IZA

DISCUSSION PAPER SERIES

IZA DP No. 3528

Calibration and IV Estimation of a Wage Outcome Equation in a Dynamic Environment

Christian Belzil
Jörgen Hansen

June 2008

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Calibration and IV Estimation of a Wage Outcome Equation in a Dynamic Environment

Christian Belzil
Ecole Polytechnique, France,
CIRANO and IZA

Jörgen Hansen
Concordia University,
CIRANO, CIREQ and IZA

Discussion Paper No. 3528
June 2008

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 3528
June 2008

ABSTRACT

Calibration and IV Estimation of a Wage Outcome Equation in a Dynamic Environment*

We consider an artificial population of forward looking heterogeneous agents making decisions between schooling, employment, employment with training and household production, according to a behavioral model calibrated to a large set of stylized facts. Some of these agents are subject to policy interventions (a higher education subsidy) that vary according to their generosity. We evaluate the capacity of Instrumental Variable (IV) methods to recover the population Local Average Treatment Effect (LATE) and analyze the economic implications of using a strong instrument within a dynamic economic model. We also examine the performances of two sampling designs that may be used to improve classical linear IV; a Regression-Discontinuity (RD) design and an age-based sampling design targeting early career wages. Finally, we investigate the capacity of IV to estimate alternative "causal" parameters. The failure of classical linear IV is spectacular. IV fails to recover the true LATE, even in the static version of the model. In some cases, the estimates lie outside the support of the population distribution of returns to schooling and are nearly twice as large as the population LATE. The trade-off between the statistical power of the instrument and dynamic self-selection caused by the policy shock implies that access to a "strong instrument" is not necessarily desirable. There appears to be no obvious realistic sampling design that can guarantee IV accuracy. Finally, IV also fails to estimate the reduced-form marginal effect of schooling on wages of those affected by the experiment. Within a dynamic setting, IV is deprived of any "causal" substance.

JEL Classification: B4, C1, C3

Keywords: dynamic discrete choice, dynamic programming, treatment effects, weak instruments, instrumental variable, returns to schooling

Corresponding author:

Christian Belzil
Groupe d'Analyse et de Théorie Economique
CNRS UMR 5824 - University of Lyon 2
93, chemin des Mouilles - B.P.167
69131 Ecully cedex
France
E-mail: belzil@gate.cnrs.fr

* We thank Michael Keane, John Rust, Jim Heckman, Kostas Tatsiramos, Bas van der Klauuw, Pieter Gautier, and Gerard van den Berg.

1 Introduction

We evaluate the accuracy of classical linear IV estimation of an outcome equation, in a context where the endogenous variables originate from a population of highly heterogeneous agents who solve a (partial equilibrium) multi-state dynamic programming model. Some of these agents are subject to policy interventions, which vary according to their intensity. An econometrician, who does not know the data generating process (aside from the functional form of the outcome equation), and who has access to data on choices, outcomes and on a policy shock indicator, estimates the outcome equation by IV. Our task is to evaluate the capacity of IV to recover the relevant treatment effect parameters. As we focus on outcome equations that are affected by multiplicative heterogeneity (a random coefficient specification), the corresponding treatment effect parameter is the Local Average Treatment effect (LATE) introduced in Imbens and Angrist (1994).

In order to link our analysis with the applied microeconometric literature, we choose a well defined empirical setting. Precisely, we analyze the behavior of forward looking agents making time allocation decisions between schooling, work, training and household production over a 33 year period. The key distinction between work and training is at the level of human capital accumulation; each state requires full-time work but training entails a higher degree of skill accumulation, as well as a higher disutility (a higher psychic cost). The underlying model is calibrated to a set of well recognized stylized facts (or common conjectures) about life cycle human capital accumulation and the analysis focuses on point estimation of the return to schooling.[1]

We consider two different versions of a basic intertemporal model. In the dynamic version, accumulated skills affect not only wages, but also the utility (psychic) costs of investing in skill accumulation (given individual heterogene-

[1]We do so because the return to schooling is the most frequently estimated parameter in the microeconometrics literature. See Belzil (2007), for a survey.

ity). In the static version, we remove the dynamic impact of accumulated skills on future choices, but retain the original heterogeneity structure that causes a spurious correlation between current and future choices. In words, the static model is characterized by spurious dynamics.

In our analysis, the policy shock takes the form of a higher education subsidy. Precisely, we simulate the effects of two different levels of the subsidy. By letting the intensity of the policy vary, we can therefore investigate the relationship between the accuracy of IV (defined as the difference between the estimate obtained from a large sample and the population treatment effect parameter) and the statistical power of the instrument. As we know the true data generating process, we simulate a history of sequential decisions for each individual, and dissect the correlation between the policy shock indicator and the population composite error term, which also depends on the power of the instrument. We provide an economic interpretation of the notion of an instrument's statistical power.

Subsequently, we examine the performances of two different sampling designs; a Regression-Discontinuity (RD) design as well as an age-based sampling design which targets early career wages that may be used to perform (and improve) IV. We discuss the relative advantages and disadvantages of each design.

Finally, we investigate the capacity of IV to estimate alternative population "causal" parameters, such as the reduced-form marginal effect of education on wages for those who have been affected by the policy shock.

It is important to understand that some aspects of the results reported in the paper are general in nature, while others pertaining to the population of agents analyzed in the paper, are model specific. Throughout the paper, we focus on the general interpretation of the results.

Overall, the failure classical linear IV is quite clear. Because IV confounds post-schooling wage growth with skill acquisition while in school, IV fails to recover the true LATE, even in the static version of the model. In some cases, IV estimates lie outside the support of the population distribution of returns to schooling, and are almost twice as large as the population LATE. Because of the trade-off between the statistical power of the instrument and dynamic self-selection caused by the policy shock, IV performance is practically uncorrelated with the power of the experiment. Within a dynamic environment, access to a "strong instrument" is not necessarily desirable.

In an intertemporal setting such as the one analyzed in the paper, a sampling design can be characterized in two dimensions; namely a cross-sectional dimension (which types of individuals are selected), and a time/age dimension (when are individuals sampled). We show that IV estimates are not only sensitive to cross-sectional composition, but are also highly sensitive to the timing of outcome measurement. In general, it is impossible to dissociate IV from a particular timing of wage measurement (for a fixed population). Yet, the LATE parameter is itself invariant to the timing of wage measurement by definition. There appears to be no obvious realistic sampling design that can guarantee IV accuracy.

Given that IV fails to estimate the population LATE parameter, it is interesting to investigate what treatment effect IV does estimate. Because IV disregards post-schooling choices, a conjecture is that it may estimate some reduced-form marginal effect of schooling on wages. As empirical labor economists often use the term "causal" parameter when referring to those parameters that are associated to a subset of the population affected by some experiment, it is natural to compare IV with the marginal effect of education for those who have been affected by the policy shock. However, our results indicate that IV also fails to estimate these reduced-form marginal effects. Interestingly, IV estimates are typically much closer to OLS estimates of the effect of education computed on the control group, or computed on a population that contains both the control group and the treatment group, than they are to OLS estimates computed on the sub-population of individuals affected by the experiment. Within a dynamic setting, IV appears to be deprived of any "causal" substance.

The paper is structured as follows. In Section 2, we present some background material as well as a review of the literature. The population behavioral model is outlined in section 3. We discuss model calibration in Section 4. Section 5 is devoted to the experimental design. In Section 6, we discuss IV identification within a dynamic model. We discuss the implication of using strong vs weak instruments in Section 7. In Section 8, we discuss the empirical results that followed the implementation of classical IV. Section 9 is devoted to the implementation of alternative sampling designs. In Section 10, we investigate other potential population parameters that IV may estimate. Section 11 concludes.

2 Background Material and Related Literature

When estimating outcome equations plagued with endogenous variables, micro econometricians typically choose between two fundamental estimation methods; Instrumental variables (IV) estimation or structural estimation. The fundamental paradigm of IV estimation is reliance on a variable (usually a policy shock) that is assumed to be correlated with an endogenous variable of interest but uncorrelated with the error term of the outcome equation so to obtain independent variation. In Microeconometrics, this policy shock is sometimes referred to as a "Natural Experiment" and is usually labeled as an "exogenous variable". Because such events are relatively uncommon, empirical studies always rely on a single experiment, and therefore use only one instrument.

In the ideal context where i) the outcome equation is linear in variables, ii) the error term of the outcome equation is additive, (iii) the instrument affects individual choices in a same (unique) direction and iv) the time elapsed between the realization of the instrument and the realization of the error term of the outcome equation precludes any form of intertemporal substitution (i.e. orthogonality conditions are met), the desirability of IV is well established.[2]

[2]See Heckman and Vytlacil (2005) and Heckman, Vytlacil and Urzua (2006).

However, in virtually all micro econometric applications, the parameter of interest is estimated from data on outcomes measured much after the effect of the instrument (policy change) has set in. If the true data generating process i) contains a wide (possibly unknown) number of endogenous variables, or (ii) is characterized by a high-memory law of motion, IV estimates are difficult to interpret.[3]

While differences between structural and IV approaches have been at the forefront of microeconometric theory in recent years (Heckman and Vytlacil, 2005), the optimal use of policy instruments, within a dynamic context, has virtually been evacuated by microeconometricians. Most of the debate between advocates of IV and structural approaches is about concepts that relate to the treatment of heterogeneity and, more precisely, the role of monotonicity (or the degree of separability) in the first stage model. However, both approaches are based on a-priori moment (orthogonality) conditions which are virtually always argued upon in a more or less static background. Economic analysis of those orthogonality conditions required to identify the model is practically never performed.[4]

This paper belongs to a new, but growing, stream of the microeconometrics literature that aims at providing an econometric interpretation of several IV estimates reported in the empirical literature. Rosenzweig and Wolpin (2000) is the seminal piece in this branch of the literature. They present a survey of the economic literature using natural experiments (mostly in labor economics and in development economics) and present an economic analysis of the implicit assumptions made in the IV literature. Keane (2007) discusses the notion of identification and points out the need for a theoretical model in order to interpret

[3]See Keane (2007).

