0.001)	-0.005 (0.001)	0.015 (0.001)
Type 2	-0.122 (0.002)	-0.112 (0.002)	-0.326 (0.002)	-0.319 (0.002)
Type 3	0.111 (0.002)	0.072 (0.002)	0.024 (0.002)	0.0005 (0.002)
Type 4	0.028 (0.003)	-0.074 (0.002)	-0.045 (0.003)	-0.107 (0.002)
Constant	-0.180 (0.007)	0.225 (0.001)	0.223 (0.008)	0.466 (0.001)
R^2	0.096	0.072	0.162	0.155

Note: The dependent variables are $\frac{H_{30}}{L_{30}+M_{30}+H_{30}}$ for the first two columns and $\frac{M_{30}}{L_{30}+M_{30}+H_{30}}$ for the last two columns.

Table 3

Wage Growth Regressions

	Specification 1	**Specification 2**	**Specification 3**	**Specification 4**
Schooling	0.003 (0.0001)	0.004 (0.0001)	-	-
High Hours	0.010 (0.001)	-	-	-
Medium Hours	0.032 (0.001)	-	-	-
AFQT	0.002 (0.0002)	0.001 (0.0002)	0.006 (0.0002)	-
Type 2	-0.003 (0.001)	-0.014 (0.001)	-0.013 (0.001)	-0.013 (0.0005)
Type 3	0.030 (0.0004)	0.031 (0.0004)	0.026 (0.0004)	0.026 (0.0004)
Type 4	0.018 (0.001)	0.017 (0.001)	0.002 (0.001)	0.001 (0.0005)
Constant	-0.032 (0.002)	-0.026 (0.002)	0.032 (0.002)	0.031 (0.0002)
R^2	0.089	0.075	0.067	0.056

Note: The dependent variable in each specification is defined as $\frac{w_{i30}-\lambda_i-\lambda_i^S.S_{i30}}{L_{i30}+M_{i30}+H_{i30}}$. Standard errors in parentheses.

Table 4

OLS Wage Regressions using Simulated Data

	Specification 1	**Specification 2**	**Specification 3**
Schooling	0.116 (0.001)	0.094 (0.001)	0.107 (0.001)
Experience	0.076 (0.001)	0.075 (0.001)	0.012 (0.001)
AFQT	-	0.084 (0.002)	0.010 (0.001)
Type 2	-	-	-0.363 (0.005)
Type 3	-	-	0.467 (0.003)
Type 4	-	-	0.273 (0.005)
Constant	0.323 (0.020)	0.637 (0.016)	0.891 (0.020)
R^2	0.140	0.160	0.360

Note: The dependent variable in each specification equals the predicted log wage at age 30. Standard errors in parentheses.

Table 5

Decomposing the Total Effects of Education on Wages by Age

	Partial Effect	**Experience Loss Effect**	**Total Effect**
Population Mean age 25	0.125 (0.0002)	-0.033 (0.00005)	0.091 (0.0002)
Population Mean age 30	0.148 (0.0003)	-0.040 (0.0001)	0.108 (0.0002)
Population Mean age 35	0.174 (0.0003)	-0.040 (0.0001)	0.134 (0.0002)
Population Mean age 40	0.204 (0.0004)	-0.042 (0.0001)	0.162 (0.0003)
Number of observations		119,900	

Note: The partial effect is defined in equation 10 while the experience loss effect is defined in equation 11. The total effect is the sum of the partial effect and the experience loss effects. Standard errors in parentheses, computed using 100 bootstrap replications.

Table 6A

The Distribution of Counterfactual Changes in Schooling

	Change in Years of Schooling				
Policy	0	1	2	3 or more	Total
Compulsory Schooling					
Number of Individuals	98,712	11,270	4,421	5,497	119,900
Proportion	82.3%	9.4%	3.7%	4.6%	100%
Reduction in the Cost of College					
Number of Individuals	94,129	15,923	5,756	4,092	119,900
Proportion	78.5%	13.3%	4.8%	3.4%	100%

Table 6B

Summarising Education Policies

The Identity of Compliers				
	Compulsory Schooling		**Reduction in the Cost of College**	
	Compliers	**Non-Compliers**	**Compliers**	**Non-Compliers**
Structural Parameters				
Mean Wage Intercept	0.884	-0.190	-0.140	0.039
Mean return schooling (entry)	-0.893	0.192	0.111	-0.031
Mean Utility of School	-1.290	0.277	0.494	-0.139
Changes in Human Capital and Hours of Work at age 30				
%△ in Hours Worked	11.3%	-	-3.35%	-
%△ in Schooling	17.9%	-	12.1%	-
△ in Schooling (years)	1.91	-	1.63	-
%Δ in Wage	19.0%	-	16.9%	-
Number of Individuals	21,582	98,318	26,259	93,641

Note: The structural, individual specific parameters above are standardized. Hours worked are measured using the following formula: $\sum_t d_{tl} * 1000 + d_{tm} * 2000 + d_{th} * 3000$.

Table 7

IV Estimates of the Wage Return to Schooling

	Policy					
	Compulsory Schooling			Reduction in the Cost of College		
		Population Estimand			Population Estimand	
Specification	**Estimate (Std. err.)**	**Compliers**	**Non-Compliers**	**Estimate (Std. err.)**	**Compliers**	**Non-Compliers**
IV_I	0.1358 (0.007)	0.0677	0.1395	0.1293 (0.007)	0.1270	0.1269
IV_II	0.1752 (0.007)	0.0943	0.1768	0.1741 (0.007)	0.1603	0.1625
IV_III	0.1646 (0.006)	0.0943	0.1768	0.1612 (0.006)	0.1603	0.1625

Note: In the first specification (IV_I), experience is not modeled and the estimand equals the total effect of education on wages for compliers. In the second specification (IV_II), experience is exogenous and the estimand is the partial effect of education on wages for the compliers. Finally, in specification three (IV_III), experience is endogenous (age is used as instrument) and the estimand is the partial effect of education on wages for the compliers.