The Return to Schooling and Experience and the Ability Bias in Structural Models: A Survey*

Christian Belzil
Centre National de Recherche Scientifique
(GATE, Ecully, France)
Institute for the Study of Labor (IZA),
CIRANO and CIREQ.

May 3, 2005

Abstract

This papers contains a survey of the recent literature devoted to the returns to schooling within a structural econometric framework. In the literature, agents are forward looking, labor market wages are modeled within a Mincerian framework and the related parameters are given a skill production function interpretation, there is a rich multi-dimensional heterogeneity (unobserved) specification and statistical inference is made from panel data. I first present a simple theoretical model of lifecycle time allocation between human capital, search and employment activities that illustrates the nature of the structural approach and I review a set of papers representative of the literature. The following discussion is centered upon three main objectives: the identification of the key dimensions of the structural literature

*This paper has been largely influenced by several discussions I had with Zvi Eckstein. I thank him for many ideas and suggestions. Michael Bognanno, Donghoon Lee, Jorgen Hansen, Susumu Imai, Lance Lochner, Claude Montmarquette, Robert Sauer and Uwe Sunde provided useful comments.

that distinguishes it from the experimental (reduced-form) literature, a comparison between structural and reduced-form estimates and the comprehension of the source of discrepancy between structural and reduced-form estimates. The empirical results stress the very high correlation between the utility of attending school and labor market skills (in a single skill framework). In a multiple skill framework, white collar skills (as opposed to blue collar skills) are highly correlated with the utility of attending school. The importance of non-linearities as well as population heterogeneity in the returns to schooling is also a highlight of the structural literature. Finally, and more importantly, most estimates point to the coexistence of relatively low returns to education (compared to the experimental literature) and relatively high returns to labor market experience.

Key Words: *Returns to Schooling, Human Capital, Ability Bias, Dynamic Programming, Dynamic Self-Selection.*

JEL Classification: J2-J3.

1 Introduction

The return to schooling is probably one of the most investigated parameters in modern economics. It plays a central role in microeconomic models of human capital accumulation and is also important in the empirical growth literature. At the policy level, increasing the level of education in the population is generally regarded as a desirable goal and knowing the effect of schooling on wages or labor market productivity is therefore a major concern to most policy makers.

Because schooling decisions are potentially affected by unobserved individual skills and tastes which are also correlated with individual wages, economists have been reluctant to associate a structural interpretation to the positive correlation between schooling and wages. Indeed, the sign of the statistical bias that may occur when measuring the causal effect of schooling on wages by a simple correlation or by OLS techniques, typically referred to as the "ability bias", has been the object of much debate over the last 40 years.

In view of the general interest in skill formation policies, it is important to obtain a measure of the true causal effect of schooling on wages. Indeed, this objective has been at the forefront of the empirical labor economics literature for many years.

As is the case for all econometric models plagued with endogenous variables, there are two alternative approaches in estimating the return to schooling and experience. One approach is to rely on exogenous variables that are correlated with schooling but uncorrelated with the error term of the wage equation and "instrument out" schooling so to obtain an independent variation. A second approach consists of modeling schooling decisions jointly with wage outcomes. In order to achieve this, the econometrician must build a model in which both the causal and the spurious effects of education on wages are separately identifiable. In modern econometric jargon, this approach is referred to as "structural".[1]

Until recently, the literature on the return to schooling had been completely dominated by empirical strategies based on experimental and instru-

[1] In the econometrics literature, the term "structural" is sometimes used to designate modeling strategies which are quite different from those surveyed in this paper. What I mean by structural will become clear later.

mental variable (IV) techniques or OLS regressions augmented with an observable measure of market ability. However, the structural literature concerned with human capital accumulation and schooling has expanded rapidly in the past 10 years. It now offers a different perspective on the issues surrounding the measurement of the returns to schooling and experience. For this reason, it deserves some attention.

This paper contains a survey of the recent literature in which the returns to both schooling and post-schooling human capital investments (including labor market experience) are estimated within a structural econometric framework. Structural stochastic dynamic programming (SSDP) models are based on the fundamental idea that agents are forward looking. In the structural approach, the estimated parameters are also those used to solve the agent's optimization problem and therefore provide a clearer connection between economic theory and the data than do reduced-form estimates. The implementation of a structural model therefore requires two fundamental steps; the solution of the dynamic programming problem (for a given set of parameters) and an iterative optimization procedure over the relevant parameter space in order to maximize an objective function (or minimize a distance).[2]

As of now, SSDP estimation techniques have been applied to several areas of labor economics, including job search, retirement decisions, fertility decisions, intertemporal labor supply and human capital accumulation behavior. Within a human capital accumulation framework, SSDP estimation techniques, along with a rich heterogeneity specification, may provide a transparent illustration of several dynamic self-selection issues central to the analysis of education and skill formation policies.

The range of applications related to human capital accumulation is large; it includes occupation choices (Keane and Wolpin, 1997), job search and endogenous experience accumulation (Wolpin, 1992), the effect of cohort size on wages (Lee, 2002), school drop-out behavior (Eckstein and Wolpin, 1999), the intergenerational education correlation (Belzil and Hansen, 2003), training and occupation choices of immigrants (Cohen-Goldner and Eckstein, 2003 and 2004), the relation between risk aversion and schooling (Belzil and

[2]Within a structural framework, it is also easy to simulate counterfactual policy experiments and therefore to evaluate relevant policies. Eckstein and Wolpin (1989) and Rust (1997) present comprehensive surveys of this literature as well as the solution and estimation techniques involved.

Hansen, 2004), education financing (Sauer, 2003, and Keane and Wolpin,1999), cohort effects in schooling attainments (Magnac and Thesmar, 2002b), racial differences in schooling attainments (Cameron and Heckman, 2001), and post-schooling human capital accumulation and wage inequality (Heckman, Lochner and Taber, 1998). As well, structural methods may prove to be a very powerful and flexible method of estimating the returns to schooling and to evaluate the ability bias (Belzil and Hansen, 2002). Within a correlated random coefficient wage regression model, structural methods may also be used to solve the OLS/IV puzzle often encountered in the literature (Belzil and Hansen, 2005).

In this paper, I survey papers in which forward looking agents make schooling and/or other human capital accumulation decisions based on a set of environmental parameters which include the return to schooling, the return to labor market experience and other surrounding parameters. Although the comparison with reduced form estimates will turn out to be a useful exercise, it is not my objective to survey the existing IV literature because there already exist comprehensive surveys.[3] At this stage, it is only important to note that estimates reported in the IV literature are typically high (estimates lying between 10% and 15% per year are often reported for the US) and that there is a surprising discrepancy between these estimates and those, much lower, obtained in a structural framework.[4] This discrepancy is difficult to reconcile for three main reasons. First, neither estimation method is nested within the opposite one. Second, each estimation method has its own weaknesses and, in particular, they both require strong behavioral assumptions. Finally, these diverging estimates are obtained even when comparable (if not identical) data sets are used.

Apart from assuming forward looking behavior, the papers surveyed herein share most (if not all) of the four following aspects. First and foremost, in all of them, the wage return to schooling is a central parameter. That is the case even when estimating the return to schooling is not the main objective of the author(s). Second, in almost all of the papers, labor market wages

[3] Card (1999 and 2001) presents an in-depth survey of the IV literature devoted to the returns to schooling. Rosezweig and Wolpin (2000) present a survey which is more critical of the economic literature using natural experiments. They discuss the returns to schooling and the ability bias as a special case.

[4] See Card (2001) for a discussion of several point estimates reported in the experimental literature.

are modeled within a Mincerian framework and the parameters are therefore given a skill formation interpretation. Third, in many of the papers, there is a rich a multi-dimensional unobserved (to the econometrician) heterogeneity specification. The relatively high dimensionality of the heterogeneity specification reflects the large number of states considered and is particularly useful to illustrate the mechanics of the ability bias. Finally, in almost all the papers surveyed, statistical inference is made using panel data.

The use of panel data is not innocuous. In practice, this means that in most of the papers surveyed, the authors maximize the joint likelihood of schooling attainments, employment decisions and wages observed over the lifecycle. In some cases, the likelihood may be simulated. As it will be clear later, the allowance for multidimensional unobserved heterogeneity (present in many papers) will imply that the objective function is a mixed likelihood, in the spirit of Heckman and Singer (1984). The use of mixed likelihood techniques reflects the desire to reach a level of flexibility as close as possible to a pure non-parametric methodology.

At the outset, and strictly speaking, it must be clear that there is no such thing as "structural estimate of the returns to schooling". Structural estimation does not identify new parameters of the Mincerian wage equation. In this survey, I nevertheless use the term "structural estimates" to refer to the effect of schooling on wages obtained from econometric models where schooling is endogenous and where the optimal level of schooling is the solution to an intertemporal model. The term "reduced-form" will be reserved for those estimates obtained in the IV literature (sometimes referred to as the experimental literature). The three terms (reduced-form, IV and experimental) may therefore be used interchangeably.

As the focus of this survey was on the return to schooling and experience as well as the ability bias, I shall concentrate on these parameters. That is not to say that the contribution of the structural approach is confined to what follows. Obviously, the structural approach identifies parameters which are not identifiable in the experimental literature. There are a wide range of issues such as the effects of liquidity constraints on schooling enrollment, the effect of risk aversion on schooling decisions or the effects of counterfactual tuition policy changes that are examined typically within a structural framework. These topics are interesting in their own right. However, estimates of the returns to schooling and experience are found in both structural and experimental papers. Their comparison is therefore particularly enlightening.

The main objectives of the paper are (i) to build a simple theoretical model of lifecycle time allocation between human capital accumulation, search and employment activities that illustrates the nature of the structural approach, ii) to summarize the main estimates found in the literature, (iii) to identify the key dimensions of the structural literature that distinguishes it from the experimental literature , (iv) to provide a comparison between the structural estimates and those reported in the experimental literature and, finally,(v) to comprehend the source of the discrepancy (ies).

The paper is structured as follows. In section 2, I survey some of the earlier literature on earnings and schooling. In Section 3, I present a theoretical model that unifies the literature surveyed herein. In Section 4, I discuss the key distinctions between the structural approach and experimental approach. In the following section, I briefly review some of the key papers that have estimated the returns to schooling and experience in a structural framework. In Section 6, I present a brief comparison of the structural estimates of the returns to schooling and experience and their reduced-form counterparts. Other key findings of the structural literature are discussed in Section 7. In section 8, I identify the main reasons why structural and reduced-form estimates diverge systematically. Finally, in conclusion, I identify some interesting topics of research where the structural approach may prove to be useful.

2 The Earlier Literature on Earnings and Schooling: From Mincer to the Experimental Literature

Before introducing the theoretical model and discussing the key distinctions between the structural literature and the experimental literature, it is useful to survey some historical aspects of the literature on earnings and schooling. As a way of introducing the structural approach, it will also be informative to discuss some of the weaknesses of the experimental literature.

Obviously, the literature devoted to the return to schooling can hardly be distinguished from the Mincerian wage equation. For the sake of the presentation, I consider two distinct periods. First, I review briefly the creation and the development of the Mincerian wage regression model. A second period covers a large number of empirical studies that used Instrumental Variable (IV) techniques to infer the returns to schooling. These papers belong to an even larger set of empirical research using the "Natural Experiment" approach.

2.1 The Mincerian Wage Function

The correlation between education and wages has been analyzed in seminal pieces by Becker (1964, 1967) and Mincer (1958, 1974). However, the first formal representation of the human capital accumulation process as an intertemporal optimization problem is due to Ben-Porath (1967). In this strand of the literature, it is customary to assume that wages are driven by competitive forces. If one assumes that individuals are paid their marginal product (according to their individual specific level of skills), offered wages reflect the spot market value (the rental price) of a unit of skill multiplied by the total stock of accumulated skills. That is

$$W_t = P_t \cdot K_t \tag{1}$$

where W_t is labor market wage, P_t is the rental price of a skill and K_t represents the human capital (total number of skills). Human capital accumulation (or skill acquisition) is rendered possible by combining inputs such as innate ability and time spent in school or time spent in the labor market. It is the representation of the relationship between accumulated skills (as an output) and education, experience and abilities (as inputs) that gave rise to the celebrated Mincerian wage regression;

$$\log W_t = w_t = \varphi_0 + \varphi_1(S_t) + \varphi_2(Exp_t) + \varepsilon_t^K \tag{2}$$

where $\varphi_1(S_t)$ represents the effects of schooling, $\varphi_2(Exp_t)$ represents the effects of experience and ε_t^K may be seen as an idiosyncratic productivity shock. Note that $\varphi_2(Exp)$ is typically defined for an exogenous post-schooling human capital accumulation rate and assumed to be concave in accumulated years of labor market experience.

2.2 The Experimental Literature

By the early 1970's, the estimation of the returns to schooling using Mincerian wage regressions had become one of the most widely analyzed topics in applied econometrics. In his survey of the earlier literature, Griliches (1977) pointed out several econometric problems that arise in estimating the returns to schooling and, in particular, those pertaining to the measurement of both schooling and ability. Until then, substantial effort had been devoted to the estimation of the return to schooling with control variables measuring (or proxies of) unobserved ability. These measures are typically IQ scores or scores of a similar nature, such as from the Armed Forces Qualifications (AFQT) Test. These are most likely imperfect measures of labor market skills. Casual empiricism reveals that AFQT scores (available in the popular National Longitudinal Survey of Youth) are more strongly correlated with schooling attainments than they are with wages. [5]

More interestingly, Griliches recognized that the endogeneity of schooling decisions, virtually ignored until then, was a serious issue which might have prevented economists from uncovering the true causal effect of education on earnings. Subsequently, a wide range of empirical papers using instrumental variable (IV) techniques have been published. A large segment of this literature is based on "institutional features" of the education system. Card (1999 and 2001) presents an extensive survey of this literature and discusses the main conceptual issues within a unifying theoretical structure in which individuals compare the benefits of schooling with the costs of schooling born early in the life cycle.[6]

At the econometric level, the main issue may be illustrated by the following cross-sectional regression function, which is a simplified version of the Mincerian wage offer equation,

$$w_i = \beta_0 + \beta_1 \cdot Schooling_i + \eta_i \tag{3}$$

Ignoring post-schooling labor market experience, it is clear that the discrepancy between OLS and IV estimates is a reflection of the correlation between schooling and unobserved ability (η_i). A positive (negative) correlation is associated with a positive (negative) ability bias. In the IV literature, the

[5]See Belzil and Hansen (2003).