[4]van den Berg (2007) investigates a dynamic model in which it is optimal for the agent to acquire the value of the intended instrumental variable. This provides a foundation of exclusion restrictions in terms of economic behavior which can be used to describe the policy evaluation settings in which instrumental variables are likely or unlikely to make sense.

IV estimates obtained in a dynamic environment.[5]

Todd and Wolpin (2006) estimate a dynamic model of schooling and fertility, using a social experiment that took place in Mexico. Although the authors are not primarily concerned with the estimation of an outcome equation, they show that a treatment group may be used to validate a structural model estimated on a control group. Taken as such, the results reported in Todd and Wolpin (2006) imply that the instrument may be disregarded by the econometrician at the inference level, but should be used as a way to provide out-of-sample fit.

While all of these papers discuss methodology to some extent, none of them can provide conclusive statements about IV accuracy (or lack thereof). As of now, it is fair to say that applied micro-econometricians (especially in Labor Economics) prefer the IV approach (Keane, 2007). As empirical Labor Economics is largely based one a "one endogenous (choice) variable/one endogenous state (outcome) variable/one instrument" paradigm, the preference for IV is often rationalized by the "minimal amount of (parametric) assumptions" required to apply IV.[6]

These pessimistic remarks are the point of departure of this research. Precisely, we start from the principle that in order to evaluate the performance of IV in a dynamic context, three conditions must be met. First, our analysis requires to know the true underlying data generating process. Second, the data generating process must be realistic, and therefore must share a large number

[5]Indeed, Keane (2007) refers to the IV literature as the "Atheoretical approach to Econometrics". In his comment on Keane (2007), Rust (2007) presents some provocative toughts on the relative impopularity of the structural approach in micro-econometrics.

[6]The fascination for static models in microeconometrics (at least in labor economics) has no pendant in macroeconomics, which has relied on dynamic representations of the economy for several decades. However, advocates of dynamic general equilibrium models of the business cycle have recently questioned the validity of Structural Vector Autoregressive Regressions (SVAR) models as an alternative to structural modeling. See Chari, Kehoe and McGrattan (2007).

of characteristics with observational data used in the applied literature. For instance, virtually all parameters should be individual specific, and the time horizon should be long enough to allow us to generate cross-section data representing a relatively wide (and realistic) age dispersion.[7] Finally, to reproduce the conditions that are usually faced by applied econometricians, IV estimates should be relatively imprecise.

For these reasons, the first step of the analysis (the choice of the population parameters used to generate artificial data) can hardly be achieved realistically by estimation. Instead, we rely on informal calibration.[8] The details are found in subsequent sections. The behavioral model is presented and discussed in the next section.

3 The Behavioral Model

Before setting up the model, we discuss some requirements that must be imposed on the theoretical structure.

First, we disregard modeling a labor market with multiple sectors. We do so because the empirical IV literature is based on the availability of a single instrument (one natural experiment). Even in absence of any post-schooling endogenous variables, the econometrician using IV would not be able to estimate

[7] In a companion paper, we compare estimates of an outcome equation obtained from a standard linear IV model with those obtained from a benchmark structural dynamic model of choices and outcomes (a dynamic multinomial logit model of choices, with normal wages), estimated by simulated maximum likelihood. The structure resembles the popular empirical IO model of Berry, Levinhson and Pakes (1994).

[8] Treating a model structure as a known DGP and performing IV out of simulated data is not new. This is done, among others, in Imai and Keane (2004), within an intertremporal labor supply model, and in Belzil and Hansen (2005), within a dynamic model of schooling attainments.

a sector/occupation specific return to schooling.[9]

Second, the model should rule out general equilibrium effects. If not, the wage distribution would not be invariant to policy change, and the monotonicity property (the requirement that individuals react in a same direction) may be lost. IV would then fail by construction.[10]

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

In total, we selected the following 8 attributes:

1. Schooling activities must be located mostly at the beginning of the time horizon. Individuals should rarely return to school, after having worked in the market.

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

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

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

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

6. Household production must be a relatively rare event

[9]See Keane and Wolpin, 1997.

[10]This issue is discussed in Heckman, Lochner and Taber (1998). They investigate tuition policies within a calibrated lifecycle general equilibrium model of human capital formation.

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

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

8. The average return to schooling of those affected by the education subsidy should exceed the population average.

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

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

The seventh and eighth characteristics may be somewhat more controversial. The seventh one is sometimes referred to as the classical "ability bias" hypothesis. It implies that the correlation between wages and schooling is an over-estimate of the true effect of schooling on productivity. We choose a positive (as opposed to a negative) bias because of its intuitive aspect. It would obviously be possible to define the model structure differently, or to modify the dynamic structure, so to imply a negative OLS bias. This would have no implication for our analysis.

Finally, the eight one may also be controversial because it is far from a stylized fact. Indeed, it is a pure conjecture. However, it is the most common interpretation that empirical labor economists offer for the incidence of high IV estimates of the return to schooling, which are usually reported in the literature. For this reason, we build our model under this specific assumption.[12]

[12]There may be other features of life cycle wages that may be occasionnally cited, but as pointed out recently in Heckman, Lochner and Todd (2006) and Belzil (2008, forthcoming), many widely accepted features of the standard Mincer wage equations are rejected when tested formally.

3.1 Model Structure

The baseline model is a stochastic dynamic discrete choice model of labor supply/human capital accumulation over the life-cycle. There are 33 periods to allocate between the 4 mutually exclusive states. The states are Schooling (s), work with a low rate of skill accumulation (e), work with a high rate of skill accumulation (a), and Household Production (h). The corresponding capital letters (S_t, E_t, A_t, H_t) are used to measure the number of periods accumulated in each state. There is a maximum of 11 years of schooling attainable. In observational data, the pendant of state e could be full time employment, while the pendant of state (a) could be work, with on-the-job training. The distinction between Full-time employment (e) and Work and Training (a) is therefore in the intensity of human capital accumulation (a is the high intensity mode). We assume that the utility of school changes with grade level and we consider 3 distinct levels; 1 to 4, 5 to 8, and 9 to 11.[13]

Individuals are risk neutral and maximize the expected value of lifetime net earnings, over the entire life-cycle. The state-specific utilities are defined below.

3.2 School

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

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

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

[13] We interpret the second level (5 to 8) as the pendant of college education.

of interruption is $t - S(t)$). The term α_i^S represents individual heterogeneity in taste for schooling (academic ability). Finally, ε_{it}^S is a stochastic i.i.d. shock.

3.3 Household Production

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

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

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

3.4 Employment and Training

The utility of **work without training**, U_{it}^E, and the utility of work with training, U_{it}^A, are constructed as the difference between the wage rate and the monetary costs of occupying a specific state. Precisely, U_{it}^E and U_{it}^A, and their related **costs**, $C_{it}^e()$ and $C_{it}^a()$, are given by the following equations;

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

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

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

3.5 Market Productivity

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

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

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

3.6 The Bellman Equations

The choices are summarized in the binary indicators, d_{tk}. Precisely, $d_{tk} = 1$ when option k (w, h, w, a) is chosen. Given the Markovian structure of the model, the solution to the problem is obtained using recursive methods, and optimal choices may be characterized by a Bellman equation (Bellman, 1957).

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

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

or, more compactly, as

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

where β is the discount factor, and where Ω_t is the set containing all state variables known by the agent at t. The law of motion maps current choices (d_{kt}) and current state variables (Ω_t) onto future state variables (Ω_{t+1}).

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

3.7 The Distribution of Individual Heterogeneity and Random Shocks

- The full heterogeneity vector, $\nu_i = \{\alpha_i^S, \alpha_i^H, \lambda_i, \delta_i, \theta_i, c_{0i}^a, \beta_i\}$ is distributed according to a multi-variate discrete distribution with 20 vectors of support points;[15]

$$\nu_k \sim \{\alpha_k^S, \alpha_k^H, \lambda_k, \delta_k, \theta_k, c_{0k}^a, \beta_k; p_k\} \text{ for } k = 1, 2, ..20 \tag{8}$$

  where p_k is the population proportion of type k.

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

3.8 Model Solution

As is relatively common in the literature, we solve the Bellman equations using simulated realizations of the random shocks. The Bellman equations need to be solved for each single type separately. Our solution method is exact in the sense that we do not use any approximation or interpolation methods. More details are found in Section 4 and Section 8.

4 Calibration of the Model

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

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

As a starting point, we choose hourly wages as the benchmark utility. To choose the preference parameters, we relied mostly on the structural literature, in order to obtain a realistic range of the relevant parameters (when possible). Then, we simulated the model and adjusted the parameters until the final values enabled us to match the population characteristics or the population moments that we stated as desirable.[16]

4.1 Outcome Equation and skills

The distribution of returns to schooling is centered at 0.06 (a value close to estimates reported in the structural literature). However, we allow for a high degree of dispersion (as reported in the IV/LATE literature). The support of the distribution of returns ranges between 0.00 and 0.12 (see Table A1). These numbers therefore reflect estimates reported in both the structural and the IV literature.[17]

The average returns to work experience (0.01) and to work with training (0.03) are chosen to reflect the fact that human capital accumulation is more intensive in state a than in state e. We treat the utility of school, the cost of on the job training and the return to education as driven by an academic skill, and enforce a perfect correlation between these components. However, to deviate from a trivial ability bias structure, we assume that both the wage intercept and the return to work experience may be driven by skills that may be non academic, and enforce a weak correlation between these two components, and the other academic heterogeneity components (the utility of school, the cost of on the job

[16] We did not proceed with a formal calibration procedure, in which a set of precise moment conditions are imposed, because the population characteristics that we target are more qualitative than quantitative. For instance, the positive ability bias and the concavity of the wage profile would need to be represented as inequality conditions.

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

training and the return to education).[18]

4.2 Post Schooling Dynamics

The parameters $\{c_{1a}, c_{2a}, c_{3a}, c_{4a}\}$ and $\{c_{1e}, c_{2e}, c_{3e}, c_{4e}\}$ are capturing the effect of accumulated schooling, home time, employment and training on the cost (or disutility) of work/training (a) and work (e). The vectors are equal to {-0.30, 0.00, 0.00, -0.05) for state a, and {-0.20, 0.00, -0.05, -0.05) for state e. The parameter values for c_{1a} (-0.30) and c_{1e} (-0.20) imply that accumulated schooling reduces both the cost of investing in human capital and the cost of labor market work. The larger effect of education on the cost of training is a reflection of the academic nature of the work/training activity. The non-negative values for c_{3e} and c_{4e} allow us to introduce some dynamics in the decision to work. The null values for c_{2a} and c_{2e} imply the absence of skill depreciation.

4.3 Preference Heterogeneity and Discount Rates

To reflect preference heterogeneity, we allow discount rates to differ across individuals. They range between 0.00 and 0.10. The average discount rate (0.05) is standard. This form of preference heterogeneity may also be re-interpreted as a way to approximate the effects of liquidity constraints. It is important to do so, because in the IV literature, the high IV estimates are often conjectured to arise because they reflect the LATE parameters of a subpopulation of individuals affected by liquidity constraints, or of a subpopulation of individuals who have high discount rates. As economics offers no guidance for the choice of a correlation between discount rates and individual skills, we started the calibration procedure by imposing quasi orthogonality between discount factors and

[18] In the single skill model of Belzil and Hansen (2002), the correlation is above 0.9. In a multiple skill model (such as in Keane and Wolpin, 1997), the correlation between the utility of attending school and white collar skills would also be very high.

other heterogeneity components, and adjusted the correlations in order to match population characteristics. In other words, and as opposed to the correlations between various market skills and the costs of training and schooling, we regarded the correlations between discount rates and other heterogeneity components as secondary. Indeed, they do not play a key role in our analysis.

4.4 Heterogeneity vs Ex Ante risk

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

4.5 The Control Groups

In order to proceed further, we build one control group for the dynamic model, and one for the static model. The dynamic version of the model is identified by the parameter values that were discussed in 4.1, 4.2, 4.3, and 4.4 , and reported in Appendix (Table A1). In the dynamic model, the correlation between current and future choices is driven by persistent unobserved heterogeneity, as well as a causal effect of current choices on the cost of choosing future actions. The static model is a restricted version of the dynamic model. It is obtained by setting the c_{1j}, c_{2j}, c_{3j} and c_{4j} to 0 for $j = a$ and e. In the static model, the correlation between current and future choices is explained solely by persistent individual heterogeneity. In other words, there is spurious dynamics.

In order to construct population data for the control group, each type of individuals is duplicated in 250 different realizations of the random vector (for a total of 5,000 units). For each model, we simulate 33 years of data on wage outcomes and choices for a total 5000 individuals.

We provide a summary of individual choices for each model in Table 1A (Dynamic model) and Table 1B (Static model). To do so, we compute the number of accumulated periods in each state for each model . The frequencies display the desired features that we advocated in Section 4. For instance, schooling is chosen mostly in the first 10 years. Average schooling attainments are higher in the dynamic version (5 years) of the model than in the static one (3.8 years).

Home production is rarely chosen (it accounts for approximately 10% of total time allocation in the dynamic model). However, it is interesting to note that the incidence of inactivity increases in the static version of the model. The average number of periods spent at home, equal to 4.4 in the dynamic model, increases to 8.7 in the static version. This is explained by the larger opportunity cost of being involved in non-productive states, which characterizes the dynamic model. Finally, the incidence of work and training is found to decrease as one approaches the end of the life cycle, and therefore illustrates a decreasing rate of skill accumulation.

Although the decreasing incidence of training is established in Table 1A and Table 1B, we also performed OLS regressions of log wages on education, and potential experience, in order to double check concavity. As the procedure used to simulate data is a central element of our analysis, it will be presented in details in Section 8. For the moment, it is sufficient to point out that regressions are computed on a cross section of 5,000 observations. In other words, we used a single wage per individual. We do this simply because most empirical studies reporting OLS or IV use cross-section data, even when panel data are available. To do so, we select a period between period 5 and period 33 using a random number generator. In Table 2A, we present summary statistics that characterize the wage distribution in the control group for both the dynamic and static models. The results of these OLS regressions are reported in Table 2B. OLS estimates are in the neighborhood of 0.12, which is comparable to OLS estimates obtained from various cross-sections of the NLSY (in the US) and are therefore higher than the population average return. The average growth rate is between 1% and 1.5% per year of potential market experience (an estimate also close to numbers found for the US).[19]

5 Characterizing the Experiment

The higher education subsidy experiment consists of giving a transfer payment to those attending grade levels 5,6,7 and 8. This boils down to a reduction in the monetary costs of attending those specific grade levels.

5.1 The Statistical Power of the Experiment

The recent literature on IV estimation has pointed out statistical issues that pertain to the use of instruments that are weakly correlated with the endogenous variable of interest.[20] To introduce the statistical power of the experiment in our analysis, we simulate the effects of various levels of education subsidies. We impose a reduction of the cost parameter (α_1^S) of the order 1 dollar (low level), and 3 dollars (high level). These numbers generate F statistics that vary between 3 (low intensity static model) and 58 (high-intensity dynamic model).[21] The related standard errors will be discussed later.

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

[20] See Staiger and Stock (1995).

[21] For instance, a first stage regression F statistic around 10 is sometimes viewed as ideal by practionners.

At the outset, it should be clear that choosing a subsidy that is realized at a grade level (level 5), which is around the average schooling attainment for the control group, puts IV in almost ideal conditions. For instance, if we implemented policy interventions that are realized either in the neighborhood of the minimum or the maximum schooling attainment (say a mandatory schooling attainment, or a subsidy conditional on attending grade level 9,10 and 11), IV would suffer a larger risk of lying outside the support of the returns to schooling distribution. This is simply because the average return of those affected by some "extremist" policy, would also be more likely to lie in the neighborhood of the extreme values of the support of the distribution.

In Table 3, we summarize the main effects of the policy interventions. At this stage, it is important to understand that these quantities are meant to summarize the counterfactual effects of implementing a new policy. For this reason, they are computed using the same realizations of the random shock vector that were used to generate the control group. In other words, we evaluate the counterfactual effects of the new policy by computing individual decisions of the control group under the old regime, as well as under the new regime, while holding individual random shock histories constant. This guarantees that the Monotonicity condition will hold.

In order to illustrate the implications of moving from a lower to a higher subsidy, we report a F statistic computed from a regression of schooling on the instrument, as well as the density of the sub-population (the population proportion) affected by the experiment.

In the dynamic model, the experiment increases average years of schooling by 0.25 in the low subsidy regime, and 0.5 year in the high subsidy regime. As a consequence, the density of the population affected, as well as the F statistic, both increase as we move toward a more generous policy. The F statistics of the dynamic model are equal to 12.3 (low intensity) and 57.9 (high intensity). The fraction of the population affected by the experiment goes from 0.18 to 0.32.

Overall, the experiment is weaker in a static environment, as the F statistics range between 3.3 and 45.4. This is easily explained. In absence of a causal effect of education on the costs of acquiring future skills, the incentive to get more educated is smaller. For this reason, the changes in schooling (0.13 and 0.50) and the proportions of individuals affected (0.09 and 0.26) are smaller.

5.2 The Population Local Average Treatment Effects

We now turn to the population Local Average Treatment Effect (LATE). We use the standard definition of the LATE parameter and evaluate it as the average return to schooling for those who are affected by the experiment (those for whom post-experiment schooling is not equal to pre-experiment schooling).[22] In our analysis, an individual is defined as a type and a particular random shocks history. Because of its intrinsic counterfactual nature, computing the LATE requires to hold the vector of realized random shocks constant for each individual.[23]

Within a dynamic setting, the LATE associated to a specific policy change depends on the type of intervention, on its intensity, and on the underlying model.

First, the subsidy affects the behavior of those who would obtain a lower level of schooling ex-ante through the future component of the utility of attending school.

At the same time, the subsidy increases the continuation probabilities of those who would have reached higher education even in absence of the subsidy. That is, for a given random shock, a higher subsidy increases the probability of continuing

[22]Imbens and Angrist, 1994, introduced the notion of LATE in an IV context. Björklund and Moffitt (1987) and Heckman and Vytlacil (2005) introduced the notion of marginal treatment effect, which generalizes the LATE parameter. Finally, Belzil and Hansen (2007) estimate LATE parameters within a stochastic dynamic programming model.

[23]Because the LATE parameters are computed from 5,000 realizations of the vector of random shocks, their sampling variability turn out to be very small. As a consequence, we treat it as a population parameters.

further for those already attending higher education.

Finally, in a multi-variate heterogeneity framework, individual choices (and therefore individual reaction to policy changes) are a non-trivial function of all individual endowments (such as discount rates, skills and tastes). For all these reasons, the effect of a change in policy intensity (say a movement from a low intensity to a high intensity policy) is difficult to predict. It will depend on the difference between the two different sets of individuals affected by the policy.