[6]The model is intertemporal but is non-stochastic. It also does not allow for forward looking behavior beyond school completion. It borrows from Becker (1967).

ability bias is only indirectly investigated through the discrepancy between IV and OLS estimates, assuming that the linear model is correct. As we will see later, in structural models, the ability bias may be evaluated directly. With estimates of the utility of attending school and market ability, it may be obtained by simulation methods.

While a positive correlation between schooling and labor market ability is usually expected, Card (1999, 2001) reports that a large number of studies find that IV estimates exceed OLS estimates by a wide margin. For instance, IV estimates lying between 10% and 15% per year are often reported for the US. While estimates obtained for other countries may be lower, the tendency to obtain IV estimates which exceed their OLS equivalent has been observed in many different data sets and for many different countries.

These results have led economists to search for potential explanations. So far, two main explanations have been put forward for the OLS/IV discrepancy. The first one is the existence of potential (significant) measurement error in schooling. However, the measurement error argument often advanced is typically set within a classical measurement error framework which ignores the correlation between schooling levels and the measurement error itself and also ignores the discrete nature of the schooling variable. A second set of explanations relates to potential heterogeneity in the returns to schooling. As recognized in the experimental literature, in presence of heterogeneity in the returns, the IV estimate is inconsistent for the population average. Indeed, there is a large econometric literature concerned with the interpretation of IV estimates when the slopes are individual specific. In the context of a random coefficient model, the IV estimator is sometimes referred to as a Local Average Treatment Effect (LATE).[7] The LATE should be understood as a measure of the returns to schooling for the sub-population affected by the experiment (the instrument). It is often postulated that the high returns are explained by the fact that those individuals more likely to react to an exogenous policy change are those who are at the margin of deciding to enter college before the policy change and that they have higher returns to schooling than average. Typically, the IV estimate represents an average value for a sub-population which has been influenced by the instrument, but the density of this sub-population is not identified.

As we will see later, the "marginal" interpretation often given to the

[7] See Imbens and Angrist (1994).

LATE estimator is typically attached to a static interpretation of most natural experiments. Indeed, within a dynamic context, it is no longer possible to impute individual reactions to policy changes only to those who are at the margin of deciding to continue beyond a specific grade level (say college). The structural approach may help shed light on this rather fundamental issue.

After this brief historical review, I turn to the core of the paper. In particular, I will now present a simple life cycle model of human capital, search and work which will illustrate some of the key dimensions of the structural approach.

3 A Stochastic intertemporal Model of Human Capital Accumulation, Job Search and Employment

In this section, I introduce the theoretical model which illustrates most of the key aspects of the structural literature. The model stresses choices between four main states: schooling, home production, search (while unemployed and while employed) and labor market work. It focusses on the fact that human capital accumulation continues beyond school completion and, therefore, it differs from the theoretical models found in Card (1999, 2001) and Becker (1967), which are commonly used as a theoretical guideline for interpreting the existing empirical work based on IV techniques. I ignore issues related to school financing and liquidity constraints partly because the key issues related to the estimation of the returns to schooling and experience do not necessitate their introduction and partly because empirical evidence seems to point to the fact that liquidity constraints are not really important.[8]

3.1 The Model

Consider a population of homogenous individuals who maximize discounted expected lifetime utility over a finite horizon T by making sequential decisions over their time allocation. Each time period is indivisible. There are four main activities. schooling (S) , home production (H), unemployed search (U) and labor market work, which is itself divided in three mutually exclusive activities which may potentially magnify earnings; full time work which implies on-the-job learning (WL), work and on-the-job Training (WT) and work and on-the-job search (WS).

3.1.1 Income and Costs

The per-period utility, denoted $U(.)$, is function of the state specific net earnings. Individuals maximize expected discounted lifetime utility (the discount

[8] In the literature, this is sometimes justified by the strong empirical correlation between schooling and parents' education (long run factors) and the weaker correlation between schooling and family income. For a discussion, see Keane and Wolpin (1999) and Cameron and Heckman (1998 and 2001). Education financing is analyzed in Sauer (2004).

factor is denoted β). The net earnings are equal to the difference between the gross earnings (ξ_t^k) and the monetary equivalent costs (ω_t^k) of choosing each particular option, where k is a state specific indicator. Without loss of generality, I assume that the $\xi_t^{k\prime}s$ are non-stochastic but that the $\omega_t^{k\prime}s$ are affected by a random component. That is

$$\omega_t^k = \bar{\omega}^k + \varepsilon_t^k \tag{4}$$

where the $\varepsilon_t^{k\prime}s$ are i.i.d. shocks. For simplicity, I assume that the mean state specific costs are time invariant. The assumption that the gross earnings are non-stochastic is also innocuous since, in most data sets, only net earnings may be identified.

The control variable, d_{kt}, is equal to 1 when state k (S, H, U, WL, WT or WS) is chosen and 0 if not. The state specific earnings and costs are described as follows.

- When in school (when $k = S$), ξ_t^S may be interpreted as an exogenous transfer provided by the parents. I assume that attending school incurs a psychic cost (a disutility). Its monetary equivalent, ω_t^S, should be inversely related to school ability (or motivation)

- When at home (when $k = H$), ξ_t^H represents the monetary value of home production.

- When unemployed and searching full-time (when $k = U$), ξ_t^U refers to unemployment compensation and ω^U to the monetary value of all direct and psychic costs of job search.

- When at work (when $k = WL, WT$ or WS), the relevant income is the labor market wage, denoted W, minus the monetary equivalent costs of choosing the particular state, $\omega_t^{(.)}$. First, ω_t^{WT} is the monetary equivalent of the disutility of being trained. The term ω_t^{WS} may be viewed as the cost of search activities while employed. Finally, I assume that learning on the job is costless and set ω_t^{WL} to 0.

- The stock of human capital obeys the following laws of motion,

$$S_t = \sum_{s=1}^{t-1} d_{Ss} \tag{5}$$

$$WS_t = \sum_{s=1}^{t-1} d_{WS,s} \tag{6}$$

$$WL_t = \sum_{s=1}^{t-1} d_{WL,s} \tag{7}$$

$$WT_t = \sum_{s=1}^{t-1} d_{WT,s}. \tag{8}$$

$$H_t = \sum_{s=1}^{t-1} d_{Hs} \tag{9}$$

and it is convenient to summarize the stock of post schooling human capital into an "Experience" vector (Exp_t). That is

$$Exp_t = (WS_t, WL_t, WT_t, H_t) \tag{10}$$

3.1.2 Labor Market Productivity

The individual productivity (the marginal product) is reflected in the quantity of skills, K_t, accumulated by date t. As is the case in the human capital literature, K_t is connected to various inputs (education and post schooling activities) through a production function (Ben-Porath, 1967 and Mincer, 1974). I assume that

$$\log K_t = \varphi(S_t, Exp_t, TR_t) = \varphi_0 + \varphi_1(S_t) + \varphi_2(Exp_t) + \varepsilon_t^K \tag{11}$$

where ε_t^K is a purely random productivity shock reflecting idiosyncratic movements in productivity, health or any other determinant of individual productivity and where Exp_t may contain all possible post schooling activities . As such, the form of the skill production function is general enough to encompass possible non-linearities in schooling and post-schooling human capital but it imposes a form of separability between schooling and post-schooling human capital. For simplicity, I assume that learning takes place even in periods of on-the-job training and job search. This may obviously be justified by the fact that most panel data sets are based on interval of one year and that periods of on-the-job training are typically short. Given time spent in the market, accumulated periods of training raise expected wages in excess of the return to learning.

3.1.3 The Structure of the Labor Market: Introducing Search Frictions

In a model of human capital accumulation, the connection between labor market skills (K_t) and wages (W_t) is central but it may differ substantially according to the nature of the labor market. While the competitive market assumption is prevalent in the applied literature, I will now allow for the existence of labor market frictions.

At the outset, I assume that firms post wages and that workers search for the best offers (wages) from an exogenously determined wage offer distribution[9]. To introduce true wage dispersion, it is necessary to distinguish between purely random productivity shocks (ε_t^K) and firm specific effects. It will also be useful to distinguish between the current wage, W_t, and the potential wage offer, W_{tj}^o, where j is the firm indicator.

- More precisely, I assume that the logarithm of the offered wage (w_{tj}^o) is represented by

$$w_{tj}^o = \log K_t + \varepsilon_j^w \tag{12}$$

  where ε_j^w is the mean 0 firm specific term representing wage dispersion. In a search context, it is clearly possible to remove the random productivity shock and assume that worker productivity is non-stochastic. To avoid the confusion between observed wages and wage offers (productivity), I refer to the density of wage offers as $\phi^O(.)$.

- While involved in unemployed search ($k = U$), the individual receive at most one offer with probability θ^U and no offer with probability (1-θ^U).

- While involved in on-the-job search ($k = WS$), the individual receive at most one offer with probability θ^W and no offer with probability (1-θ^W).

- To be general, I assume that the job offer arrival rate may depend on accumulated human capital and skills; that is[10]

$$\theta_t^k = \theta^k(S_t, Exp_t) \text{ for } k = U \text{ and } WS \tag{13}$$

[9] In Wolpin 2003, skill formation policies are discussed within an equilibrium search model (with endogenous wage dispersion).

[10] I do not provide a formal argument for the possibility that the offer probability may depend on accumulated human capital and skills.

3.2 The Value Functions and the Reservation Wages

To characterize the sequence of optimal decisions taken by an individual, it is useful to define the relevant information set, $\Omega(t)$, as follows

$$\Omega(t) = (\xi_t^{(k)}, \varepsilon_t^K, \omega_t^{(k)}, W_t, W_t^o, S_t,\ Exp_t) \tag{14}$$

As is standard in the literature, I assume that the $\varepsilon_t^{(k)\prime}s$, ε_t^K and ε_j^w(when an offer is received) are known when the decision is made (at the beginning of period t) but that the future values are unknown (although their distribution is known). Of course, their past values are also known but they are irrelevant when the decision is made at t. Once a wage offer from firm j is accepted, ε_j^w is known for all future periods. The optimal choices are summarized in the various state specific value functions. Following Bellman (1957), these value functions are generally written as

$$V_t^k(\Omega_t) = U(\xi_t^k - \omega_t^k) + \beta EV_{t+1}(\Omega_{t+1} \mid d_{kt} = 1) \tag{15}$$

where

$$EV_{t+1} = E\max\{V_{t+1}^S, V_{t+1}^H, V_{t+1}^U, V_{t+1}^{WL}, V_{t+1}^{WT}, V_{t+1}^{WS}) \tag{16}$$

and where the expected value is a taken with respect to the future $\varepsilon_t^{(k)\prime}s$ and ε_t^w.

For each possible state, the dependence of Ω_{t+1} on $d_{kt} = 1$ illustrates how human capital is accumulated.

- **School and Home Production:** For the value functions of school attendance, the dependence of Ω_{t+1} on $d_{kt} = 1$ reflects the increment in K_t induced by an increase in one unit (one year) of schooling. For home production, it may reflect a reduction if, for instance, labor market skills depreciate during inactivity.

- **Labor Market work and Search**: When the individual chooses on-the-job search, the dependence of Ω_{t+1} on $d_{WSt} = 1$ reflects both the possibility of receiving no acceptable offer as well as the expected utility of working at a new wage next period.[11] The dependence of Ω_{t+1}

[11] The particular form of the value function for search problems are relatively well known and are surveyed in Eckstein and van den Berg (2003).

on $d_{WTt} = 1$ reflects the increase in skill (and potential wages) caused by an increase in both Exp_t and TR_t. Finally, the dependence of Ω_{t+1} on $d_{WLt} = 1$ reflects the increase in skills caused by accumulated experience (learning).

It is now straightforward to define the relevant reservation wages. Given the allowance for both employed and unemployed search, there will be two reservation wages. One will dictate how individuals move from one job to the next and another one will be central to explain movements from the non-productive states (school, home production and search) to labor market work (post-schooling human capital accumulation).

Obviously, for an individual holding a job with wage rate W_t, any outside job offer paying a wage rate W_t^o exceeding the current wage will be acceptable. The acceptance probability $(\Pr W_t^o > W_t)$ is denoted $\Pi_t(.)$.

However, the most interesting reservation wage is the one that will dictate the decision to join the labor market or not. At the empirical level, this reservation wage becomes the key threshold value which regulates the observability of wage offers to the econometrician. To define it, let's consider the state specific wage offers that equate the value of each non-work states (Schooling, home production and unemployment) to the minimum of the different value functions of labor market work and denote these wages W_t^S, W_t^H and W_t^U. These values, W_t^S, W_t^H and W_t^U are such that

$$\begin{aligned} V_t^S(.) &= Min\{V_t^{WL}(W_t^S), V_t^{WT}(W_t^S), V_t^{WS}(W_t^S)\} \\ V_t^H(.) &= Min\{V_t^{WL}(W_t^H), V_t^{WT}(W_t^H), V_t^{WS}(W_t^H)\} \\ V_t^U(.) &= Min\{V_t^{WL}(W_t^U), V_t^{WT}(W_t^U), V_t^{WS}(W_t^U)\} \end{aligned} \tag{17}$$

Let's denote the minimum wage acceptable by W_t^*, then clearly

$$W_t^* = Max\{W_t^S, W_t^H, W_t^U) \tag{18}$$

At any period t, an employed individual must earn a wage which is at least as large as W_t^*. In turn, using this reservation wage, it is easy to derive the wage distribution observed at any point in time. The distribution of observed wages, $\phi_t(W_t)$, is given by

$$\phi_t(W_t) = \frac{\phi_t^o(W_t)}{\Pi(W_t^*)} \cdot I(W_t \geq W_t^*) \tag{19}$$

where $I(w_t \geq w_t^*)$ is the indicator function.

The distinction between $\phi_t^o(.)$ and $\phi_t(.)$ is not innocuous. In theory, it may be possible to learn about the return to schooling from either $\phi_t(.)$ or $\phi_t^o(.)$ since both distributions depend on accumulated schooling. However, the reader should remember that the effect of education on market productivity is contained in $\phi_t^o(.)$.

3.3 Choice Probabilities and Solution

The fundamental input required at the estimation step is an expression for the probabilistic choice of option k. In general terms, the probability that option k is chosen is simply

$$\Pr(d_{kt} = 1) = \Pr\{U^k(.) + \beta EV_{t+1}(. \mid d_{kt} = 1) \geq$$

$$U^s(.) + \beta EV_{t+1}(. \mid d_{st} = 1)\} \text{ for all } s \neq k \tag{20}$$

In the case where $U_t^k()$ is linear in the $\varepsilon_t^{k\prime}s$, this expression takes the form of a non-linear discrete choice expression and boils down to the probability that the ε_t^k exceeds a threshold value. A simple illustration is provided in Section 3.4.