For each subsidy intervention, we have computed the Local Average Treatment Effects (LATE). These are also summarized in Table 3.

As is evident from the numbers provided in Table 3, the LATE parameter does practically not vary with the intensity of the experiment, despite the relative differences in the population density of individuals affected. This is true both in the dynamic model (0.0710 and 0.0714) and the static model (0.0789 and 0.0781).

As we will see later, this is an important feature of the model that we have calibrated. It indicates that, in our specific model, individual returns to schooling are not the only determinant of individual reactions to policy change. In particular, individual comparative advantages in labor market work, in training, as well as differences in discount rates may also be relevant.

6 IV Identifying Conditions in a Dynamic Environment

In order to comprehend what **IV is** in a dynamic setting, it is useful to re-express the Mincer equation as follows:

$$w_{it} = \lambda_i \cdot (\sum_{j=1}^{t} d_{k=S,j}) + \varphi_{it}(.) + \alpha^W + \varepsilon_{it}^W \tag{9}$$

where

$$\varphi_{it}(.) = \delta_i \cdot (\sum_{j=1}^{t} d_{k=E,j}) + \theta_i \cdot (\sum_{j=1}^{t} d_{k=A,j})$$

The econometrician who estimates the wage equation using a natural experiment uses a policy shock indicator, denoted Z_i, as a source of identifying condition. As IV, which is defined as $(Z'S)^{-1}Z'w$, naturally arises in a classical linear regression framework with an additive error term, it is informative to consider the economic implications of the IV identifying conditions in a dynamic environment.

To achieve identification, the econometrician must disregard all post schooling choices and collapse them in a composite error term.

Using a cross-sectional notation, one starts from

$$w_i = \lambda_0 + \bar{\lambda} \cdot S_i + \varepsilon^*_{it} \tag{10}$$

where λ_0 is an intercept term, $\bar{\lambda}$ is the population average, and where ε^*_{it} (the composite error term) is equal to

$$\varepsilon^*_{it} = \varphi_{it}(.) + \omega_{\lambda i} \cdot S_i + \varepsilon^W_{it} \tag{11}$$

with

$$\omega_{\kappa i} = \lambda_i - \bar{\lambda}$$

The error term ε^*_{it} is composed of three distinct elements.

First, ε^W_{it} is a purely stochastic innovation. It plays no role in our analysis.

Second, the term $(\omega_{\kappa i} \cdot S_i)$ is the classical representation of the error term in a correlated random coefficient wage regression model.[24]

Finally, the term $\varphi(.)$ collapses the effects of all post-schooling choices made until date t.[25] It depends on the actual sequence of individual choices, on the heterogeneity distribution, and indirectly, on both schooling and the policy shock

[24]This term is the central piece in the analysis of Imbens and Angrist (1994) and Heckman and Vytlacil (2005).

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

indicator (Z). Obviously, it also depends on age (or calendar time). As a consequence, the correlation between $\varphi_{it}(.)$ and the policy shock also changes as individuals evolve over time, and so does the correlation between ε^*_{it} and the policy shock . So, from now on, it will be convenient to think about the composite error term as $\varepsilon^*_{it}(Z)$.

7 What does a Strong Instrument Mean in a Dynamic Model?

In the weak instrument literature, a high correlation between the instrument and the endogenous variable is viewed as desirable for two main reasons. First, a strong instrument increases the precision of IV.

A second reason has to do with the asymptotic bias. In a model with additive heterogeneity, the IV asymptotic bias is equal to $(plim\frac{Z'S(Z)}{N})^{-1}plim(\frac{Z'\varepsilon^*(Z)}{N})$. That is, for a given (fixed) level of correlation between the instrument and the error term, a strong instrument reduces the IV bias. Obviously, when the error term is viewed as an individual (fixed) heterogeneity endowment, it is possible to think about variations in the strength of an instrument (variations in the intensity of the policy shock), for a fixed level of correlation between the instrument and the error term. However, this argument is inherently static. We now pay attention to the desirability of a strong instrument, within a dynamic setting.

As most applied econometricians interpret IV in a framework where heterogeneity is multiplicative (and correlated with the regressor), the degree of inaccuracy should measure the distance between IV and its estimand (the Local Average Treatment Effect). So, from now on, in order to avoid confusion, we reserve the term "inaccuracy" for the IV-LATE difference $(\hat{\lambda}_{iV}-\lambda_L)$, while the term "asymptotic bias" is reserved for the IV-population average difference $(\lambda_{IV}-\bar{\lambda})$.

In a dynamic setting, the weak/strong instrument distinction becomes more complicated. To see the argument, it is sufficient to recognize the dependence of both $S(.)$ and $\varepsilon^*(.)$ on Z .

Re-write the IV-LATE difference as

$$plim(\lambda_{IV}(Z) - \lambda_L(Z)) = (\bar{\lambda} - \lambda_L(Z)) + (plim\frac{Z'S(Z)}{N})^{-1}plim(\frac{Z'\varepsilon^*(Z)}{N}) \quad (12)$$

Within a class of policy experiment (for instance, a policy that implements a higher education subsidy), a variation in the intensity of the incentive has several implications.

First, an increase in intensity changes the correlation between schooling and the policy shock (the first stage regression). At the same time, this change in intensity may induce a change in the population local average treatment effect, $\lambda_L(Z)$.

Second, and independently from the potential change in the LATE, an increase in policy intensity also changes the correlation between the instrument and the error term, since individual post-schooling choices are also affected by the variation in intensity.

Third, even in the sole presence of persistent heterogeneity in the utility of post-schooling choices (in the static version of the model with spurious dynamics), a change in the intensity of Z will automatically change the correlation between Z and the error term. This is because the change in schooling observed for a subset of the population will translate into a change in the correlation between Z and $\varphi_{it}(.)$, which is explained by the fact that these individuals have different endowments (returns to schooling, returns to training and returns to work experience). It is important to understand that this may happen even if differences in individual returns to schooling between those affected and those who are not, are very small.

As a consequence, having access to a strong instrument may not always be a blessing. There is no guarantee that increasing the power of the instrument will increase IV accuracy.

For example, consider an hypothetical case where a movement from Z_{low} to Z_{high} denotes an increase in intensity, and assume, without loss of generality, that IV over-estimates the LATE in both cases. An increase in IV accuracy (i.e. $plim(\lambda_{IV}(Z_{low}) - \lambda_L(Z_{low})) - plim(\lambda_{IV}(Z_{high}) - \lambda_L(Z_{high})) > 0$) requires that

$$\begin{aligned} & (plim\frac{Z'_{high}S(Z_{high})}{N})^{-1}plim(\frac{Z'_{high}\varepsilon^*(Z_{high})}{N}) - \\ & (plim\frac{Z'_{low}S(Z_{low})}{N})^{-1}plim(\frac{Z'_{low}\varepsilon^*(Z_{low})}{N}) \\ < \; & \lambda_L(Z_{high}) - \lambda_L(Z_{low}) \end{aligned} \tag{13}$$

In words, an increase in accuracy associated to a stronger instrument, will always require some adjustment condition governing the co-movements between the true LATE parameters, the first stage regression, and the correlations between the instrument and the population error term.

As an illustration, suppose an example where (i) $\lambda_L(Z_{high}) - \lambda_L(Z_{low})$ is close to 0 (which is actually the case in our model), and where (ii) the correlation between the error term and the instrument is positive under both the high-intensity and the low-intensity regime, the implementation of the high intensity policy must imply either a decrease in the instrument-error term correlation, or an increase limited by the proportional increase in the correlation between schooling and the instrument. That is

$$plim(\frac{Z'_{high}\varepsilon^*(Z_{high})}{N} < plim(\frac{Z'_{low}\varepsilon^*(Z_{low})}{N}) \cdot c$$

where

$$c = \frac{plim\frac{Z'_{high}S(Z_{high})}{N}}{plim\frac{Z'_{low}S(Z_{low})}{N}} > 1$$

Obviously, these inequalities would have to be adjusted for cases where IV under-estimates its target, or for cases where increasing policy intensity would reduce the correlation between the instrument and the endogenous variable of interest. However, the main conclusions would remain the same. Access to a stronger instrument would only increase IV accuracy under some specific conditions.

As is going to be illustrated below in the context of our dynamic skill accumulation model, those specific conditions are not imposed by IV estimation. As noted earlier, IV has been developed in a classical linear regression framework with an additive error term. The moment conditions that define IV are those that characterize the orthogonality of the policy shock with respect to the econometrician's error term. This implies that, for a given model structure, and given a precise target (the LATE parameter), there may exist an optimal degree of policy intensity associated to a specific class of policy intervention.[26]

We can now proceed with a formal implementation of different IV strategies on our calibrated model.

8 Computing Classical IV

We now turn to the implementation of classical IV estimates.

8.1 Creating a Cross-Sectional Data Set

To mimic empirical analyses reported in the IV literature, we must achieve two things. First, we must append a treatment group to the control group which has already been defined. However, as it is the case in observational data, the treatment and control groups should be representative of the same population (display the same distribution of types) but should still be composed of different individuals. Second, we must construct a sub-population of individuals who work

[26]This point is obviously different from the notion of optimal instruments that relates to efficiency arguments. For instance, in the example analyzed in this paper, there may be an optimal amount of higher education subsidy that would lead to the most accurate IV estimate of the LATE parameter (minimum bias). However, the theoretical existence of such an optimal instrument strength is a conjecture which we do not investigate any further.

(at least once) in the labor force.

In order to build the treatment group, we proceed as we did for the control group. Again, each type of individuals is duplicated in 250 different realizations of the random vector (for a total of 5,000 units). We simulate 33 years of choices and wage outcomes under the new policy, as described in the previous section. As we did for OLS regressions of wage outcomes on the control group, we select one wage per-individual over the entire life-cycle. More precisely, individual wages are selected randomly between period 5 and period 33, because most actual cross-section data sets contain wages that are realized over the entire life cycle, but do not include very young workers.[27] To do so, we use a uniform random number generator. We end up with 10000 observations (5000 in control and 5000 in treatment).

In the second step, we use simulated choices, and construct a sub-population of individuals who are either in state e or state a at the actual period randomly selected. These conditions must be met when IV is actually implemented on observational cross-section data.

It is important to understand that sampling the Labor Force entails two different types of selection; namely a classical cross-sectional dimension (what type of individuals do we sample) and a time dimension (when are individual wages measured).[28]

Before going any further, two issues have to be clarified. First, and because applying IV to different sub-populations (different sampling regimes) may imply changing the true LATE parameter, a logical estimation strategy requires to select the particular LATE in which the econometrician is interested. For the sake of the presentation, we will first assume that the econometrician is interested in the

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

[28]An entry wage is defined as a wage offer for an individual who has, until that time, accumulated no work experience and no on-the-job training.

population LATE (the control group). Indeed, in the current model, it turns out that the LATE parameter is relatively inelastic with respect to cross-sectional sampling. This facilitates the task of the econometrician.[29] This very small elasticity of the LATE parameter is a feature of any life-cycle model of human capital accumulation in which forward looking agents take into account that the largest share of the life-cycle is spent in the labor market. In other words, rational individuals may take their decisions primarily on their post-schooling comparative advantages.

Second, because IV estimates are theoretically inconsistent, their sampling variability cannot be evaluated by the usual formulas. Instead, we evaluate it using bootstrap methods.

8.2 IV Estimates Obtained from the Labor Force

For the sake of realism, we first focus on IV estimates which have been obtained using the population of labor force participants. The results are in Table 4. Apart from IV estimates and their standard error, we report OLS estimates of the effect of education on wages (log), as well as the difference between IV and both the population LATE parameter and the population average (namely $\hat{\lambda}_{iV} - \lambda_L$, and $\hat{\lambda}_{iV} - \bar{\lambda}$). We also report several correlations between the instrument and various components of the composite error term. These include the correlation between the instrument and (i) the composite error term ($Corr(Z_i,\ \varepsilon^*_{it})$), (ii) post-schooling choices ($Corr(Z_i,\ \varphi_{it}(.))$, and (iii) the product of schooling times individual returns ($Corr(Z_i, \omega_{\lambda i} \cdot S_i)$). Finally, we also report the correlation between the instrument and schooling, as well as the first stage regression F Statistic.

[29] Obviously, it would be possible to construct another model in which the LATE would be more sensitive to sampling. The econometrician would then need to decide which LATE raises more interest. As far as we know, this issue (rather fundamental in our opinion) is practically never discussed in empirical IV papers using a LATE interpretation.

8.2.1 The Dynamic Model

As a starting point, we examine the implementation of the low intensity experiment within the dynamic model. The reduction in the total number of observations (ranging between 7000 and 8000 individuals according to policy intensity) obviously reflects that the labor force is only a subset of the general population. The IV estimate is equal to 0.10. As indicated earlier, the population LATE (as well as the Labor Force LATE) is virtually three percentage points lower (0.07).[30] IV is not only inaccurate, but is also very imprecise (with an estimated standard error (0.0374). This high degree of inaccuracy prevails while the correlation between the instrument and schooling (equals to 0.05) implies a reasonably high first stage F-statistic (17.9). Interestingly, and as is the case in many empirical applications, the OLS estimate computed from the labor Force population (equal to 0.09) is inferior to the IV estimate.

We now focus on the potential reduction of this inaccuracy. According to the weak-instrument literature, increasing the strength of the instrument is desirable. We now examine the inaccuracy that would result from the implementation of the high subsidy policy. After all, the F Statistic (which is equal to 63.9) obtained with the high-subsidy is three times as large as with the low subsidy. As noted before (equation 15), and because both in the low level intensity and in the high level intensity cases, IV over-estimates the LATE, a reduction in asymptotic bias requires that the change in the population LATE induced by the instrument $(\lambda_L(Z_{high}) - \lambda_L(Z_{low}))$ must be larger than the difference in asymptotic biases $(As.Bias_{high} - As.Bias_{low})$.[31] However, this is not the case. The IV obtained

[30] If the LATE parameters had been computed from the labor force population, the low and high intensity values would be equal to 0.0708 and 0.0713 respectively. This illustrates well the very low elasticity of the LATE parameter with respect to policy intensity (in the current model).

[31] As noted earlier, inequalities have to be adjusted according to whether IV over-estimates or under-estimates its target.

with the high subsidy is equal 0.1026. Despite an increase in precision (the standard error is around 0.019), IV is even more inaccurate, since the difference in asymptotic biases is strictly positive while the difference between the high and low intensity LATE is virtually equal to 0 (0.0004). So, an increase in statistical power of the experiment has translated into a reduction in accuracy.

There are two issues that need to be explained. The first one is the failure of IV. The second one is the fragility of the weak/strong instrument paradigm.

To understand the failure of IV, it is important examine the composition of the error term. Once a new policy sets in, a subset of the population reacts by increasing schooling. Given this increase, two mechanisms are playing at the same time.

First, an increase in schooling (caused by the policy shock) raises the attractiveness of employment and training, at the detriment of household production. This automatically changes $Corr(Z_i,\ \varphi_{it}(.))$. This first mechanism reflects the causal effect of schooling on subsequent skill accumulation, after conditioning on individual heterogeneity.

However, as individual who react are also endowed with different returns to experience and training, this also translates into a correlation between the policy shock indicator and the error term. This is the second mechanism. It is a pure composition effect. It is present even if schooling does not cause future skill accumulation, after controlling for individual heterogeneity.

Given this, the fragility of the relationship between IV accuracy and instrument's power, may easily be explained. As noted earlier, within a dynamic setting, a change in policy intensity not only changes the correlation between the instrument and education choices, but also changes the correlation between the instrument and the error term. In this specific example (the dynamic model applied to the labor force), an increase in the correlation between Z and post-schooling choices from 0.0074 to 0.0099 (as seen in Table 4), is sufficient to raise the correlation between Z and the error term from 0.012 to 0.024. This reduces

IV accuracy. This is all happening despite a huge increase in the significance of the first stage regression.

8.2.2 The Static Model

We now consider the static version of the model. The IV estimate resulting from the low subsidy experiment is equal to 0.0292. This is far below the population LATE, which is equal to 0.0789 with the low intensity policy. However, IV is also particularly imprecise (as indicated by its standard error equal to 0.026). With the higher intensity policy, the IV estimate increases to 0.0858, to reach a value that exceeds the LATE, which is equal to 0.0781 with the low intensity policy.[32] Compared to what was observed for the dynamic model, or for the low intensity version of the static model, this may appear as a reasonable performance since the difference between $\hat{\lambda}_{iV}$ and λ_L is 0.0077.

This very high sensitivity of IV is again well illustrated by the correlations between Z and various components of the error term. Both the correlation between the instrument and the product of schooling times individual deviations from the population returns $(Corr(Z_i, \omega_{\lambda i} \cdot S_i)$, and the correlation between the instrument and post-schooling choices $(Corr(Z_i,\ \varphi_{it}(.)))$ are strongly affected by the change in intensity. They go respectively from 0.0010 to 0.0236, and from -0.0062 to 0.0050. As a result, increasing the power of the instrument is beneficial to IV accuracy in the static version of the model, even though it is detrimental in the dynamic version. In other words, the capacity to increase IV accuracy with a stronger instrument is model dependent. This illustrates the fragility of the weak-strong instrument paradigm, within a dynamic context.

[32] For the static version of the model, the LATE parameters computed from the labor force population (equal 0.0780 for the low intensity case, and to 0.0781 for the high intensity case) are practically equivalent to those computed for the general population.

8.3 Some Preliminary Conclusions

At this stage, the performance of IV appears very bad. Except for the static version of the model with high policy intensity, IV seems incapable of estimating the population LATE parameter. Before considering alternative sampling designs, the following remarks should be noted. A natural guess is to impute the failure of IV to classical composition (selection) effects. In theory, low productivity individuals are less likely to be observed in the sub-population analyzed by applied econometricians (other things equal). In our model, individual attraction toward work is explained by returns to work experience (including training), individual returns to schooling and even discount rates. In Table 7, we report the average endowments of all sub-populations analyzed in the paper, along with the population average. Indeed, these numbers indicate that the Labor Force is composed of individuals who have higher returns to schooling, as well as high returns to work and training (state a). While implementing an IV on the labor force requires cross-sectional selection, it is not the only reason for IV failure.[33]

There is another explanation. As pointed out earlier, classical IV is not only dependent on cross-sectional sample composition, but also on the time period at which outcomes are measured. While there exists a single LATE for a given population (or sub-population), there is a multiplicity of IV estimates that may be associated to the same population. This is due to the extremely high number of possible individual/age combinations that are possible to construct if the econometrician uses cross-section data (if he/she samples only one wage per individual). Theoretically, the same group of cross-sectional units sampled at two different points in time, would give different IV estimates, simply because the cor-

[33]For instance, as an experiment, we also computed an IV on the general population using both the control group and the treatment group, using the same time sampling procedure. Precisely, we used individual wages regardless of work decisions (10,000 observations). The performance of IV turned out to be as bad as in the Labor force (the four different IV estimates were 0.0353, 0.0789, 0.1205 and 0.1289).

relation between the instrument and post-schooling choices evolves with calendar time. As most cross-section data used to infer wage returns to schooling contain a relatively balanced age distribution (approximately uniform), it follows that statistical inference based on this specific sampling structure may be dependent on this very specific pattern.