However, even in such a case, the technical difficulties arise in the evaluation of the $EV_{t+1}(.)$. When the model is set in a finite horizon framework, the solution method will be based on recursive techniques (Bellman, 1957). In general, there are no closed-form expression for it, and numerical methods need to be used. There is indeed a relatively large literature devoted to the solution methods.[12] However, in specific cases, it may be expressed in closed form. This is the case, for instance, when the number of states is limited (Belzil and Hansen, 2002) or when the random terms follow an extreme value distribution (Rust, 1987).

[12] For a discussion, see Keane and Wolpin (1994), Hotz and Miller (1993) and Eckstein and Wolpin (1989,b).

3.4 Some Illustrative examples

In order to fix ideas, it may be useful to consider two specific examples. In the first one, I illustrate how the schooling decisions may be affected by the relative values of the return to schooling and experience as well as the discount rate. In the second one, I illustrate how inference about productivity growth may be confounded by the presence of search frictions.

For each of these examples, I ignore stochastic variation in productivity. For simplicity, assume that all the stochastic shocks affecting the cost of choosing each option are standard normal with CDF $\Phi(.)$ and density $\phi(.)$. Finally, suppose that agents maximize lifetime earnings ($U(\xi_t^{(k)}) = \xi_t^{(k)}$).

3.4.1 Example 1: The Determinants of Schooling Attendance

Assume that individuals live for 2 periods and that everyone works in period 2. In period 1, individuals choose only between School (S) and Work (WL). This means that labor market experience enhances wages only through learning (there is no formal training opportunities). Since search is not an option, assume that the market is competitive and normalize the skill price to 1. Then,

$$\log W_t = \log K_t$$

and, finally, assume that

$$\log K_t = \varphi_0 + \varphi_1 \cdot S_t + \varphi_2 \cdot WL_t$$

without loss of generality, assume that schooling in period 1 has no effect on wages.

The intertemporal problem, as solved in period 1, is the decision between going to school ($d_{S1} = 1$) and therefore work one period (in period 2) and to work two periods ($d_{S1} = 0$). First, note that the relevant value functions are

$$V_2^{WL}(S_2 = 1) = \exp(\varphi_0 + \varphi_1) \text{ and } V_2^{WL}(S_2 = 0) = \exp(\varphi_0 + \varphi_2)$$

$$V_1^{WL} = \exp(\varphi_0) \ + \beta \cdot \exp(\varphi_0 + \varphi_2)$$

$$V_1^S = -\omega^S - \varepsilon_2^S + \beta \cdot \exp(\varphi_0 + \varphi_1)$$

$$\begin{aligned} \Pr(d_{S1} &= 1) = \Pr(V_1^S > V_1^{WL}\ ) = \\ &\Phi(-\omega^S - \exp(\varphi_0) + \beta \cdot \exp(\varphi_0) \cdot (\exp(\varphi_1) - \exp(\varphi_2))) \end{aligned} \quad (21)$$

Within this simple model, (21) is just a particular case of equation (20).

As simple as this example may be, it is still a useful illustration of the key differences between structural models (in which agents are forward looking) and the more static representations of the schooling decision.

Evaluating how the school attendance probability changes with the key parameters, we obtain

$$\frac{\delta pr(d_{S1} = 1)}{\delta \varphi_0} = \phi(.) \cdot (\exp(\varphi_0)(\beta \cdot (\exp(\varphi_1) - \exp(\varphi_2)) - 1) \quad (22)$$

$$\frac{\delta pr(d_{S1} = 1)}{\delta \varphi_1} = \phi(.) \cdot (\beta \cdot \exp(\varphi_{0+}\varphi_1)) > 0 \quad (23)$$

$$\frac{\delta pr(d_{S1} = 1)}{\delta \varphi_2} = -\phi(.) \cdot (\beta \cdot \exp(\varphi_{0+}\varphi_2)) < 0 \quad (24)$$

$$\frac{\delta pr(d_{S1} = 1)}{\delta \beta} = \phi(.) \cdot \ \exp(\varphi_0) \cdot (\exp(\varphi_1) - \exp(\varphi_2)) \quad (25)$$

These expressions imply that school attendance increases with the return to schooling (23) and decreases with the return to labor market experience (24). The relationship between school attendance and market ability (the wage intercept), which is found in (22), is ambiguous. It depends on all the parameters. However, for realistic values of φ_1 and φ_2, it will be negative.[13] Interestingly, the relationship between school attendance and the discount factor is also ambiguous (23). It depends crucially on the relative values of the returns to schooling and experience. Results (24) and (25) constitute key differences between structural models in which agents are forward looking (such as those reviewed in this survey) and the models found in Card

[13] This will be especially true if both φ_1 and φ_2 are close to 0.

(1999, 2001) and Becker (1967), in which agents solved for the optimal time allocation over an infinite horizon and in which post-schooling wage growth is not modeled.[14] More precisely, these models do not take into account the lifecycle opportunity cost of schooling.

3.4.2 Example 2: Wage Growth and Search Friction

In this example, retain the two period structure and individuals either work and search (WS) and work and on-the-job training (WT). For the sake of the argument, every individual enters period 1 with a wage equal to W. The problem is now an arbitrage between investing in search activities (while enjoying wage growth) and investing in on the job training. With search frictions, I now assume that the potential wage offer,W_{tj}^o, (in logs) is given by

$$\log(W_{tj}^o) = w_{tj}^o = \log(K_t) + \varepsilon_j^w \tag{26}$$

where the distribution of ε_j^w is left unspecified and

$$\log K_t = \varphi_0 + \varphi_2 \cdot WS_t + \varphi_3 \cdot WT_t$$

The relevant value functions are

$$V_1^{WT}(w) = W - \omega^{WT} - \varepsilon_2^{WT} + \beta \cdot \exp(w + \varphi_3 \cdot w) \tag{27}$$

$$\begin{aligned} V_1^{WS}(w) &= W - \omega^{WS} - \varepsilon_2^{WS} + \\ &\beta[(1 - \theta_{WS} \cdot \Pi(W)) \cdot \exp(w + \varphi_2 \cdot w) + \theta_{WS} \cdot \Pi(W) \cdot \Xi(W, \varphi_2)] \end{aligned} \tag{28}$$

where $\Xi(W, \varphi_2) = E(W_{tj}^o \mid W_{tj}^o > \exp(w \cdot (1 + \varphi_2)))$.

[14] In other words, Card (1999, 2001) does not take into account the lifecycle opportunity cost of schooling.

Clearly,

$$\Pr(d_{WS,1}=1)=\Pr(V_2^{WS}>V_2^{WT})=\Phi(0.5\cdot h(W,\theta_{WS},\beta,\varphi_2,\varphi_3)) \quad (29)$$

where

$$\begin{aligned} h(w,\theta_{WS},\beta,\varphi_2,\varphi_3) &= \beta\cdot([(1-\theta_{WS}\cdot\Pi(W))\cdot\exp(w+\varphi_2\cdot w) \\ &\quad +\theta_{WS}\cdot\Xi(w,\varphi_2)]-\exp(w+\varphi_3\cdot w))-\omega) \quad (30)\end{aligned}$$

and where $\omega=\omega^{WT}-\omega^{WS}$ and $\varepsilon_2=\varepsilon_2^{WS}-\varepsilon_2^{WT}$.

More importantly, the model points out to the divergence between the average lifecycle wage gain and the lifecycle productivity growth. The average lifecycle wage gain (in logs), denoted $\Delta\ln W$ is given by the following expression;

$$\begin{aligned} E(\Delta\ln W) &= \Pr(d_{WS,1}=1)\cdot[\theta_{WS}\cdot\Pi(W)\cdot(\Xi_l(W,\varphi_2)-w) \quad (31) \\ +(1-\theta_{WS})\cdot\varphi_2]+\Pr(d_{WS,1} &= 1)\cdot\varphi_3\end{aligned}$$

where $\Xi_l(W,\varphi_2)=E(\log W_{tj}^o \mid W_{tj}^o>\exp(w\cdot(1+\varphi_2)))$.

Clearly, the difference in expected log wages for successful job seekers must exceed φ_2.

The life cycle productivity growth (also in logs), denoted $E(\Delta\ln K)$on the other hand, is given by

$$\begin{aligned} E(\Delta\ln K) &= \Phi(0.5\cdot h(w,\theta_{WS},\beta,\varphi_2,\varphi_3))\cdot\ \varphi_2 \quad (32) \\ &\quad (1-\Phi(0.5\cdot h(w,\theta_{WS},\beta,\varphi_2,\varphi_3))\cdot\varphi_3\end{aligned}$$

and it follows that $E(\Delta\ln W)>E(\Delta\ln K)$.

The relationship between wages (observed) and accumulated human capital is therefore difficult to interpret and, without a clear theoretical structure, the observed correlations between wages and different notions of human capital constitute dissolute pieces of information. While the private returns to schooling should incorporate the non-productivity based wage growth, policy interventions aimed at skill formation must still be based on real productivity enhancement, not on actual wage gains. It is also easy to see that the complications arising in the presence of search frictions also transmit to the estimation of the return to schooling. In the hypothetical case where job offer rates are function of education level, the return to education may also be confounded by the effect of a higher rate of mobility.

3.5 Heterogeneity, Treatment Effect and Dynamic Self-Selection

Up to now, the theoretical structure focused on a single individual. This implies that individual differences in school and labor market histories would be attributable solely to differences in the pure stochastic choices experienced. In the real world, individual differences in human capital accumulation may also (perhaps mainly) reflect individual differences in tastes and abilities. To be realistic, the model must therefore allow for dynamic-self selection.

3.5.1 Introducing unobserved heterogeneity:

Population heterogeneity may be introduced by assuming that each individual is endowed with a vector of costs $\bar{\omega}_i^k$ and market ability(η_i). The vector $\{\bar{\omega}_i^k, \eta_i\}$ is a representation of the individual specific initial endowments in taste and abilities existing before the start of the human capital accumulation process. In particular, labor market ability should be understood as affecting labor market wages, even after controlling for acquired human capital. These abilities and tastes are permanent and, moreover, are typically unobserved to the econometrician. However, they are assumed to be known by the agent. In some data sets, there may exist some variables which are potentially good approximations of the inherent abilities. For instance, in the National Longitudinal survey of Youth (NLSY), some individuals were administered the Armed Force Qualification Tests (AFQT) and individual scores are therefore available.[15]

In the case where such variables are not available or, at least, are only partial measures of the relevant tastes and abilities, it will be important to

[15] The AFQT is a test that measures basic quantitative and analytical skills. However, this test is administered while individuals are in their early teenage years and individual differences may sometimes reflect differences in schooling attainment that may have already taken place. For this reason, AFQT scores are often corrected before the estimation procedure.

allow for permanent unobserved heterogeneity. I assume that

$$\{\bar{\omega}_i^k, \eta_i\} \sim \ G(.) \tag{33}$$

where $G(.)$ is the cumulative distribution function. In empirical work, the correlations implied by $G(.)$ will be fundamental in order to illustrate the ability bias, or other issues. For instance, knowing $G(.)$ will reveal if those who have high labor market ability tend to be also endowed with high value of attending school, a lower psychic costs of training or a low cost of search.

Given the allowance for population heterogeneity, the amount of skills owned by individual i at date t (and therefore the mean wage offer) has to be reinterpreted. In the most general case possible, all components of the human capital production function are potentially affected by unobserved market ability. That is

$$E(\ln K_{it}) = E(w_{it}^o) = \varphi_1(S_{it}, \eta_i) + \varphi_2(Exp_{it}, \eta_i) \tag{34}$$

When the wage offer function is linear in market ability and schooling, an increase in labor market ability (η),for a fixed level of school ability (ω_i^S), may act as a disincentive to stay in school. At the same time, for a fixed level of market ability, an increase in school ability will raise schooling. It is the possible correlation (positive) between η and and the utility of attending school $(U(\xi_t^S - \omega_t^S))$ which led many economists to conjecture that the correlation between schooling and market ability may be positive. In the case where $\varphi_1(S_t)$ is linear in schooling and φ_2 is ignored, we obtain the classical ability bias problem considered by Griliches (1977).

When the regression function is non-linear in ability, we typically obtain a Correlated Random Coefficient Wage Regression Model, which has substantially gained in popularity in recent years.[16] Indeed, estimating models with heterogenous (individual specific) slopes has attracted a substantial amount of attention within the last 15 years. As we will see in several specific cases, SSDP techniques represent an alternative to various methods popularized in the 'Treatment Effects' literature.[17]

[16] See Heckman and Vytlacil (1998) and Imbens and Angrist (1994) for a discussion.

[17] This will be clear when reviewing Cohen Goldner and Eckstein (2002, 2004) and Belzil and Hansen (2005).

3.5.2 The Structure of the Labor Market: General vs Partial Equilibrium

In the literature, the partial equilibrium approach has been favored in most empirical applications. To understand the key implications of the choice between a partial and a general equilibrium setting, it is useful to re-examine (1). In a partial equilibrium setting, the skill unit price is subsumed in the intercept term. In a general equilibrium framework, changes in wages are decomposed into changes in skill price and changes in skill levels. The change in skill price is typically identified though movements in aggregate measures of some of (or all) the inputs relevant in the production process. These may include changes in cohort sizes or changes in physical capital. Until now, the authors concerned with general equilibrium inference have only considered time series changes in skill prices but have always implicitly assumed stationarity in the skill production function.[18]

3.6 Identification and Estimation

In the literature, identification may only be considered within a particular parametric structure.[19] Structural models require to specify preferences and technology and therefore imply functional forms and parametric assumptions. Moreover, estimating a structural model with unobserved heterogeneity requires to distinguish between the distribution of random shocks and the distribution of unobserved heterogeneity.

Estimation is usually performed using maximum likelihood techniques or their simulated counterparts. This implies that, in most of the cases, estimation requires solving a number of moment conditions equal to the number of parameters. Obviously, estimation requires identifiability of the particular parametric structure. Typically, as most complicated identifiable non-linear models, structural models are locally identified. However, in the case where a specific model is estimated at a low computation time cost,

[18] To my knowledge, the time series properties of the human capital production function has not been investigated in conjunction with the well documented recent increase in wage inequality. For a discussion of several empirical issues in the wage inequality literature, see Lemieux (2004).

[19] The degree of under-identification (non-parametric) found in empirical dynamic programming models is discussed in Rust (1994) and Magnac and Thesmar (2002).

it is sometimes possible to search a larger parameter space for other local maxima.

Basically, the estimation step requires forming the joint probability of the individual specific choices and market wages (when observed) over the life cycle. For a given individual i, the joint probability is the product of all period specific probabilities; that is

$$\Pr\{(d_{ik1}, w_{i1}), ...(d_{ikT}, w_{iT})\} = \Pr(d_{ik1}, w_{i1}) \cdot ..... \Pr(d_{ikT}, w_{iT}) \tag{35}$$

where the particular form of $\Pr(d_{ikt}, w_{it})$ will differ according to the specificity of the model considered. In particular, the allowance for search frictions will imply a quite different expression than the one implied by a standard competitive labor market equilibrium.[20]

In all models allowing for unobserved heterogeneity, it is assumed that individual unobserved (to the econometrician) skills are fully known by the optimizing agent. Admittedly, this is a strong behavioral assumption. This implies forming the following mixed likelihood function:

$$L(.) = \int \Pi_{t=1}^{T} \Pr(d_{kt}, W_t \mid \bar{\omega}_i^k, \eta_i) dG(\bar{\omega}_i^k, \eta_i) \tag{36}$$

where $\theta(.)$ is a set of parameters to be estimated. In practice, (36) is maximized with respect to a discrete approximation of $G(.)$ with a fixed (known) number, say M, of types. This means that each type is endowed with a specific vector of endowments $(\bar{\omega}_m^k, \eta_m)$ and the integral is then replaced by a discrete sum such as; [21]

$$L(.) = \sum_{m=1}^{M} \Pi_{t=1}^{T} \Pr(d_{kt}, W_t \mid type\ m). \Pr(type\ m) \tag{37}$$

where $\Pr(type\ m)$ refers to the population proportion of individuals belonging to a particular type m.

[20]When modeling search activities, it may turn out to be preferable to use shorter time periods than a one year interval. For a specific treatment of search activities and related estimation issues, see Ecsktein and van den Berg (2003).

[21]This approach is in the spirit of Heckman and Singer (1984).

4 What Distinguishes the Structural Literature from the IV Literature?

Before reviewing specific contributions, it is informative to point out at least three main attributes of the structural literature that distinguish it from the natural experiment literature. These dimensions form the basis of the "comparative advantages" of the structural approach and, moreover, represent the key to understanding the reasons why structural estimates and reduced-form estimates of the return to schooling diverge as much.[22] When trying to comprehend these discrepancies (in Section 7), it will be useful to organize the discussion around these points.

First, by imposing an econometric structure in which young individuals make schooling decisions based on a set of parameters which includes their individual specific market ability(ies), structural models offer an explicit specification of the selectivity process or, at least, they provide the opportunities to quantify the importance of selectivity through simulation methods. Put differently, in structural models, the individual specific ability term(s) inferred from lifecycle (panel) data on post-schooling wages is (are) forced to be an input in the maximization problem solved by the agent. As seen earlier, when market ability enters the wage equation additively and when individuals maximize lifetime income, the resulting selectivity coincides with the celebrated notion of "ability bias". In the structural literature, all relevant correlations are identified (if identifiable) and are therefore computable. In the reduced-form literature, they may only be inferred from the OLS/IV difference, provided that the linearity assumption is valid. Obviously, this particular aspect of the structural literature comes at the cost of a strong behavioral assumption; namely that all unobserved (to the econometrician) tastes and abilities are known to the agent at the beginning of the optimization process.

Second, in many papers surveyed herein, estimates of the returns to schooling (and experience) are obtained from a more flexible specification of the Mincerian wage regression than the one typically found in the experimental literature. In some cases, this flexibility refers to the possibility that, for a given level of ability, the local returns to schooling vary with

[22] The advantages of the IV (or the Natural Experiment approach) are discussed extensively in the literature (see Card, 1999 and 2001).

grade level.[23] Obviously, this implies that the average return to schooling also varies with grade level. In other cases, it translates into the allowance for individual (or occupation) specific returns. In the experimental literature based on IV techniques, the effect of schooling is linear by construction and is summarized in a single parameter estimate.[24]

Finally, in the structural literature, there is a more realistic treatment of labor market experience as a substitute for time spent in school in order to achieve higher life cycle wages. The relative values of the return to schooling and the return to experience will play a key role in schooling decisions. In the case where the slopes of the Mincerian wage equation are allowed to be individual specific, the degree of self-selectivity is also directly linked with both the returns to schooling and experience.[25] In the experimental literature, the strong behavioral assumptions regarding post schooling optimization behavior constitutes one of its most definite weaknesses. While most authors using IV techniques based their estimation strategy on finding exogenous events (natural experiments) which are orthogonal to unobserved ability, they rarely take into account that individuals keep optimizing beyond school completion. This is surprising as labor market experience (including learning on-the-job, training or job search) represents a key substitute to schooling as a mean for enhancing life cycle wages and as most researchers do not draw inference solely from entry wages. A survey of the IV literature reveals that practically no paper presents a joint estimation of the returns to schooling and experience, other than occasional inclusion of a control for age or for potential experience. Yet, without controls for individual differences in accumulated post-schooling human capital, it is difficult to give an interpretation to the discrepancy between OLS and IV estimates of the return to schooling.[26]

To see this, consider the following simple framework where the wage equation is

$$w_{it} = \beta_0 + \beta_1 \cdot S_{it} + \delta \cdot PSHC_{it} + \eta_i \tag{38}$$

[23] Obviously, this implies that the average return to schooling also varies with grade level.

[24] However, as will be seen later, the linearity assumption is virtually always rejected (Belzil and Hansen, 2002, 2003, and Belzil, 2005).

[25] This is the case in Belzil and Hansen (2004).

[26] Many issues related to the endogeneity of work experience are discussed in Rosenzweig and Wolpin (2000)

and where $PSHC_{it}$ denotes all cumulated post schooling human capital investment activities (training, search and learning) which have taken place by date t (when wages are measured).

To illustrate the nature of the models investigated by most of the authors in the experimental literature, suppose that you wish to estimate a simplified version of (38) using instrumental methods[27], namely

$$w_{it} = \beta_0 + \beta_1 \cdot S_{it} + \eta^*_{it} \tag{39}$$

where

$$\eta^*_{it} = \delta \cdot PSHC_{it} + \eta_{it}. \tag{40}$$

Assume that you are given access to an instrument Z_i. Typically, Z_i refers to the occurrence (or non occurrence) an event that took place before t, say at $t\prime$. The point is that the instrument comes into effect at time t'; that is each individual decides to attend (or continue) school or at t' based on the realized value of Z_i. For example, $Z_{i'}$ may be season of birth, distance to college or change in mandatory schooling age. Clearly, for this instrument to be a valid, it must be that

$$Corr(Z_i,\ \eta^*_{it}) = 0 \tag{41}$$

which is a stronger condition than $Corr(Z_i,\ \eta_i) = 0$. However, in general, (41) cannot be true since Z_i is typically correlated with $PSHC_{it}$. This may easily be demonstrated using technical arguments but it can be also illustrated with more intuitive arguments.

In technical terms and in the context of the model presented in Section 2, the issue is the following. First, consider a cross-section of wages measured at period $t = t' + s$, which represent one element of the state space at the beginning of period $t' + s$ (denoted $\Omega(t' + s)$). The non-orthogonality is explained by the fact that $\Omega(t' + s)$, is clearly affected by Z_i since the $d_{kt'}s$ are affected by Z_i (because Z_i is an instrument). Therefore, this also means that actions taken at any date $t' + s$ (the $d_{kt+s'}s$) are also potentially affected by Z_i. In a timely fashion, the order of causation is the following

$$Z_i(\text{at } t') \rightarrow d_{kt'} \rightarrow \Omega(t'+1) \rightarrow d_{kt'+1} \rightarrow ...\Omega(t'+s) \rightarrow d_{kt'+s} \tag{42}$$

[27]The argument will work as well if you add an additional control for potential experience or age (as is sometimes done) .

In most empirical applications, the researcher assumes that the effect of Z_i is limited to $d_{kt'}$ only. However, in the real world, η_{it}^* is affected by the entire history of $d_{kt}'s$ from $t\prime$ to $t+s$. This illustrates the differences in the implementation of an IV strategy between an intertemporal context where individuals optimize sequentially in every periods and a more standard (static) context. In a purely static problem with two endogenous variables and only one instrument available, it is possible to include one of the endogenous variables in the error term, as long as the instrument is orthogonal to the omitted endogenous variable. In a dynamic setting, it is the orthogonality between the instrument and second endogenous variable that is automatically violated.

An intuitive illustration may be provided in the context where the instrument refers to season of birth or differences in birth outcomes (Angrist and Krueger, 1991) or distance to college (Card, 1995). Clearly, an individual who loses one potential year of schooling may then react by investing in post schooling training, in search activities and/or in any other wage enhancing activities. Similarly, an able individual willing to maximize lifetime income and who is born (or raised) far away from the nearest college may also invest heavily in wage enhancing post-schooling activities.

Rosenzweig and Wolpin (2000) discuss similar issues in a variety of micro economic models which are inherently dynamic. Indeed, they claim that the natural experiment literature is plagued with the misuse of IV techniques. The arguments developed by Rosenzweig and Wolpin for each specific model are essentially imbedded in expression (42). These arguments are powerful and difficult to refute if one is assuming sequential optimizing behavior. As such, this criticism does not imply that experiments are irrelevant pieces of information but rather that the multiplicity of instruments required to estimate a sequential model may be a serious barrier to estimation.

To a large extent, it is possible to view the structural literature on the returns to schooling as an alternative estimation method which allows the researcher to get around many of the fundamental shortcomings and make progress on several points where the experimental approach may not be conclusive. Obviously, structural estimation has its own weaknesses. In particular, it requires parametric (distributional) assumptions which are not always needed in a reduced-form framework and it is also based on strong behavioral assumptions. However, it is the key differences between the structural and the experimental approach that allow the researchers to shed new perspectives on the estimation of the returns to schooling and experience and the sign of the Ability Bias. Hopefully, this will become clear in the following section, in which I survey some important contributions.

5 An Overview of the Structural Literature

In this section, I briefly summarize some of the most important papers found in the structural literature. The contributions may be classified according to the following aspects: the endogeneity of schooling, the explicit treatment of search frictions (whether the model is set in a competitive framework or in a search economy), the allowance for general equilibrium effects, the extent to which the specific model is structural (as opposed to semi-structural) and, finally, the number of skills rewarded in the labor market.

However, for the sake of the presentation, I have grouped the papers in the following three categories:

- Partial equilibrium Models
  - Models focussing on search friction
  - Models focussing on endogenous schooling and or endogenous experience
- General equilibrium models
- Semi-structural models

Before reviewing specific papers, I discuss briefly the basic aspects of the models in each category. Table 1 summarizes the aspects of the theoretical models in all papers reviewed below.

As in most microeconometric applications, the analysis of human capital accumulation may be modeled in general equilibrium or in a partial equilibrium approach. In the structural literature, the partial equilibrium approach is dominant. The first two papers (Wolpin, 1992 and Eckstein and Wolpin,1995) focus on post schooling human capital accumulation and job mobility conditional on schooling and are set within a partial equilibrium search framework.

The set of partial equilibrium models include not only all those papers mentioned in the above paragraph but a second subset of papers including Eckstein and Wolpin, 1989, Keane and Wolpin (1997), Cohen-Goldner and Eckstein (2003, 2004), Belzil and Hansen (2002, 2005), Eckstein and Wolpin (1999) and Todd and Wolpin (2003). These papers focus mainly on endogenous schooling and endogenous work experience accumulation.

In the structural literature, few general equilibrium models have been estimated. Indeed, the second group contains only two contributions (Heckman, Lochner and Taber, 1998, and Lee, 2004).

Finally, the last three papers (Cameron and Heckman, 2001, Magnac and Thesmar (2002,b). and Belzil, 2004) are semi-structural. They are set in a dynamic framework but they do not require solving a dynamic programming problem.

Although the presentation is not classified with respect to the treatment of skill heterogeneity, it is also appropriate to discuss some issues relating to the dimensionality of labor market skills. Although most models are set in a single skill framework, it is also possible to interpret the wage/schooling relationship in a multi-skill framework. In the literature, the extension to multiple skills is usually based on the distinction between labor market activities such as those varying by occupation or industrial sector. For instance, in Keane and Wolpin (1997) and in Cohen-Goldner and Eckstein (2003, 2004), skills are associated to occupations. Each skill has a rental price and its own production function. This approach is able to take into account that schooling (and perhaps training) is more valuable in some sectors than in others.[28]

The discussion of specific papers is structured as follows. I briefly discuss the structure of the model and the specification of the Mincerian wage regression (when applicable), the role of unobserved heterogeneity and, when applicable, the implications for ability bias. After a brief review of the estimation method, I stress the main results found in the paper.

[28] Multiple skill models are developed in Willis (1986) and Heckman and Sedlacek (1985).

5.1 Partial Equilibrium models with Search Frictions

5.1.1 Wolpin (1992): Job Search and Racial Differences in Earnings

Wolpin (1992) is concerned with the estimation of the return to endogenous work experience, conditional on schooling attainment. However, the model is set in search framework. The author studies the transition from high school to full time employment during the first five years following graduation and focuses on racial differences in job search outcomes. The model is quite general. It incorporates wage dispersion, layoffs and recall probabilities, unemployment insurance benefit and a differentiated effect of work experience on wage growth depending on general versus employer specific work experience.

The model is fit on a sample taken from the National Longitudinal Survey of Youth (NLSY). Wolpin focuses solely on high school graduates and assumes that, aside from racial differences, workers are homogeneous. Wolpin assumes that once an offer is accepted, the wage remains constant and is therefore not affected by a stochastic term (unless the individual is laid off and recalled). The law of motion that links endogenous work experience with wages belongs to the class of Mincerian wage regressions, that is

$$w = \varpi_{00} + \varpi_{01} \cdot SK + \varpi_{02} \cdot GK + \varpi_{03} \cdot SK^2 + \varpi_{04} \cdot GK^2 \tag{43}$$

where w is the log of quarterly earnings and SK and GK refer to specific and general accumulated work experience respectively.