9 Implementing Alternative Sampling Designs

Given the failure of classical IV applied to the Labor Force, it is natural to investigate alternative strategies that may help improve IV estimates. Formally, the issue is to find a new sampling procedure that defines a new composite error term, say ε_{it}^{**}, such that

$$plim(\frac{Z'\varepsilon^{**}(Z)}{N}) = plim\frac{Z'S(Z)}{N} \cdot (\lambda_L(Z) - \bar{\lambda}) \tag{14}$$

As pointed out earlier, sampling data from a population generated by a dynamic structure with unobserved heterogeneity, entails both a cross-sectional and a time dimension. Precisely, the econometrician has to find a sampling procedure that either annihilates $\varphi_{it}(.)$, or annihilates its effect on the correlation between the instrument and the composite error term. Obviously, all of that must occur without perturbing the distribution of individual specific returns. This is not necessarily a simple task. For instance, any cross-sectional sampling procedure that would turn out to reduce $\varphi_{it}(.)$ for all individuals in the sample, may also change the distribution of $\omega_{\lambda i} \cdot S_i$, if it implies selecting a different sub-population (see equation 12). So, reducing $\varphi_{it}(.)$ is not necessarily sufficient to achieve higher IV accuracy.

A natural guess is to consider sampling strategies that either restrict the degree of heterogeneity, or limit the impact of the dynamic effect of schooling.

We now turn our attention to these issues. Precisely, we analyze the impact of a Regression Discontinuity design and an age-based sampling strategy that consists of sampling individuals during a specific time/age interval over the life cycle.

9.1 A Regression Discontinuity Design

Regression Discontinuity (RD) design has appeared in the recent micro-econometric literature, within the context of static models.[34] It is particularly popular in empirical labor economics. Within our framework, one group (the control group) faces the old policy, while the second group started under the old regime and experiences the new policy regime at the time when the decision to enter higher education is made. The idea is to compare the treatment group with the subset of the control group that is at the margin of entering higher education (grade level 5) under the old regime.

In the applied literature, the effectiveness of this approach relies on a potential reduction of the degree of cross-sectional heterogeneity. However, in a dynamic model where returns to schooling are individual specific, it is far from guaranteed that a RD design will perform better than classical IV.[35]

In order to implement this sampling design, we take the labor force population that was analyzed in the previous section, and select a sub-population of individuals who have completed at least 4 periods of schooling. We then construct a similar treatment group and simulate their subsequent choices under a new policy. Obviously, Regression Discontinuity designs always imply (by construction)

[34] See Hahn, Todd and van der Klaauw (2001) for a theoretical presentation and van der Klauw (2002) for an application.

[35] In general, the implementation of a specific RD design within a particular data generating process, could create a serious divergence between the population LATE parameter, and the LATE parameter of this hypothetical new population. Furthermore, the results obtained may be representative of a very limited set of the population. However, and as indicated earlier, this is not an issue in the present model.

an important reduction in sample size. In our example, the RD sub-population contains a total of 4000 to 6000 individuals.[36]

9.1.1 The Dynamic Model

Estimates are reported in Table 5. Interestingly, and as opposed to what was noted in the Labor Force population, RD design estimates are below the true population LATE parameters regardless of the power of the instrument. Indeed, in our specific example, RD design provides estimates that are closer to the population average (equal to 0.06). The estimates obtained for the dynamic model, 0.0551 (low subsidy), and 0.0552 (high subsidy), reveal a surprising degree of inelasticity of IV with respect to policy intensity.

There is a clear explanation for this. In the low subsidy case (in which the F statistics is 53.0), the correlation between the instrument and the composite error term is a small negative number that approaches 0 (-0.0018), despite a positive correlation between the instrument and the product of schooling times individual deviations from the population returns (0.0073), as well as a positive correlation between the instrument and post-schooling choices (0.0073). As a result, because the instrument appears to be nearly orthogonal with the error term, IV does not even capture the minimal degree of non-orthogonality that is present in standard static random coefficient models, and approaches the population average. This gives the illusion that the LATE is equal to the population average, but again, it is explained by mis-specification of IV. The same phenomenon occurs with the stronger policy intensity.

When compared to the standard labor force participants, the degree of inaccuracy is lower, but remains relatively high, as IV-LATE differences are equal to

[36]While this particular design implies a more significant change in population composition (a larger fraction of individuals with academic abilities), as is obvious upon examining Table 7, its effect on the true LATE parameter (just like what was observed for the labor force) was found to be quite small.

-0.0159 (low subsidy), and to -0.0162 (high subsidy). So, the RD design is not sufficient to eliminate the inaccuracies found with the workforce population. As with classical IV applied to the labor force population, a huge increase in the instrument-schooling correlation (or a huge increase in first stage F statistic) has no impact on accuracy. Therefore, and has noted earlier in other circumstances, increasing the power of the instrument does not necessarily imply a reduction in the IV-LATE difference.

9.1.2 The Static Model

Similar results are found in the static version of the model. As it was the case for the dynamic model, IV underestimates the LATE. The IV-LATE differences are equal -0.0224 (low subsidy), and -0.0178 (high subsidy) and are even closer to the population average. For this sub-population, the correlation between the instrument and post-schooling choices becomes negative (-0.004 in both cases). This negative correlation annihilates the positive correlation between the instrument and the product of schooling times individual deviations from the population returns. Again, the instrument appear to be nearly orthogonal with the composite error term.

To summarize, there is no clear evidence that a Regression Discontinuity design does better than standard IV. Still, it is important to notice that IV estimates are very sensitive to the use of a Regression Discontinuity design.

9.2 An Age-Based Sampling Design

An aged-based design originates from the intuitive conjecture that, in the case where the mis-specification of IV depends heavily on post schooling choices, IV performed on early wages (when $\varphi_{it}(.)$ approaches 0 for all individuals) may be less affected by the dynamic structure of the model than more standard IV applied to life-cycle wages. One way to annihilate $\varphi_{it}(.)$ completely would be to sample only entry wages. However, such an empirical strategy would require to discard a large proportion of most standard cross-section data sets. Indeed, we do not know of any cross-sectional study that does so. Consequently, we impose an age limit so that the period elapsed between market entrance and actual sampling is smaller than in the Labor Force sampling strategy. This approach is feasible for any applied econometricians who have access to a relatively large cross-section.

To implement this, we start from the original Labor Force population. We compress the age distribution of the potential sampling period, and select individuals only between period 5 and period 19 (as opposed to period 5 to period 33).

Because sampling early career wages still requires some selection based on contemporaneous decisions, it is impossible to say if the early career design will lead to higher IV accuracy. For instance, over-sampling early wages implies a larger dependence on schooling decision outcomes, since a larger portion of wages will now be measured between period 5 and period 10. In other words, while early wages may be purged (to some extent) of the contamination introduced by post-schooling choices, they may be more sensitive to individual decisions to enrol in higher education.

Because we still select wages of those who work, the results (displayed in Table 6) are naturally comparable with those of Table 4.

9.2.1 The Dynamic Model

We first examine the dynamic version of the model. There is clear evidence that the performance of IV, applied to a sampling design that targets individuals when they are younger, is superior to the one observed for the broader sampling window (period 5-period 33). The high and low intensity estimates (0.0727 and 0.0685) are within very small distance of their estimand (0.0710 and 0.0714). The IV-LATE differences (0.0017 and -0.0029) display the best performance that we have encountered so far.

However, and as this was the case earlier, there is no gain imputable to a statistically stronger instrument. Despite a correlation between the instrument and schooling that is multiplied by 3, and a F Statistic multiplied by 6, the high intensity IV does not perform better. Precisely, the IV-LATE distance, equals to 0.0017 with the weaker instrument, is larger in absolute value (0.0029) with the stronger policy.

9.2.2 The Static Model

To say the least, the very good performance of IV within the early-career wage sampling design (in the dynamic model) does not carry to the static version of the model. IV estimates obtained for the static version of the model (equal to 0.1297 and 0.1212) lie outside the support of the distribution of returns to schooling. The distance between IV and the relevant LATE parameters ranges from 4 to 5 percentage points. This is one of the worst performances that we have observed in our analysis. This may easily be explained. Because individuals obtain less schooling in the static version, the early-career sampling procedure will generally select a different population within the static model. For instance, a portion of individuals who would be in school in period 5 (when sampling starts) in the dynamic model, are now likely to be selected in the static version of the model. So, in the static case, the early-career sampling procedure is detrimental to IV accuracy.

To summarize, while the early career sampling design performs well in the dynamic model, it cannot be regarded as a general solution to IV estimation of the return to schooling. The gain obtained from narrowing the sampling window dominates the loss imputable to cross-sectional selection, within the dynamic version of the model, but not in the static version

10 What does IV Estimate?

Given that IV fails to recover the population LATE parameter, and that we have not found any alternative sampling design that can guarantee IV accuracy, it is natural to investigate what treatment effect can IV estimate. Because it disregards post-schooling choices, it is possible that IV only estimates some reduced-form marginal effect of schooling on wages. Such a quantity may capture indirect effects of schooling on wages, through the incidence of training and work experience.

As empirical labor economists often use the term "causal" parameter when referring to those parameters that are associated to a subset of the population affected by some experiment, a natural candidate would be the average wage gain for those who have increased their level of schooling following the higher education subsidy (for those who have been affected).[37] Formally, this would be

$$\frac{\delta E(w_{it})}{\delta S_i} |_{S_{i(Z_{low})} \neq S_{i(Z_{high})}} \tag{15}$$

Because we have two different policy intensities, we computed two different estimates for each model. As the marginal effect is age/time dependent (as is IV), we used the same sampling method used throughout the analysis, and considered only one wage per individual (between period 5 and period 33). The marginal effects are computed using simple OLS regressions. These numbers are reported in the summary table (Table 8).