The main findings are i)white high school graduates face a much higher wage dispersion that do blacks and whites therefore enjoy a much higher return to experience than blacks. ii) the probability of receiving a wage offer while employed and while unemployed are both higher for blacks than for whites iii) the value of non-market time (evaluated on a quarterly basis) is much higher for blacks than for whites. Perhaps the most original aspect of this paper is its reliance on the existence of search frictions to estimate the return to human capital (experience)

5.1.2 Eckstein and Wolpin (1995): The Return to Schooling in a Search/Matching Model

Eckstein and Wolpin (1995) are the first authors to attempt to estimate the returns to schooling within a search framework. As in EW (1989) and Wolpin (1992), education is assumed to be exogenously determined. However, the authors depart from the standard partial equilibrium search framework and estimate a search/bargaining model in the spirit of Diamond and Maskin (1979). In their model, firms and workers meet randomly (given a level of effort endogenously determined by both the firm and the worker) and sample one observation from a "match" distribution. The optimal decision rule is to search until a match value exceeding a reservation level is actually drawn. Using data on the duration to first job and accepted wages, and imposing further restrictions on the solution to the bargaining problem, the authors are able to recover some key parameters.[29]

The use data from the 1979 youth cohort of the NLSY. They perform separate estimation for blacks and whites, and distinguish between four schooling levels; high-school non-completers, high school graduates, college non-completers and college graduates. As exposed in (19), EW distinguish between offered wages and accepted wages when estimating the return to schooling. They argue that observed differences in mean accepted wages provide a distorted picture of the return to schooling since not all the firm-worker matches are accepted. To illustrate this, EW show the discrepancy between the returns to schooling measured from observed wages and those obtained from offered wages, in their sample, the differentials in mean accepted wages by schooling level ranges between 7% and 26%. Mean accepted wages are much higher than mean offered wages. EW compute internal rates of return to schooling on both accepted and offered wages. In general, those computed with offered wages are higher than those computed with accepted wages.

[29] More precisely, in order to fit the model, they impose a symmetry condition on the firm/worker solutions.

5.2 Partial Equilibrium Models with Endogenous Schooling and/or Experience

5.2.1 Eckstein and Wolpin (1989): Dynamic Labor Force Participation

Eckstein and Wolpin (1989) is the first structural dynamic programming model of human capital accumulation set in a Mincerian framework and estimated from micro-data. The authors estimate a structural dynamic model of married women's labor force participation. The authors focus on the endogeneity of work experience but they condition on schooling attainment. In their model, women experience a disutility of work but also take into account that accumulated experience raises future wages. Both fertility and schooling are exogenous. Current labor force participation affects future wages which itself affect future labor force participation. Eckstein and Wolpin (EW) pay a particular attention to intertemporal substitution and build a model which embodies intertemporal substitution both through preferences and constraints. The model is estimated using data from the National Longitudinal Survey of mature women (they restrict attention to those who were between 39 and 44 years old in 1967). This avoids modeling childbearing decisions.

EW model time allocation between 2 states (work and non-participation) from age 45 until retirement. The utility function is linear and additive in consumption so the authors may disregard borrowing and saving behavior. However, as the utility function is not intertemporally separable, it is consistent with the existence of a diminishing marginal utility of leisure (non-market time) or its opposite, namely habit persistence. The law of motion that describes human capital accumulation is classical; the effect of schooling on wages is linear and the effect of experience is quadratic such as in the most common form of the Mincerian wage regression. EW do not consider unobserved heterogeneity in market ability but focus on selectivity issues arising in the presence of heterogeneous tastes (or distastes) for work (unobserved heterogeneity is introduced as a intercept term in the instantaneous utility of work).

The results indicate that labor market participation reduces total utility and that the disutility of work increases with schooling. As well, the disutility of work rises with accumulated experience. The return to schooling is estimated to be 0.05 while the return to experience and its square are found to be 0.024 and -0.0002 respectively. Interestingly, the authors also investigate how taking into account unobserved heterogeneity may affect the return to experience. When the authors re-estimate their model with an individual specific unobserved taste for work (fixed effect), the return to experience is diminished slightly but state dependence remains important. This is the first example of the use of structural estimation in order to correct for heterogeneity bias.

5.2.2 Keane and Wolpin (1997): Life Cycle Schooling and Occupation Choices

Keane and Wolpin (1997) constitutes a seminal piece in the literature. The authors estimate a structural dynamic programming model of schooling, home production and occupation choices (blue collar, white collar and military occupations). The model is fit on a sample of young males taken from the 1979 cohort of the NLSY. The model is original in many dimensions. First, it is set in a multiple skill framework where the different skills are defined by occupation types. Second, experience accumulation is endogenous; that is individuals choose to work or not and the type of work experience to accumulate. As a consequence, Keane and Wolpin (KW) estimate a structural dynamic programming model of schooling, home production and occupation choices (blue collar, white collar and military occupations). The authors estimate occupation specific returns to schooling in a context where both schooling and occupation type are endogenous. This is a major achievement.

To put the paper in perspective, it is useful to consider the form of the Mincerian regression function. It is given by the following expression

$$Ew_{imt} = \varphi_{m1} \cdot S_{it} + \varphi_{21} \cdot Exp_{imt} + \varphi_{22} \cdot Exp_{imt}^2 + \eta_{im} \tag{44}$$

where m is the occupation indicator. It is important to note that accumulated experience in each different occupation affects the mean wage offer in every occupation. The term η_m refers to occupation specific unobserved skills by age 16. The regression function therefore allows for heterogeneity in slopes, although the heterogeneity is only allowed though specific occupations. Given occupation, the regression function is specified as a classical Mincer wage regression and it does not belong to the class of correlated random coefficient wage regression models introduced in Section 3.5.

As in most papers covered in this paper, Keane and Wolpin postulate that the per-period utility function is additive in the stochastic error term. The authors assume that the per-period utility of work is the wage rate, while the utility of attending school is denominated in monetary equivalent. There exist no closed-form solution for the value functions. They must rely on repeated numerical solutions of the value functions. Given the complexity, KW develop an approximation method. The approximation method is based on interpolation methods based on simple OLS regression techniques.[30]

Keane and Wolpin consider a rich heterogeneity specification. The utility of attending school is assumed to be affected by an individual specific unobserved heterogeneity term as well as age and grade level specific costs. This heterogeneity term is allowed to be correlated with occupation specific skills appearing in (21). As is common in the structural literature, they assume that the population distribution is approximated by a discrete distribution with a fixed (known) number of types. KW set the number of types to four. As they model choices from age 16 onward, they must condition on observed differences in schooling attainment and let the type probabilities depend on schooling.[31]

Interestingly, given the allowance for endogenous occupation choices, it is possible to view the correlation between various skills and schooling attainment as a source of occupation specific ability bias. Basically, the correlation between individual/occupation specific skills and the utility of attending school will dictate whether or not those who are more able (at one occupation) will achieve a higher level of schooling. Because the wage regression is linear in unobserved skills, the model analyzed by the authors is compatible with the existence of a negative as well as a positive ability bias.

Keane and Wolpin find that white collar skills are the most strongly correlated (positively) with taste for schooling. While they do not report correlations between skills, this may be inferred from the type specific rank for each heterogeneity component. Indeed, the type specific rank for white collar skills and taste for schooling are exactly coincident. [32]

In their preferred specification, the returns to schooling are found to be 0.070 for white collar occupation, 0.024 for blue collar occupation and 0.058

[30] It is discussed in details in Keane and Wolpin (1994).

[31] This is a way to take into account the endogenous initial conditions.

[32] This is found in Table 9, page 502.

for military occupations. Obviously, there are no comparable results in the IV literature but, when averaged over all types, these returns are low. This is especially true when compared to OLS estimates of the return to schooling applied to different waves of the NLSY, which range between 0.09 and 0.11.[33]

5.2.3 Cohen-Goldner and Eckstein (2003 and 2004): Training and Occupation Choices of Immigrants to Israel

In two companion papers, Cohen-Goldner and Eckstein (2002 and 2004) analyze the behavior of male and female immigrants newly arrived from the former Soviet union and who choose between working and attending government provided training courses. As in Keane and Wolpin (1997), the agent has the choice between working in a blue-collar job or a white collar job. In both of theses papers, the authors assume that both the job offer rate and the labor market wage depend on the occupation as well as the participation in training. This constitutes a more general framework to evaluate publicly provided training program than what is typically found in the evaluation literature.

As the papers are structured in a similar fashion, I first summarize the 2004 paper, which focuses on female immigrants. At the end, I will briefly state the findings of the 2002 paper. The data is based on a set of retrospective surveys conducted between 1992 and 1995. The surveys targeted immigrants from the former Soviet Union who came to Israel between 1989 and 1992. As schooling has been completed before the migration option was actually feasible, it is treated as exogenous. The offered wage in occupation j follows a Mincerian wage function that is

$$\bar{w}_{jt} = \varphi_{j1} \cdot S_i + \varphi_{2j} \cdot K_{1,t-1} + \varphi_{3j} \cdot K_{1,t-1}^2 + \varphi_{4j} \cdot K_{2,t-1} + \varphi_{4j} \cdot K_{2,t-1}^2 + \varphi_{6j} \cdot DT_t \quad (45)$$

where $K_{j,t-1}$ is actual work experience accumulated in occupation j and DT_t is an indicator equal to 1 if the individual has completed a training program by date t. The offer probability is also occupation specific and depends on the training indicator.

[33]Keane and Wolpin do not report the population average but based on my own calculation, and after setting initial schooling at its sample average, the population average is close 0.058.

It should be noted that φ_{6j} constitutes one of the most investigated parameters in the "treatment effect" literature. This is particularly true in the case where the effect of training is allowed to be individual specific. It is important to note that the SSDP approach proposed by Cohen-Goldner and Eckstein may be viewed as an alternative estimation method to the popular treatment effect models common in the empirical literature.[34] The key difference is that their model is set in a dynamic setting and it considers multiple ways in which training may affect labor market outcomes; namely wages, employment (job search outcomes) and timing of re-employment. The individual benefit from the availability of training is measured in this study by the increase in expected lifetime utility, which consists of the effect of training on employment, wages and preferences. As the authors use quarterly data, they model the first 20 quarters specifically and approximate the terminal value function (in quarter 21) by a linear function.

The results indicate that both schooling and experience imported from the country of origin have no effect on wages. However, an additional year of white collar experience raises white collar wages by 3.9% and blue collar wages by 2.7%. The return to blue collar experience is negligible in both sectors. More importantly, the wage return to training is almost 20% in white collar occupations but it is not statistically different from zero in the blue collar occupation. As well, the authors find that training has a positive impact on job offer probabilities in both occupations. However, the impact is much stronger on white collar job offers. Overall, the results suggest that training reduces unemployment substantially.

The authors conclude by performing policy experiments such as reducing and increasing training availability. This is achieved by lowering or increasing the probability of finding a training opportunity. They conclude that, unlike what is found in the classical evaluation literature, the social and private gain to training is large. This is true despite the virtually null effect that training has on blue collar wages. These findings are fully consistent with the existing literature. Indeed, in the training literature, it is customary to report very low estimates for the effects of training on wages (Heckman, LaLonde and Smith,1999). That is, the effect of training on the mean wage offer for less advantage workers is close to zero. However, it has been recognized that training may significantly affect the employment probability (Ham and

[34]This is also the case in Belzil and Hansen (2002 and 2005).

Lalonde, 1996).

Finally, in Cohen-Goldner and Eckstein (2002), a similar model is fit to the male counterpart sample. The fundamental behavior of male immigrants and female immigrants do not appear different. Unlike in the paper devoted to female immigrants, the authors are able to fit the dynamic programming model for two types of individuals. The results are quite similar, although the return to white collar experience and white collar vocational training (0.15 and 0.10 respectively) are higher. Fundamentally, as for female immigrants, a positive effect of vocational training is only found for white collar occupations.

5.2.4 Belzil and Hansen (2002,a): The Convexity of the Wage/Schooling Relationship and the Ability Bias

Belzil and Hansen (2002) use a dynamic programming model in order to investigate the shape of the wage schooling relationship and evaluate the ability bias as defined by the correlation between schooling achievement and the individual specific intercept term of the wage equation. In their paper, there are three states; schooling, labor market work, and an intermittent state capturing the fact that the schooling acquisition process is not necessarily continuous. As the objective is to obtain structural estimates comparable with those reported in the IV literature, they do not distinguish between occupation or sectors. The model may therefore be viewed as single skill model.

The model is implemented on a panel of white males taken from the National Longitudinal Survey of Youth (NLSY) covering the years 1979 to 1990. The sample appears to be quite close to the one analyzed in Keane and Wolpin. The Mincerian wage function is specified as

$$\bar{w}_t = \varphi_1(S_t) + \varphi_{21} \cdot Exp_t + \varphi_{22} \cdot Exp_t^2 + \eta \tag{46}$$

where $\varphi_1(.)$ is left unspecified and is estimated flexibly with spline techniques. They allow for the local returns to differ from grade level 8 to 17.[35]

Belzil and Hansen (BH) parameterize the utility of attending school as a function of parents' background variables (father and mother's education,

[35] This is the case because in the sample analyzed, virtually everyone has at least 7 years of schooling.

family income, number of sibling, regional dummies and family status indicating whether the individual was raised by both parents). However, the utility of work (the log wage rate) is not function of the parents' background variables.[36] Unobserved heterogeneity has two dimensions: heterogeneity in school ability (taste for schooling), and heterogeneity in market ability. They assume that there are K types of individuals and that each type is endowed with a pair of school and market abilities $(\bar{\omega}_i^k, \eta_i)$ for $k = 1, 2...K$ and set $K = 6$. The distribution of unobserved ability is orthogonal to parents' background by construction and should be understood as a measure of unobserved ability remaining after conditioning on parents human capital. Note that BH model schooling choices from grade seven onward, so they do not really need to take into account endogenous initial conditions.

The correlation between ability in school and ability in the market, $\text{Corr}(\bar{\omega}_i^k, \eta_i)$, is found to be very high (0.95). In order to evaluate the potential ability bias, BH perform type specific simulations of schooling decisions and wage histories. They find that the correlation between unobserved market ability and realized schooling, $Corr(\eta_i, S_i)$, is equal to 0.28. Orthogonality between market ability and realized schooling is therefore strongly rejected. This provides evidence in favor of the existence of a strong positive ability bias although the correlation is technically speaking not a structural estimate of the OLS bias.

The estimates of the Mincerian wage regression indicate that the marginal returns are generally increasing with the level of schooling up to grade 14. The local returns to college training are substantially higher than the returns to high-school education. Indeed, schooling has practically no value until grade 12. Until grade 10, the local return is below 1% per year (0.4%). It increases to 1.2% in grade 11 and to 3.7% in grade 12. Beyond high school graduation, the local return starts to increase substantially. The local return increases to 6.0% in grade 13 and 12.7% in grade 14. After a drop at grade 15 (the local return is around 10.7%), the return to grade 16 raises to 12.2%. In subsequent years (corresponding to graduate training), the local returns are estimated to be 8.8% per year. Until college graduation and contrary to what had been often postulated, the log wage regression equation is generally convex in schooling.

Aside from the estimates reported, the main contribution of the authors

[36]However, this assumption is relaxed in a companion paper (Belzil and Hansen, 2003)

is the illustration of the distortions introduced in a model built on the assumption that the local returns to schooling are constant.[37]

5.2.5 Eckstein and Wolpin (1999) School Drop Out

Eckstein and Wolpin (1999) present a structural model of high-school attendance, work and academic performance. It is important to note that EW do not focus on the return to schooling. Indeed, post high school graduation is not modeled explicitly and they do not rely on Mincerian wage regressions. However, despite a high level of complexity, the model investigated by EW offers a relatively clear picture of the role of comparative advantages in school/work decisions. For this reason, it deserves some attention.

Basically, the model is structured as follows. Young individuals receive both part-time and full-time wage offers which depend on their ability endowment. Their ability endowment is itself correlated with their abilities affecting school performance and the consumption value of attending school. Working reduces school performance as well as leisure time, which is also valued by young individuals. The model is estimated using data from the NLSY 79. The sample is smaller than the one analyzed in Keane and Wolpin (1997) The smaller sample size is explained by the age requirement. More precisely, Eckstein and Wolpin only consider young males who were less than 15 as of October 1, 1977.

EW stress four main questions. Who drops out of high-school? Why do youths drop out? Does working while attending school affects school performance? Would restrictions on employment affect the drop-out rate? The likelihood function (simulated) is the joint probability of school attainment, work hours (discretized), wages, credits and grades (school performance). As

[37]In a companion paper, Belzil and Hansen (2003) invetigate the relative importance of family background variables and individual specific abilities in explaining cross-sectional differences in schooling attainments and wages. Given scholastic ability, household background account for 68% of the explained cross-sectional variations in schooling attainments. Interestingly, more than one half of this 68% is explained by father's and mother's schooling alone. However, individual differences in wages are mostly explained by ability endowments. Parents background variables account for 27% of the explained variations in wages while unobserved abilities (orthogonal to family background variables) account for 73%. When scholastic ability is correlated with family background variables, the role of ability is even stronger. Ability endowments explain as much as 81% of wages while only 19% is explained by family background variables.

in Keane and Wolpin (1997), all heterogeneity is regarded as unobserved, and they assume that the population is represented by four types.

Basically, they find that working while in school reduces academic performance although the effect is small. They find that the school drop out decision is typically explained by at least one of the following attributes; a low school ability (or low motivation), high ability at wages that do not require high school graduation, high value of leisure or low attached value to high school graduation. Altogether these results imply that policies aimed at forcing young individuals to stay in school will not really be effective at increasing graduation rates.

Overall, the rich heterogeneity components allowed in EW illustrates the notion of comparative advantages that hides behind the positive correlation between market ability and taste for schooling reported in Belzil and Hansen (2002).

5.2.6 Belzil and Hansen (2005): The Correlated Random Coefficient Model and the OLS/IV Puzzle

In Belzil and Hansen (2005), the authors investigate the properties of a restricted version of the correlated random coefficient wage regression where the return to experience is homogenous.[38] The starting point of the paper is the following log wage function received by individual i, at time t, which is given by

$$\bar{w}_{it} = \varphi_{i1} \cdot S_{it} + \varphi_{21} \cdot Exp_{it} + \varphi_{22} \cdot Exp_{it}^2 + \eta_i \tag{47}$$

They assume that $\varphi_{1i} = \bar{\varphi}_1 + \omega_{1i}$ where $\bar{\varphi}_1$ represents the population averages. They show that the estimates of the dynamic programming model with a rich heterogeneity specification, may be used to obtain virtually all treatment effects proposed in the microeconometric literature; the average treatment effects (ATE), the average treatment effects for the treated and the untreated (ATT/AUT), the marginal treatment effects (MTE) and, finally, the local average treatment effects (LATE). More specifically, BH are able to reconcile the well known discrepancy between OLS estimates of the returns to schooling and their IV counterparts (the LATE estimators). To do so, they simulate various experiments: an overall decrease in discount rate, a

[38] The more general version is investigated in Belzil and Hansen (2002,b).

decrease in discount rate from grade 13 onward , and an increase in the utility of attending school in grade 12 [39] In turn, these artificial experiments provide the authors with valid instruments which may be used in order to estimate the returns to schooling by standard IV techniques.

The results indicate that the degree of dispersion found in the population can be reconciled with the existence of high IV (or LATE) estimates. They are comparable to those very high estimates often reported in the literature. They exceed both OLS and the structural dynamic programming estimate (the population average). Using various simulations, BH are able to illustrate the degree to which heterogeneity in the returns is mostly a reflection of the high positive correlation between the returns to schooling and the individual specific reactions. The results indicate clearly that the effects of a policy change are not solely located within a sub-population of individuals who are close to the margin of entering college but that the high IV estimates are also explained by the reactions of those who would have attended college no matter what. In the literature, the local average treatment effect is to be understood as a measure of the returns to schooling within the sub-population affected by the experiment but it is almost always interpreted as a measure of the effect of a policy experiment for those who are at the margin of entering college.

These results are encouraging. They indicate that the structural dynamic programming model with multi-dimensional heterogeneity is capable of explaining the coexistence of relatively low returns to education (on average) with very high returns for some identified sub-populations and is capable of explaining the well known OLS/IV puzzle. As well, the structural approach is capable of unifying various treatment effect measures such as the LATE, the ATT and the MTE.

5.2.7 Todd and Wolpin (2003): Child Schooling and Fertility

Todd and Wolpin's contribution is substantially different from all other papers surveyed in this article. The authors estimate a dynamic programming model of child schooling and fertility using data taken from a school subsidy experiment (PROGRESA) which took place in Mexico during the late

[39] In the literature, it is sometimes argued that differences in credit constraints may be captured in the discount rate (see Cameron and Taber, for a recent example).

1990's. This is one of the first attempts to use both structural and experimental approaches simultaneously. As pointed out by the authors, this social experiment departs completely from past policies that had been implemented and, for this reason, standard experimental methods used to evaluate various programs, are not feasible in this particular context.

Their approach, while fully structural, is also experimental in nature. They estimate a structural model of household fertility and child schooling decisions on a sample that includes both the control group and the treatment group (prior to the implementation of the experiment). In their model ,households make sequential decisions about the timing and spacing of births as well as about time allocation (schooling and work) of their children until age 15. The key identification strategy in the model comes from the household consumption equation. Todd and Wolpin assume that consumption is enhanced by children's earnings. In the presence of a child labor market, fluctuations in wages provide time varying measures of the opportunity cost of school attendance. In turn, this allows the authors to simulate various counterfactual policies, including the introduction of a school attendance subsidy such as the actual one implemented within PROGRESA.

The existence of the experiment provides them with the opportunity to use out-of-sample forecasts to validate the behavioral model estimated. Todd and Wolpin evaluate the performance of the model by comparing the predicted impact with the observed one. They find that, in general, the estimated structural parameters provide an accurate forecast of the effect of the program, as measured by the observed changes in schooling and fertility behavior.[40] Though the model is set in partial equilibrium, some important lessons may be gained from this exercise. First, it illustrates how an internally consistent dynamic behavioral model (estimated from the non-experimental portion of the data) may indeed provide valid forecast of the impact of counterfactual policies. Second, it also illustrate that, within a context where individuals or household optimize over their lifecycle, the impact of an experiment may be long lasting. If so, experimental data collection should, in general, and as it is the case in PROGRESA, cover longer periods than those usually considered in the treatment effect literature.

[40] The authors also simulate various policy experiments and also examine the long run impact of the program.

5.3 General Equilibrium Models

5.4 Heckman, Lochner and Taber (1998): Post Schooling Human Capital and Inequality

Heckman, Lochner and Taber (1998) is the first general equilibrium model with endogenous schooling. As Keane and Wolpin (1997), Heckman, Lochner and Taber (HLT) set their model in a multi-skill environment. Actually, the model allows for two skills, but skills are associated with schooling level (high school graduates and college graduates) as opposed to occupation (Keane and Wolpin, 1997).

It is convenient to think about HLT as a two-period model. Individual schooling decisions are made in period 1. The choice is to attend college or not and it is affected by a random shock. In turn, this choice dictates the sector in which an individual will be working. Subsequently, on-the-job training and saving decisions are made (deterministically) for the remaining periods.

It should be noted that the human capital production function chosen by HLT is more general than most of those used in the literature. There exist a skill specific function for each school level. The return to schooling therefore varies with age (as post schooling human capital is accumulated) and cannot be summarized in a single parameter. As a matter of fact, the model does not belong to the class of linear separable models in which the returns to experience is independent of schooling.

In their model, the wage rate is the product of the skill rental price and the amount of skills devoted to their current employment, after removing the current fraction of the period devoted to on-the-job training). There is a key difference between a general equilibrium model and a partial equilibrium model. The returns to schooling depends on how many other people go to college since the wage rate decreases with school attendance. In partial equilibrium, this is ignored.

HLT pay a particular attention to the effect of a simulated change in tuition policies on college enrollment, a topic beyond the scope of this survey.[41] A notable difference between HLT and the rest of the papers in the literature is the modeling of post schooling human capital accumulation. As opposed to models where experience is only modeled through the decision to work or

[41]For a survey of the literture concerned with tuitio policies, see Wolpin (2000).

not (say like Keane and Wolpin, 1997), the portion of time endowment spent on on-the-job training activities is decided by the workers.

In HLT, there is no unobserved heterogeneity. Difference in skills/motivation are measured by AFQT scores. As the authors assume the existence of a specific human capital production function for each schooling level, the wages are not represented by a classical wage regression function with separability. Schooling facilitates on-the-job training and the return to schooling is changing with age.

As stated before, the returns to schooling reported in HLT are not directly comparable to those found in Mincerian wage regression. The returns are age specific and take into account the causal effect of schooling on training opportunities. As well, their estimates are not computed for a marginal year but for a high school/college graduate differential. When put back on a per-year basis, their estimates range between 8% per year and 17% per year. Given this, these estimates are not particularly high. They are indeed compatible with those reported in Belzil and Hansen (2002).[42]

5.4.1 Lee (2004): Cohort sizes, Occupation Choices and Wages

Lee (2004) estimates a dynamic general equilibrium model of career decisions in the spirit of Keane and Wolpin (1997). Using CPS data on employment, cohort sizes schooling attainment and occupation choices, Lee investigates how cohort size affects wages and investigates how a counterfactual tuition policy changes would change college enrollment. From the individual optimization side, his model is quite close to Keane and Wolpin (1997). Basically, individual make schooling and occupation choices made between age 16 and 65 based on the skill endowment at age 16 and on current as well as future skill prices. Lee focusses on two main skills; blue collar and white collar occupations. However, in his model, skill prices are determined endogenously in the market. They depend on the aggregate supply of white and blue collar workers as well as on capital. The marginal product of each skill is derived through an aggregate production function. In order to estimate the model, he assumes that individuals have perfect foresight about future skill prices

As in Keane and Wolpin, his wage regression is linear in schooling and quadratic in accumulated experience and he estimates returns to schooling

[42]In Belzil and Hansen (2002), the return to college education (4 years) averages 10.4% per year of college.

which are occupation specific. Given the complexity of the model, he assumes that his population is composed of two types only. Aside from the general equilibrium aspects, Lee also needs to approximate the value functions numerically He uses approximation techniques which appear similar to those used in Keane and Wolpin (1997).

The main results are consistent with was reported before. First, he finds a positive correlation between white collar skill and taste for schooling but a negative correlation between the blue collar skill and taste for schooling. As in Keane and Wolpin (1997) and Belzil and Hansen (2002), he finds relatively low returns to schooling. The return to schooling in white collar occupation is 0.079 while it is equal to 0.048 in the blue collar occupation. As well, the return to experience upon entrance in the market exceeds the return to schooling (it is 0.094 for white collar experience and 0.022 for blue collar occupation). Interestingly, the returns are quite close to those obtained by Keane and Wolpin (1997), even though they are obtained from a different data set.

5.5 Some Alternative Strategies: Semi-Structural Models

Until now, the presentation has focused solely on the structural approach. There are alternative methods which are well suited to study the dynamics of schooling decisions These methods are semi-structural in the sense that their implementation does not require optimization.[43]

5.5.1 Cameron and Heckman (2001): Family Background and Racial Differences in Schooling Attainment

Cameron and Heckman (2001) model schooling attainment as a sequence of decisions made at each age and at each grade level. Dropping the individual subscript, the point of departure is a linear approximation of the choice specific value functions (equation 15).

$$V_{a,j,c} = Z'_{a,j,c} \cdot \beta_{a,j,c} + \varepsilon_{a,j,c} \tag{48}$$

[43] In the literature, there seems to be no formal definition of the term "semi structural".

where $V_{a,j,c}$ is the intertemporal utility of choosing option c, given grade level j, at age a. $Z'_{a,j,c}$ is a vector of observed variables and $\beta_{a,j,c}$ is the corresponding vector of parameters. Cameron and Heckman use a factor loading structure and assume that

$$\varepsilon_{a,j,c} = \alpha_{a,j,c} \cdot \eta + \nu_{a,j,c} \tag{49}$$

where the $\nu'_{a,j,c}s$ follow an extreme value distribution and η is a mean 0 and unit variance random variable and the $\alpha_{a,j,c}$ are parameters . The $\nu'_{a,j,c}s$ and η are assumed to be independent. It follows that the probability of choosing option $c\prime$ at age a, with j grade level completed is

$$pr(\arg\max V_{a,j,c} = c') = \frac{\exp(Z'_{a,j,c'} \cdot \beta_{a,j,c'} + \alpha_{a,j,c'} \cdot \eta)}{\sum_c \exp(Z'_{a,j,c} \cdot \beta_{a,j,c} + \alpha_{a,j,c} \cdot \eta)} \tag{50}$$

and is therefore a generalization of McFadden's conditional logit model. Cameron and Heckman investigate out the distribution of η using non-parametric methods.

The model is fit on a sample of young black, hispanic and white males taken from the NLSY 1979. The authors pay a particular attention to the effect of family income versus long-run factors (such as parents' schooling) and examine racial differences in schooling attainments. They conclude that family background variables other than family income have a large impact on grade level transition and that family income is more important in explaining earlier grade transition than later ones. For this reason, they conclude that short-term credit constraints are probably not important. With respect to racial differences, their estimates imply that racial differences are almost fully accounted for by differences in family background. Given background variables, minorities are even more likely than whites to graduate from high school and attend college.