As intuition would suggest, the marginal effects of schooling on wages exceed the population LATE parameters. For the dynamic model, the marginal effects are equal to 0.1797 and 0.1617. For the static model, the marginal effects are equal to 0.1780 and 0.1110. In all cases, the marginal effects exceed the corresponding

[37]The notion of causality used by many empirical labor economists is different than the usual definition used by economists (namely a parameter that captures the effect of a counterfacatual change in one variable, holding other variables fixed).

IV estimates.[38]

At this stage, the relevant question is whether or not IV estimates tend to be closer to these marginal affects, than they are to the LATE parameters. Without loss of generality, we focus our attention on the IV estimates obtained from the labor force. The answer is obvious. A quick examination of Table 8 reveals clearly that IV estimates are not closer to the reduced-form marginal effects of those affected (equation 15) than they are to the relevant LATE parameters.

This raises an interesting question. While (15) is defined for the sub-population affected by the experiment, its unconditional version is worth examined. Indeed, OLS estimates of the effect of education on wages found in columns 1 and 4 of Table 2, or those reported for all different sampling designs (found in Table 8) are most likely good estimates of this reduced-form effect. To avoid confusion with the marginal effects of schooling that depend on individual reactions to policy changes, we refer to marginal effects computed from simple OLS (and regardless of counterfactual outcomes) as "unconditional marginal effects". As noted earlier, and aside from those computed for the very specific RD design sub-population, these OLS estimates gravitate around 0.10, depending on the sub-population considered. They range between 0.08 (for the early career sampling), and 0.12 (for the control group).

The interesting issue is now to determine if IV estimates are closer to the conditional or the unconditional marginal effects. The answer is also clear. In general, IV estimates are much closer to unconditional OLS estimates of the effect of education, than their conditional counterparts. This is the case for both IV estimates obtained in the dynamic model, and for the low-intensity IV estimate associated to the static model. For the high intensity policy applied to the static

[38]Indeed, these marginal effects could also have been computed from each sub-population (the control group, the general labor force, or the RD design group). As the results are quite similar, and for the sake of clarity, we use only the population marginal effects as potential estimand in order to illustrate our points.

model, IV is closer to the conditional marginal affect than the unconditional, but, it is still far from both. Indeed, as we had noticed earlier, the high intensity/static model was virtually the only case where IV approached the population LATE.

Interestingly, these results carry to both the RD design and the early career sampling strategies. Indeed, because IV estimates obtained from the Regression Discontinuity design are always below the population LATE parameter, the difference between IV estimates and the marginal effects of schooling (either the conditional or the unconditional one) are even larger.

To conclude, given the failure of IV to recover the population LATE parameter, there is no evidence that IV is even capable of estimating any reduced-form marginal effect of schooling on wages for those affected by the policy experiment. Within a dynamic setting, IV therefore appears to be deprived of any "causal" substance.

11 Concluding Remarks

In many fields of economics (especially in Labor Economics), it is common to estimate an outcome equation under the maintained hypothesis that outcomes measured over the life-cycle are affected by a single endogenous (choice) variable. In such a case, and under certain conditions, IV may converge to some interpretable parameter. This is exemplified in the empirical literature devoted to the estimation of returns to schooling, in which inference is typically based on a single instrument, and in which high IV estimates are said to arise because they reflect the LATE parameters of a subpopulation of individuals affected by liquidity constraints, or of a subpopulation of individuals who have high discount rates.

This line of reasoning is made of two different conjectures. One is about the link between the underlying heterogeneity structure and individual choices. Precisely, it says that those affected by the experiment have high returns to schooling, but would have not reached higher education in absence of policy exposure, because they may be endowed with high discount rates (or face liquidity constraints). The second one is about the capacity of IV to recover the LATE parameter.

The analysis presented in this paper has illustrated that, within a dynamic setting, the second argument may be particularly wrong. As seen earlier, our model implies a LATE parameter that exceeds the population average, for exactly these same reasons.[39] Yet, IV estimates are, in some cases, much larger than the relevant treatment effect parameter, and in other cases (like in the RD design), much smaller.

In a dynamic setting, IV seems to estimate neither the population LATE parameter, nor any reduced-form "causal" effect of schooling on wages. Unless the econometrician is capable of finding a sampling design that annihilates the effect of post-schooling dynamics (so to recover the basic properties of a static correlated random coefficient wage regression model in which the correlation between the instrument and the error term is solely explained by the correlation between schooling and individual returns), he/she must implicitly make at least one of the following assumptions; a precise correlation structure characterizing unobserved heterogeneity, very specific distributions of random shocks, a highly restrictive law of motion, or particular preferences. Without any implicit "functional form restrictions", equation (14) is not verified, and IV is deprived of any "causal" substance. Access to a "strong" instrument is not a solution.

The analysis presented in the paper was confined to one specific parameter; namely the return to schooling. It need not be the case. We believe that our conclusions could be transported to a wide class of parameters that economists typically estimate using life-cycle data on choices and outcomes. The economic

[39] It is generally impossible to say if the first conjecture is correct. We performed our analysis within a theoretical framework where the LATE exceeds the population average. Obviously, it would be possible to construct a model structure where the opposite would prevail.

returns to work experience, or to on-the-job training, the effect of work interruptions on female wages, the impact of unemployment duration or unemployment incidence on post-unemployment opportunities, the effect of job displacement on earnings, or the effect of child birth on female labor market outcomes, are all examples of parameters that applied econometricians may be tempted to estimate by IV. Because all of these parameters could easily be interpreted within a truly dynamic theoretical structure, we conjecture that classical IV could suffer the same problems that have been encountered in the present paper.

Obviously, it is the econometrician's prerogative to assume a particular data generating process, or to prefer a static representation of the labor market over a dynamic one. After all, with life-cycle data on choices and outcomes, it is not possible to establish whether or not the data generating process is dynamic or static. This is particularly true because identifying the present component from the future component of an intertemporal utility function is difficult, and because the exact structure of the law of motion is difficult to establish if the dynamics of the underlying model is explained by multi-dimensional unobserved state variables (or parameters). Every estimation method must rely on fundamentally subjective assumptions. So, every micro-econometric model is mis-specified, and the choice between IV and structural methods is also subjective. This is a triviality. However, interpreting IV estimates of an outcome equation outside of the "one endogenous variable/one instrument" paradigm, is not an innocuous task. This cannot be debated.

Table 1A

Life Cycle Choices in the Dynamic Model

Average Number of Periods Accumulated in each State, by Date.

	state **(s)**	state **(e)**	state **(a)**	state (h)
Date	**School**	**Work**	Work/**Training**	**Home**
1	0.824	0.068	0.000	0.108
2	1.568	0.190	0.000	0.242
3	2.272	0.353	0.000	0.375
4	2.956	0.528	0.000	0.516
5	3.474	0.786	0.087	0.653
6	3.936	1.085	0.189	0.790
7	4.351	1.392	0.335	0.922
8	4.727	1.712	0.505	1.056
9	4.891	2.047	0.877	1.185
10	4.988	2.417	1.276	1.319
15	5.045	4.451	3.545	1.959
20	5.045	7.358	4.981	2.616
25	5.045	11.442	5.223	3.290
30	5.045	15.736	5.259	3.960
33	5.045	18.333	5.259	4.363

Table 1B

Life Cycle choices in the Static model

Average Number of Periods Accumulated in each State, by Date.

Date	**state (s)** **School**	**state (e)** **Work**	**state (a)** **Work/Training**	**state (h)** **Home**
1	0.639	0.115	0.014	0.232
2	1.186	0.307	0.015	0.492
3	1.720	0.511	0.015	0.754
4	2.247	0.713	0.017	1.023
5	2.631	0.954	0.122	1.293
6	2.949	1.229	0.254	1.568
7	3.240	1.499	0.432	1.829
8	3.503	1.800	0.605	2.092
9	3.652	2.094	0.893	2.361
10	3.750	2.409	1.218	2.623
15	3.815	4.094	3.168	4.632
20	3.815	6.491	4.455	6.250
25	3.815	9.890	4.707	6.588
30	3.815	12.766	4.780	7.639
33	3.815	15.699	4.781	8.705

Table 2A

Summary Statistics of Life Cycle Wages

	Dynamic Model		Static Model	
Date	**Mean**	**Std dev.**	**Mean**	**Std dev.**
1	5.60	2.95	5.56	2.89
2	5.92	3.25	5.93	3.17
3	6.29	3.39	6.26	3.35
4	6.62	3.65	6.54	3.51
5	6.97	3.95	6.91	3.91
6	7.34	4.10	7.14	4.04
7	7.80	4.52	7.62	4.55
8	8.24	5.17	8.04	4.91
9	8.67	5.51	8.36	5.44
10	9.05	5.85	8.70	5.77
15	11.26	8.27	10.66	8.28
20	13.12	11.02	12.43	10.59
25	13.62	10.89	13.29	11.65
30	14.55	12.06	13.88	11.95
33	15.23	12.80	14.53	12.63

Table 2B

OLS regressions on Simulated Wages

	Dynamic Model			**Static Model**		
		estimate (t-ratio)			estimate (t-ratio)	
education	0.1220	0.1328	0.1332	0.1206	0.1316	0.1334
	(51.9)	(20.10)	(22.11)	(53.6)	(21.0)	(20.18)
experience		0.0123	0.1090		0.0153	0.0934
		(4.23)	(3.55)		(4.18)	(3.08)
experience^2		-	-0.0057			-0.0046
			(3.18)			(2.66)
# of individuals	5000	5000	5000	5000	5000	5000
R^2	0.26	0.28	0.30	0.25	0.27	0.33

Note: The regressions are computed on a cross section of wages generated from simulated choices in the control group. The cross section contains 5,000 observations (20 types multiplied by 250 different realizations of the full vector of random shocks).