5.5.2 Magnac and Thesmar (2002,b): Cohort effects in Schooling Attainments in France (1980-1993)

Magnac and Thesmar (2002,b) also estimate a dynamic schooling model in which does not require to solve (or approximate) value functions. They use the fact that there exists a one-to-one mapping between value functions and choice probabilities (Hotz and Miller, 1993) in order to analyze the degree of under-identification of dynamic discrete choices (Magnac and Thesmar, 2002,a). Their constructive identification proofs suggest a simple estimation method.[44] In their empirical application, MT investigates three competing explanations for the increase in schooling attainments observed between 1980 and 1993 in France. These factors are the increase in the return to education, a decrease in the direct and psychic costs of schooling and a decrease in academic requirements (an increase in the success probability given enrolment).

The model is set up as a standard optimal stopping problem and is estimated from data taken from the "Enquete Formation Qualification professionnelle" performed in 1993 by INSEE. They focus on cohorts of individuals born between 1963 and 1973. MT show that flexible regression methods may be applied to the sample analog of the net gain of staying in school for one period, and may be used to resolve all questions mentioned above. Interestingly, the authors find that, in France, the increase is most likely explained by the decrease in academic selectivity.

5.5.3 Belzil (2004): The Functional Form of the Mincer Wage Equation

Belzil (2004) uses a semi-structural model of grade transition and wages to estimate the returns to schooling and experience. The econometric model is based on two distinct components; a reduced-form dynamic model of schooling attainment based on the hazard specification of the transition from one grade level to the next with observed and unobserved heterogeneity (as in Cameron and Heckman, 1998 and 2001) and a non-linear Mincerian wage regression model with observed and unobserved skill heterogeneity. Indeed, the mixed likelihood function maximized by the author may be seen as a reduced-form equivalent of (24) or (25), where the schooling decisions probabilities involving various Bellman equations are replaced by hazard rates. The main objective behind the model specification is to restrict the number of behavioral, parametric and distributional assumptions to a minimal level and, in particular, to estimate the returns without enforcing selectivity on all heterogeneity components.

[44] The method develop by Magnac and Thesmar (2002,a) uses the fact that value functions, after suitable normalization, are functions of choice probabilities only. It is therefore simpler than the method propose by Hotz and Miller (1993).

Belzil specifies the Mincerian wage regression quite generally; skill heterogeneity affects the intercept term, the return to schooling and the return to experience, ii) the local return to schooling may vary with grade level (the return to college may be different than the return to grade school or high school), iii) the returns to experience depend on accumulated schooling and iv) the distribution of wage offers are conditionally (on skill heterogeneity) heteroskedastic.

The model rejects all simplifying assumptions common in the empirical literature. Belzil finds that the degree of convexity of the wage regression, as measured by the difference in the local returns to schooling before and after high school graduation, is only slightly dependent on the allowance for skill heterogeneity. However, The convexity is acute and the log wage regression remains highly convex, even after conditioning on unobserved and observed skills. Skill heterogeneity is also found to be quite important and there is more cross-sectional variability in the returns to experience than in the returns to schooling. Belzil finds that those endowed with high returns to schooling will also be endowed with high returns to experience, although the correlation is quite small. He also finds that ignoring non-linearity inflates the cross-sectional variance in the returns to schooling.

After conditioning on skill heterogeneity, there is a positive correlation between accumulated schooling and the individual specific returns to experience. This is consistent with the view that accumulated schooling may have a causal effect on wage growth. Finally, wages are found to be homoskedastic in schooling but heteroskedastic in experience. This result is consistent with the possibility that senior workers are less exposed to business cycle fluctuations or other stochastic shocks affecting the labor market.

6 Comparing Structural Estimates with IV Estimates

In order to summarize, it is useful to put the key results obtained in the structural literature in perspective with those found in the experimental literature. I believe the key findings of the structural literature are the following:

1. The returns to schooling are relatively low. They are indeed much lower than those reported in the experimental literature and, interestingly, they are low in the context where they are assumed to be homogenous as well as in the context where population heterogeneity is explicitly taken into account. Table 3 summarizes the structural estimates of the returns to schooling in all models in which schooling is endogenously determined. In a linear setting, the marginal effect of schooling averages around 5% per year of schooling. Interestingly, the returns are also low when non-linearities are allowed, although the local returns in college average around 10% per year. However, when averaged over the relevant range (from grade 7 to college graduation), the average is also around 5%. These estimates are typically lower than their OLS counterpart. A set of estimates reported in the experimental literature is summarized in Table 5. These are representative of a large body of work and include Angrist and Krueger (1991), Card (1995), Lemieux and Card (2000) and Staiger and Stock (1997). For the most part, IV estimates reported in Table 5 exceed OLS estimates and are close or exceed 10%. Except for Lemieux and Card (obtained with Canadian data), these estimates have been obtained with US data.

2. The return to labor market experience is relatively high, at least when compared to schooling. In table 4, I report a set of parameter estimates for the return to a first year of experience as well as the return to training (when applicable). These estimates are taken from models where experience accumulation is endogenously determined and range between 10% and 20% for the US. While the overall return will be affected by the degree of concavity, it is still informative to examine the return to experience within a short period from entrance in the labor market. Despite this relatively wide range of variation, the returns are higher than the return to schooling. Because the returns to experience

are rarely investigated in the experimental literature, these estimates are difficult to compare.

3. The return to schooling, experience and other forms of general training (including government provided training courses) is much lower in blue collar occupations. For instance, the return to schooling is much below 5% in blue collar occupations but between 7% and 8% in white collar occupations. Because training intensity may vary substantially across programs, there is no natural unit of time upon which the return to training may be based, but evidence suggests that the return to government sponsored training is also much lower in blue collar occupations. As intuitive as these results may be, formal econometric results obtained in a context where occupation choice is endogenous exist only in the structural literature (Keane and Wolpin, 1997, Lee, 2004 and Cohen-Goldner and Eckstein, 2004 and 2005). Basically, the training literature has emphasized the existence low estimates for the effects of training on wages (Heckman, LaLonde and Smith,1999).

7 What else do we Learn from Structural Estimation?

As argued earlier, structural estimation does not only offer a credible alternative to the experimental approach to estimating the return to human capital, but it also provide estimates of relevant parameters (or functions of parameters) that are particularly relevant to the analysis of skill formation policies. The following findings, unlike for the estimates of the returns to schooling, experience and training, have no history in the experimental literature.

1. As postulated by many economists, in a single skill framework where market ability enters the wage equation linearly, schooling and market ability are positively correlated. That is the sign of the ability bias is positive. Moreover, this correlation is explained by the very high correlation between the per-period utility of attending school (the taste for schooling) and market ability (Belzil and Hansen, 2002, Keane and Wolpin, 1997). Again, evidence on the sign of the ability bias found in the experimental literature is based only on the discrepancy between

OLS and IV estimates, assuming a linear effect of schooling on log wages.

2. In a multiple skill framework, there is a high correlation between the taste for schooling and white collar market skills. The correlation between blue collar skills and taste for schooling is much lower (Keane and Wolpin, 1997, and Lee, 2005).

3. The literature finds support for the correlated random coefficient correlation specification of the Mincer wage equation. However, the relatively low average return in the population seem to coexist with very high returns for a subset of the population. Indeed, the structural literature can explain relatively well the discrepancy between OLS estimates of the return to schooling and their IV counterparts (the local average treatment effects). Off course, the experimental literature based on IV techniques, the level of cross-sectional dispersion in the returns to schooling (or experience) is not identifiable.[45]

4. The private gain to training (general) may be large. This is true despite the virtually null effect that training has on blue collar wages and it may largely be explained by the fact that training substantially increases the job offer rates in both white collar and blue collar occupations as well as for the unemployed. In the experimental literature, the effect of training on job offer arrival rates are practically never investigated.

8 Why do Structural Estimates and IV Estimates of the Return to Schooling Differ?

This is, of course, the natural question to ask at this stage. In this section, I will focus solely on the returns to schooling.[46] As stated earlier, neither estimation method is directly nested within its competing one. Furthermore,

[45]However, there is also a recently growing literature that focuses on non-parametric instrumental methods (see Heckman and Vytlacil, forthcoming).

[46]To a certain extent, comparing estimates of the return to experience would be a more difficult task. After all, very few reduced-form papers estimate the return to experience and, furthermore, post-schooling human capital accumulation embodies a wide range of activities.

IV methods will typically provide a single point estimate for the return to schooling whereas SSDP techniques may provide a single estimate, a grade specific estimate or a population distribution, depending on the dimensionality of the heterogeneity terms or on the degree of non-linearity incorporated in a particular model. As a consequence, a formal answer cannot really be given. Nevertheless, I believe that the elements presented below are worth considering. For the sake of the discussion, it is informative to distinguish the case in which the return is a single parameter (the effect of schooling is linear on log wages and everyone faces the same parameter) from more general cases where the return varies with the level of schooling or varies across individuals.

8.1 The Return to Schooling as a Single Parameter

As a first step, it is natural to search the intrinsic nature of the structural approach for obvious defects which could, for instance, lead to low returns to schooling. This is a difficult task. A survey of the structural papers in which an estimate of the ability bias is either directly or indirectly available would reveal that the correlation between the utility of attending school and labor market ability is always unrestricted. For that matter, none of the structural estimates presented herein seem to be systematically biased toward a positive ability bias. Similarly, the existence of potential (significant) measurement error, often put forward as an explanation for the OLS/IV discrepancy, is very unlikely to be an explanation. My understanding of the literature is that the measurement error argument often advanced is typically set within a classical framework which ignores the correlation between schooling levels and the measurement error itself and also ignores the discrete nature of the schooling variable. As well, structural models are intrinsically non-linear and the estimates are obtained from the solution of a large number of moment conditions.

A second set of answers refers to the reliability of the experimental approach. Instrumental variables (IV) techniques may be applied in a context where the instrument is only weakly correlated with schooling attainments (Staiger and Stock, 1997). Indeed, before the late 90's, most empirical researchers concentrated their efforts on finding an instrument uncorrelated with neglected ability, but the power of the instrument chosen was practically never investigated. In the presence of weak instruments, reported estimates may be at best imprecise and, at worst, seriously biased (or inconsistent). The large bias is explained by the magnifying effect that the weak correlation between the instrument and the endogenous variable may have on the possible correlation (non-zero) between the instrument and the error term of the regression. As a consequence, the validity of very high returns to schooling, reported in a simple regression framework, may be seriously questioned. Carneiro and Heckman (2003) presents an in-depth analysis of various instruments used in the literature.

As seen earlier, another explanation may simply be that the Mincerian wage regression incorporates two many endogenous variables and that the dimensionality of the vector of instruments required is larger than the number of instruments actually available. Put differently, it is the lack of control for the endogeneity of post-schooling work experience that may be the cause. However, as appealing as this explanation may be, in the presence of a diversity of post-schooling opportunities such as training, search and learning, it is difficult to predict the sign of the correlation between omitted post schooling wage growth and both schooling and specific instruments. Only further work will clarify this issue

8.2 The Return to Schooling is Non-linear

When the effect of schooling depends on schooling level, the discrepancy is much easier to explain. A survey of the reduced-form literature reveals that when IV techniques are chosen, the log wage regression is usually assumed to be linear in schooling and, perhaps, quadratic in experience. However, there is no obvious reason to presume that the local returns to schooling are independent of grade level. As individuals with a lower taste for schooling tend to stop school earlier, OLS (or IV) estimates of the return to schooling, which impose equality between local and average returns at all levels of schooling, will be strongly affected by the relative frequencies of individuals with high and low taste for schooling.[47] More precisely, if there are large differences in local returns between various grade levels, the OLS estimate (measuring an average log wage increment per year of schooling) will tend to be biased toward the local returns at schooling attainments that are the most common in the sample data. Therefore, the difference between the average return to

[47] This issue is sometimes referred to as the Discount Rate bias.

schooling obtained from structural estimates (Belzil and Hansen, 2002) and IV estimates is easily explained.

8.3 Heterogeneity in the Returns

As with the non-linear case, the discrepancy between structural and reduced-form estimates is more easily explained than in the single parameter case. As recognized in the experimental literature, in presence of heterogeneity in the returns, the IV estimate is inconsistent for the population average and the resulting estimate may reflect the returns of a sub-population only. Indeed, there is a large econometric literature concerned with the interpretation of IV estimates when the slopes are individual specific. In the context of a random coefficient model, the IV estimator is sometimes referred to as a Local Average Treatment Effect (LATE).[48] The LATE should be understood as a measure of the returns to schooling for the sub-population affected by the experiment. It is often postulated that the high returns are explained by the fact that those individuals more likely to react to an exogenous policy change are those who are at the margin of deciding to enter college before the policy change and that they have higher returns to schooling than average.

As seen earlier, this may be verified within a structural framework. Belzil and Hansen (2005) present a characterization of several counterfactual experiments and show how the reactions are correlated with the individual specific returns to schooling. The examples provided in BH (2005) illustrate the key distinction between the interpretation that must be given to IV estimates set in a discrete (2 period) framework and estimates obtained within a dynamic model. In a simple two period model, the individuals who would have decided to attend college without exposure to the instrument are obviously not affected by policy changes. In the literature, the discussion of the IV interpretation is often pitched within such a framework and, for this reason, the high estimates are imputed solely to those who would not have attended college without exposure to the instrument and, more precisely, those "who are at the margin" of choosing college enrollment. In a fully dynamic setting, the issue is more complicated. If one considers introducing a college attendance subsidy (paid over 4 years of college), the subsidy will not only affect those who would not have attended college but also some of those who

[48]See Imbens and Angrist (1994).

would have entered college even without being exposed to the experiment by increasing their continuation (graduating) probabilities. It is therefore important to understand that the effects of a policy change are therefore not solely located within a sub-population of individuals who are close to be at the margin ex-ante.

9 Conclusion and Avenues for Future Research

In conclusion, it is probably fair to say the discrepancy between structural and IV estimates is primarily a reflection of the differences in objects estimated by the researchers. As is already well known, classical IV methods have been developed in the context of static regression models in which slopes are common to all individuals. They are ill equipped for estimating population averages and they do not arise as a natural estimation strategy in a context where agents continuously optimize in a dynamic environment, even though they are capable of measuring the returns to schooling for sub-populations defined in function of a reaction to a policy change.