Table 3

The effects of a Higher Education Subsidy in the Population: Local Average Treatment Effects of Education and Instrument Power

Level of Subsidy	**Low**	**High**
Dynamic Model		
$\Delta Schooling$	0.25	0.54
LATE (λ_L)	0.0710	0.0714
Proportion affected	0.175	0.315
F Stat	12.3	57.9
Static Model		
$\Delta Schooling$	0.13	0.50
LATE (λ_L)	0.0789	0.0781
Proportion. affected	0.093	0.258
F Stat	3.3	45.4

Note: The higher education subsidy is 1$ in the low subsidy experiment, and 3$ in the high subsidy experiment. $\Delta Schooling$ is the difference between the average years of schooling in the treatment group and the control group. The LATE is defined as $E(\lambda \mid S_{i(Z_{low})} \neq S_{i(Z_{high})})$ and the proportion affected is the fraction of the population for whom $\Delta Schooling_i \neq 0$. The F statistic is computed from an OLS regression of schooling on the policy shock exposure indicator.

Table 4

Classical IV in the Labor Force

Model	**Dynamic**		**Static**	
Level of the Subsidy	**Low**	**High**	**Low**	**High**
$\hat{\lambda}_{iV}$	**0.1000**	**0.1026**	**0.0292**	**0.0858**
st-error of $\hat{\lambda}_{IV}$	0.0374	0.0190	0.0260	0.0170
$\hat{\lambda}_{OLS}$	0.0917	0.0905	0.0921	0.0908
$\hat{\lambda}_{iV} - \lambda_L$	0.0288	0.0312	-0.0497	0.0077
$\hat{\lambda}_{iV} - \bar{\lambda}$	0.0398	0.0426	-0.0308	0.0258
$Corr(Z_i,\ \varepsilon^*_{it})$	0.0119	0.0240	-0.0015	0.0168
$Corr(Z_i,\ \varphi_{it}(.))$	0.0074	0.0099	-0.0062	0.0050
$Corr(Z_i, \omega_{\lambda i} \cdot S_i)$	0.0097	0.0104	0.0010	0.0236
$Corr(Z_i,\ S)$	0.0481	0.0905	0.0067	0.0882
F Stat	17.9	63.9	0.3	54.7
# of individuals	7717	7738	7031	6981

Table 5

Alternative Sampling Designs:

Regression Discontinuity at Grade Level 5

Model	**Dynamic**		**Static**	
Level of the Subsidy	**Low**	**High**	**Low**	**High**
$\hat{\lambda}_{iV}$	**0.0551**	**0.0552**	**0.0565**	**0.0603**
st error of $\hat{\lambda}_{iV}$	0.0367	0.0171	0.0600	0.0162
$\hat{\lambda}_{OLS}$	0.1331	0.1419	0.1398	0.1358
$\hat{\lambda}_{IV} - \lambda_L$	-0.0159	-0.0162	-0.0224	-0.0178
$\hat{\lambda}_{IV} - \bar{\lambda}$	-0.0049	-0.0048	-0.0035	0.0003
$Corr(Z_i,\ \varepsilon^*_{it})$	-0.0018	-0.0038	-0.0009	0.0003
$Corr(Z_i,\ \varphi_{it}(.))$	0.0064	0.0149	-0.0037	-0.0037
$Corr(Z_i, \omega_{\lambda i} \cdot S_i)$	0.0073	0.0181	0.0097	0.0351
$Corr(Z_i,\ S)$	0.0959	0.2148	0.0588	0.2240
F Stat	53.0	278.4	15.3	227.6
# of individuals	5708	5758	4004	4310

Table 6

Alternative Sampling Design:

Early Career Sampling (period 5–period 19)

Model	**Dynamic**		**Static**	
Level of Subsidy	**low**	**High**	**low**	**High**
$\hat{\lambda}_{iV}$	0.0727	0.0685	0.1297	0.1212
st error of $\hat{\lambda}_{iV}$	0.0530	0.0215	0.0707	0.0245
$\hat{\lambda}_{OLS}$	0.0801	0.0800	0.0871	0.0863
$\hat{\lambda}_{IV} - \lambda_L$	0.0017	-0.0029	0.0508	0.0431
$\hat{\lambda}_{IV} - \bar{\lambda}$	0.0127	0.0085	0.0697	0.0612
$Corr(Z_i,\ \varepsilon^*_{it})$	0.0029	0.0047	0.0127	0.0320
$Corr(Z_i,\ \varphi_{it}(.))$	-0.0104	0.0080	-0.0022	-0.0018
$Corr(Z_i, \omega_{\lambda i} \cdot S_i)$	0.0031	0.0127	0.0117	0.0210
$Corr(Z_i,\ S)$	0.0351	0.0854	0.0247	0.0685
F stat	8.5	50.5	3.9	29.7
# of individuals	6887	6871	6430	6292

Table 7

The Heterogeneity Distribution in Various Sub-Populations

	α^S	α^H	λ	θ	δ	c_0^a	β
General Population (control group)	0.5000	2.8000	0.0600	0.0100	0.0300	5.1300	0.9485
Labor Force (Table 4)	0.5751	2.8003	0.0637	0.0121	0.0318	5.0895	0.9495
Regression Discontinuity (Table 5)	0.7748	2.7931	0.0751	0.0109	0.0376	4.9817	0.9549

Table 8

A Summary of Results

Population parameters, IV, and OLS

Model	**Dynamic**		**Static**	
Level of the Subsidy	**Low**	**High**	**Low**	**High**
LATE (λ_L)	0.0710	0.0714	0.0789	0.0781
Pop Average ($\bar{\lambda}$)	0.06	0.06	0.06	0.06
$[\lambda_{\min}, \lambda_{\max}]$	[0.01,0.12]	[0.01,0.12]	[0.01,0.12]	[0.01,0.12]
$\frac{\delta E(w_{it})}{\delta S_i} \mid_{S_{i(Z_{high})} \neq S_{i(Z_{low})}}$	0.1797	0.1617	0.1780	0.1110
IV (Work Force)	0.1000	0.1026	0.0292	0.0858
IV (RD Design)	0.0551	0.0552	0.0565	0.0603
IV (early career)	0.0727	0.0685	0.1297	0.1212
OLS (Work Force)	0.0917	0.0905	0.0921	0.0908
OLS (RD Design)	0.1331	0.1419	0.1398	0.1358
IV (early career)	0.0801	0.0800	0.0871	0.0863
OLS (control group)	0.1220		0.1206	

Appendix: Table A1 - The Heterogeneity Distribution

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

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

Table A2

The correlation between the heterogeneity components

	cost of school	value of home time	return to education	return to experience	return to training	cost of training	discount factor
	α^S	α^H	λ	θ	δ	c_0^a	β
α^S	1.00	-0.009	0.857	0.111	0.857	-1.00	-0.106
α^H		1.00	-0.043	0.087	-0.043	0.009	-0.046
β			1.00	0.110	1.00	-0.857	-0.076
θ				1.00	0.110	-0.111	0.112
δ					1.00	-0.857	-0.076
c_0^a						1.00	0.106
$\frac{1}{1+r}$							1.00

References

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

[2] Belzil, Christian (2008) "Testing the Specification of the Mincer Wage Equation", forthcoming in *Annals of Economics and Statistics*

[3] Belzil, Christian (2007) "The Return to Schooling in Structural Dynamic Models: A Survey of the Literature" *The European Economic Review, vol 51, 5,*

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

[5] Belzil, Christian and Hansen, Jörgen (2007) "A Structural Analysis of the Correlated Random Coefficient Wage Regression Model" *Journal of Econometrics, vol 140.*

[6] Belzil, Christian and Hansen, Jörgen (2005) " Structural Analysis of the Correlated Random Coefficient Wage Regression Model with an Application to the OLS-IV Puzzle, IZA working Paper 1585

[7] Berry, Steve, Levinsohn James, and Ariel Pakes (1995) "Automobile Prices in Market Equilibrium" Econometrica, Vol. 63, No. 4, 841-890.

[8] Björklund, A. and Robert Moffit (1987) "The Estimation of Wage Gains and Welfare Gains from Self-Selection Models", *Review of Economics and Statistics, 69, 42-49.*

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

[10] Hahn, J, Todd, P. and van der Klaauw (2001) "Identification and Estimation of Treatment Effects with a Regression Discontinuity Design", *Econometrica 69, 201-209.*

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

[12] Heckman, James, Lance Lochner and Christopher Taber (1998) "General Equilibrium Treatment Effects: A Study of Tuition Policy, *American Economic Review, vol 88, no. 2*

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

[14] Imai, Susumu and Michael P. Keane (2004) "Intertemporal Labor Supply and Human Capital Accumulation", *International Economic Review*, vol 45, 2, 601-639.

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

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

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

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

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

[20] Rust, John (2005) Comments on: "Structural vs. Atheoretic Approaches to Econometrics, by Michael Keane", Unpublished Manuscript

[21] Staiger, Douglas and James H. Stock (1997),"Instrumental Variables Regression with Weak Instruments" *Econometrica, 65*: 557-586.

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

[23] Todd, Petra and Kenneth 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,*Vol. 96, No. 5, (December)

[24] van den Berg, Gerard (2007) "An Economic Analysis of Exclusion Restrictions for Instrumental Variable Estimation", IZA Working Paper 2585.

[25] Van Der Klaauw, W. (2002) "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-discontinuity Approach" *International economic Review 43, 1249-1287.*