SSDP techniques, on the contrary, reflect the desire to model endogenous decisions as well as outcomes for randomly sampled individuals. For this reason, structural estimates are more naturally associated with classical measures of central location. The evidence reviewed in this paper seems to suggest that point estimates obtained within a structural framework are a better indicator of the population average than IV estimates. Interestingly, most structural estimates of the return to schooling are also smaller than their OLS counterparts and therefore imply that OLS estimates probably over-estimate the population average return. In a certain sense, structural estimation has revived the interest in the notion of "ability bias" and has brought credibility to the classical hypothesis that the observed correlation between wages and schooling is an over-estimate of the true causal effect of schooling on wages.

Finally, I would like to identify interesting avenues for future research where the structural approach may prove to be useful. First, as stated earlier, structural models are typically estimated under the maintained hypothesis that persistent unobserved (to the econometrician) heterogeneity is in the information set of the agent from the start of the optimization process. This is perhaps the strongest behavioral assumption made in the structural human capital accumulation literature. Whether it is a valid assumption or not is debatable. It would be interesting to develop estimation framework which allows for gradual learning about academic and, in particular, labor market skills.[49]

Second, despite the fact that most individuals spend a much larger share of their productive life in the market than in school, the structural literature (just like the experimental literature) has focused on the endogeneity of schooling. When treated as endogenous, work experience is modeled through occupation choices or through a simple discrete labor supply. Little is known about the relationship between schooling and the intensity of post-schooling human capital accumulation (on-the-job training decisions as well as work intensity). This is an area where progress is likely to be made in a near future.

Third, the role of search frictions on both schooling and training decisions remains largely hypothetical. At the empirical level, those who have modeled human capital accumulation within a search framework have conditioned on schooling attainments. The effect of wage dispersion and schooling decisions has been completely ignored. There are no compelling reasons to do so.

Finally, despite the general focus put on human capital theory, it is relatively well known that lifecycle wage growth may be partly disconnected from productivity growth. In the presence of incentive based employment contracts, just like in the case of search frictions, the interpretation given to post schooling wage growth becomes problematic (Lazear, 1997). It would be interesting to investigate how schooling decisions are made in a context where promotions and human capital accumulation are alternative methods to enhance life cycle income.

Off course, there are no reason why these topics should be investigated solely by structural applied econometricians. Each of these questions are interesting in their own right and, indeed, structural estimation is often viewed as controversial in empirical labor economics.[50] However, a realistic representation of the human capital accumulation process must recognize that skills may be enhanced by a wide variety of different tasks and that individuals are constantly faced with investment or search opportunities over their life

[49] This issue is analyzed within a reduced-form framework in Cunha, Heckman and Navarro (2005).

[50] It is much less so in empirical industrial organization and in dynamic macro economic theory.

cycle. This means that the occurrence of exogenous events happening at one particular point in time will rarely be sufficient to uncover the key economic parameters that characterize human capital accumulation over the entire life cycle. For this reason, the structural approach should be seen as a key tool for understanding skill formation behavior and for performing relevant policy evaluations.

References

[1] Angrist, J. and Alan Krueger (1991) "Does Compulsory School Attendance Affect Schooling and Earnings" *Quarterly Journal of Economics*, 106, 979-1014

[2] Becker, Gary and B. Chiswick (1966) "Education and the Distribution of Earnings" *American Economic Review, 56, 358-69.*

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

[4] Belzil, Christian and Hansen, Jörgen (2004) "Risk Aversion, Earnings Dispersion and Education" *Research in Labor Economics, volume 23.*

[5] Belzil, Christian and Hansen, Jörgen (2002,a) "Unobserved Ability and the Return to Schooling", *Econometrica, vol 70, 575-591.*

[6] Belzil, Christian and Hansen, Jörgen (2003) "Structural Estimates of the Intergenerational Education Correlation", *Journal of Applied Econometrics, vol18: 679-696.*

[7] Belzil, Christian and Hansen, Jörgen (2002,b) "A Structural Analysis of the Correlated Random Coefficient Wage Regression Model" *IZA Working Paper.*

[8] Belzil, Christian and Hansen, Jörgen (2005) "A Structural Analysis of the Correlated Random Coefficient Wage Regression Model with an Application to the OLS/IV Puzzle"

[9] Cameron, Stephen and Heckman, James (1998) "Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males" *Journal of Political Economy, 106 (2)*, 262-333.

[10] Cameron, Stephen and Heckman, James (2001) "The Dynamics of Educational Attainments for Black, Hispanic, and White Males" *Journal of Political Economy, 109 (3)*, 455-499

[11] Card, David (2001) "The Causal Effect of Education on Earnings" Handbook of Labor Economics, edited by David Card and Orley Ashenfelter, North-Holland Publishers.

[12] Card, David (1995) "Using Geographic Variation in College Proximity to estimate the Return to Schooling in Aspects of Labour Market Behaviour: Essays in Honor of John Vanderkamp, Toronto, University of Toronto Press, 201-222.

[13] Cunha, Flavio, James J. Heckman and Salvador Navarro (2005) "Separating Uncertainty from Heterogeneity in Life Cycle Earnings" IZA Working Paper 1437.

[14] Cohen-Goldner, Sarit and Zvi Eckstein (2004) "Estimating the Returns to Training and Occupational Experience: The Case of Female Immigrants" IZA Working Paper no 1225.

[15] Cohen-Goldner, Sarit and Zvi Eckstein (2002) "Labor mobility of Immigrants: training, Experience, Language and opportunities" IZA Working Paper no 519.

[16] Eckstein, Zvi and Gerard van den Berg (2003) "Empirical Labor Search: A Survey, IZA Working Paper # 929.

[17] Eckstein, Zvi and Kenneth Wolpin (1999) "Youth Employment and Academic Performance in High School", *Econometrica* 67 (6,)

[18] Eckstein, Zvi and Kenneth Wolpin (1989a) "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models". Journal of Human Resources, 24, 562{598.

[19] Eckstein, Zvi and Kenneth Wolpin (1989b) “Dynamic Labour Force Participation of Married Women and Endogenous Wage Growth”. Review of Economic Studies, 56, 375{390.

[20] Eckstein, Zvi and Kenneth Wolpin (1995) “Duration to first job and return to schooling: estimates from a search-matching model”, Review of Economic Studies, 62, 263{286.

[21] Ham J.C. and R.J. LaLonde (1996), The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training, Econometrica, 64(1), 175-207.

[22] Heckman, J.J., LaLonde, R.J. and J.A. Smith (1999), ”The Economics and Econometrics of Active Labor Market Programs”, in Handbook of Labor Economics, Vol. 3, A. Ashenfelter and D. Card (Ed.), North-Holland: Amsterdam.

[23] Heckman, James, Lochner, Lance and Petra Todd (2001) “Fifty Years of Mincer Earnings Regressions” forthcoming in *The Journal of Political Economy*

[24] Heckman, James and Singer, Burton (1984) “A Method for Minimizing the impact of distributional Assumptions in Econometric models for Duration Data, *Econometrica, 52 (2):* pp 271-320.

[25] Heckman, James and E. Vytlacil (2004) “Structural Equations, Treatment Effects and Econometric Policy Evaluations” forthcoming in *Econometrica.*

[26] Heckman, James and E. Vytlacil (1998) “Instrumental Variables Methods for the Correlated Random Coefficient Model”, *Journal of Human Resources*, Volume 33, (4), 974-987.

[27] Heckman James and Sedlacek, G (1985) “Heterogeneity, Aggregation, and Market Wage Functions: An Empirical Model of Self-Selection in the Labor Market‘, *Journal of Political Economy, 93 (6)*, 1077-1125.

[28] Imbens, Guido and J. Angrist (1994) “Identification and Estimation of Local Average Treatment Effects”, *Econometrica, 62, 4,467-76.*

[29] Keane, Michael P. and Wolpin, Kenneth (1997) "The Career Decisions of Young Men" *Journal of Political Economy, 105 (3)*, 473-522.

[30] Lazear, Edward (1997) "Personnel Economics: Past Lessons and Future Directions" *Journal of Labor Economics*" vol 17, no 2, pp 199-236

[31] Lemieux, Thomas (2004) "Increasing Residual Wage Inequality: Composition Effects, Noisy Data, or Rising Demand for Skill? Working Paper, University of British Columbia and NBER, May 2004

[32] Lemieux, Thomas and David Card (2000) "Education, earnings and the Canadian GI Bill", *Canadian Journal of Economic*s

[33] Magnac, T. and D. Thesmar (2002,a): "Identifying Dynamic Discrete Decision Processes," *Econometrica*, vol. 70, 2002, p. 801-816. *Econometrica.*

[34] Magnac, T. and D. Thesmar (2002,b) "Analyse économique des politiques éducatives: l'augmentation de la scolarisation en France de 1982 à 1993 " *Annales d'Économie et de Statistique*, vol. 66, 2002, p. 1-35.

[35] Manski, Charles and John Pepper (2000) "Monotone Instrumental Variables: with an Application to the Returns to Schooling" *Econometrica, 68 (4), 997-1013*

[36] Rosen, Sherwin (1977) "Human Capital: A Survey of Empirical research" Research in Labor Economics, Edited by R. Ehrenberg, JAI Press, vol 1.

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

[38] Roy, Andrew "Some thoughts on the Distribution of Earnings" *Oxford Economic Papers, vol 3 (June), 135-146.*

[39] Rust, John (1994) "Structural Estimation of Markov Decision Processes" in *Handbook of Econometrics,* ed. by R. Engle and D. McFadden. Amsterdam; Elsevier Science, North-Holland Publishers, 3081-4143.

[40] Sauer, Robert (2004) "Education Financing and Lifetime Earnings" Working Paper, *Review of Economic Studies.*

[41] Taber, Christopher (1999) "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability", *Review of Economic Studies*.

[42] Todd, Petra and K. I. Wolpin (2003) "Using a Social Experiment to Validate a Dynamic Behavioral Model of Child Schooling and Fertility: Assessing the Impact of a School Subsidy Program in Mexico" Working Paper, University of Pennsylvania.

[43] Willis, R. and S. Rosen (1979) "Education and Self-Selection", *Journal of Political Economy, 87, S-7-S36. .*

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

[45] Willis, R. and S. Rosen (1979) "Education and Self-Selection", *Journal of Political Economy, 87, S-7-S36.*

[46] Willis, Robert (1986) "A Survey and reinterpretation of human Capital Earnings Functions", Handbook of Labor Economics, Vol 1, Chap. 10, Edited by O. Ashenfelter and R. Layard

[47] Wolpin, Ken (1992) "The Determinants of Black-White Differences in Early Employment Careers: Search, Layoffs, Quits, and Endogenous Wage Growth" *Journal of Political Economy, 105 (3)*, 473-522.

[48] Wolpin, Ken (2003) "Wage Equations and Education Policy" in Dewatripont and L.P. Hansen, *Advances in Economics and Econometrics,* Cambridge University Press.

Table 1
The Economic Structure of the Models

	# of Skills	**Labor Market**	**Equilibrium**
Eckstein and Wolpin (89)	one	Competitive	partial
Wolpin (92)	one	Search	partial
Eckstein and Wolpin (95)	one	Search	partial
Keane and Wolpin (1997)	multiple	Competitive	partial
Eckstein and Wolpin (99)	multiple	Competitive	partial
Belzil and Hansen (2002)	one	comptetitive	partial
Belzil and Hansen (2003)	one	competitive	partial
Lee (2004)	multiple	Competitive	general
Heckman Lochner and Taber (1998)	multiple	Competitive	general
Belzil and Hansen (2005)	single	Competitive	partial
Cohen-Goldner (2002)	multiple	Search	partial
Cohen-Goldner (2004)	multiple	Search	partial
Todd and Wolpin (2003)	single	competitive	partial

Table 2
Specification of the Effect of Schooling in the Mincerian Wage Offer Function

	Functional Form	**Heterogeneity**
Eckstein and Wolpin (89)	linear	none
Keane and Wolpin (1987)	linear	intercept
Eckstein and Wolpin (99)	-	intercept
Belzil and Hansen (2002)	non-linear	intercept
Belzil and Hansen (2003)	non-linear	intercept
Lee (2004)	linear	intercept
Heckman, Lochner and Taber (1998)	-	intercept
Belzil and Hansen (2004)	linear	intercept ret. schooling, ret to expe.
Belzil and Hansen (2005)		wage inter., ret. schooling,

Table 3
Returns to schooling in Models with Endogenous Schooling

	Sample	**return to schooling**
Keane and Wolpin (1997)		
	blue collars (males)	0.024
	white collars(males)	0.070
	military (males)	0.058
Belzil and Hansen (2002)		
	in high school (males)	0.007
	in college(males)	0.104
	average (males)*	0.046
Lee (2004)		
	blue collar (males/Females)	0.048
	white colar(males/Females)	0.079
Belzil and Hansen (2005)		
	pop. average(males)	0.043

*The average refers to the average taken from grade 6 to grade 16.

** The average refers to the population average (over all types) of a linear effect of schooling on wages.

Table 4
Return to a first year of experience
and Return to post-schooling training

		return to experience	return to Training
Eckstein and Wolpin (89)	Females	0.050	-
Cohen-Goldner and Eckstein (2004)			
	blue collar (females)	0.00	0.00
	white colar (females)	0.16	0.19
Keane and Wolpin (1997)			
	blue collars (males)	0.215	-
	white collars (males)	0.247	-
	military (males)	0.120	-
Lee (2004)			
	blue collar (males)	0.094	-
	white colar (males)	0.096	-
Cohen-Goldner and Eckstein (2002)			
	blue collar (males)	0.07	0.15
	white colar (males)	0.08	0.10

Note: The return to experience reported is actually the return to a first year of experience

Table 5
Some Estimates of the Return to Schooling in the Experimental literature

	Sample	Parameter/std. error OLS	IV
Angrist and Krueger (1991)	**1970-80,US Census**		
	1920-29 cohort	0.070	0.101
		(0.000)	0.033)
	1930-39 cohort	0.063	0.060
		(0.000)	(0.030)
	1940-49 cohort	0.052	0.078
		(0.000)	0.039)
Card (1995)	**NLS young Men,1966 cohort**	0.073	0.132
		(0.004)	(0.05)
Lemiex and Card (2000)	**1971/81 Canadian Census**	0.070	0.164
		(0.002)	(0.053)
Stock and Staiger (1997)	1980 US Census men		
	1930-39 cohort	0.063	0.098
		(0.000)	(0.015)
	1940-49 cohort	0.052	0.088
		(0.000)	(0.015)