



Centre National
de la Recherche
Scientifique

GATE
Groupe d'Analyse et de Théorie
Économique
UMR 5824 du CNRS



DOCUMENTS DE TRAVAIL - WORKING PAPERS

W.P. 06-09

**The Return to Schooling
in Structural Dynamic Models:
A Survey**

Christian Belzil

October 2006

GATE Groupe d'Analyse et de Théorie Économique
UMR 5824 du CNRS
93 chemin des Mouilles – 69130 Écully – France
B.P. 167 – 69131 Écully Cedex
Tél. +33 (0)4 72 86 60 60 – Fax +33 (0)4 72 86 60 90
Messagerie électronique gate@gate.cnrs.fr
Serveur Web : www.gate.cnrs.fr

The Return to Schooling in Structural Dynamic Models: A Survey*

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

October 11, 2006

*This paper has been largely influenced by several discussions I have had with Zvi Eckstein. I thank him for many ideas and suggestions. Comments from three anonymous referees were very useful in preparing the last version of the paper. Discussions with James Heckman, Michael Keane, Thierry Magnac and Ken Wolpin are also gratefully acknowledged. Finally, several individuals, including Michael Bognanno, Barry Chiswick, Bruno Crepon, Donghoon Lee, Gordon Fisher, Jorgen Hansen, Susumu Imai, Thierry Kamionka, Winfried Koeniger, Francis Kramarz, Guy Laroque, Marco Leonardi, Lance Lochner, Claude Montmarquette, Peter Mueser, Robert Sauer, Kjell Salvanes and Uwe Sunde provided useful comments at various stages of the preparation of the manuscript. I thank them all.

Abstract

This paper contains a survey of the recent literature devoted to the returns to schooling within a dynamic structural framework. I present a historical perspective on the evolution of the literature, from early static models set in a selectivity framework (Willis and Rosen, 1979) to the recent literature, stimulated by Keane and Wolpin (1997), and which uses stochastic dynamic programming techniques. After reviewing the literature thoroughly, I compare the structural approach with the IV (experimental) approach. I present their commonalities and I also discuss their fundamental differences. To get an order of magnitude, most structural estimates reported for the US range between 4% and 7% per year. On the other hand, IV estimates between 10% and 15% per year are often reported. The discrepancy prevails even when comparable (if not identical) data sets are used. The discussion is focussed on understanding this divergence. The distinction between static and dynamic model specifications is a recurrent theme in the analysis. I argue that the distinction between the IV approach and the structural approach may be coined in terms of a trade off between behavioral and statistical assumptions. For this reason, and unless one has very specific knowledge of the true data generating process, it is neither possible, nor sensible, to claim which approach to estimation is more flexible. More precisely, I show that structural and IV approaches differ mainly at the level of i) the compatibility of the underlying models with truly dynamic behavior, ii) the role of heterogeneity in ability and tastes, iii) the consideration of post-schooling opportunities, and (iv) the specification (and interpretation) of the Mincer wage equation.

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

JEL Classification: **J2-J3.**

1 Introduction

The return to schooling is one of the most investigated parameters in modern economics. It plays a central role in microeconomic models of human capital accumulation and it is important in the empirical growth literature. The return to schooling is also at the center of the debate on the recent increase in wage inequality. 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, both of 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 “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. This objective has been at the forefront of the empirical labor economics literature for many years and, most likely, it will continue to be there for some time.

This paper contains a survey of the recent literature in which returns to schooling are estimated using structural dynamic programming techniques. The evolution of the structural literature is split into three fundamental stages. The first stage consists of seminal work on schooling and earnings by Becker (1964) and Mincer (1958). The second stage, which culminates in Willis and Rosen (1979), is largely influenced by the econometric self-selection literature (Heckman, 1976). Finally, the recent literature, stimulated by Keane and Wolpin (1997), uses stochastic dynamic programming techniques, and forms a third stage of the evolution of the structural literature. It is the focus of this survey.

In order to be consistent with the terminology found in the existing literature, I use the term “return to schooling” to refer to the causal effect of schooling on wages, even though it is only partially related to the concept of a rate of return on an investment; nevertheless, it is widely accepted in labor economics.¹ Until recently, the literature on the return to schooling had been completely dominated by studies using instrumental variable (IV) techniques or OLS regressions augmented with an observable measure of market ability.² However, the literature concerned with human capital accumulation, and based on structural stochastic

¹Issues surrounding the distinction between internal rates of return and wage growth rates are discussed in Heckman, Lochner and Todd (2005).

²In this paper, the expressions “experimental approach” and “IV approach” may be used interchangeably. The term “experimental” refers to the use of “Natural experiments” and has little to do with the recent literature on laboratory experiments.

dynamic programming (SSDP) techniques, has expanded rapidly in the past 10 years. It now offers a different perspective on the issues surrounding the measurement of returns to schooling and it provides a transparent illustration of several dynamic self-selection issues which are central to the analysis of education and skill formation policies. Obviously, the structural and the experimental approaches to estimation are based on completely different philosophies and are, at a technical level, fundamentally different.

Aside from these “ideological” differences, there is also a surprising discrepancy between the relatively high estimates of the return to schooling reported in the IV literature and those, much lower, obtained in a structural framework. To get an order of magnitude, most structural estimates obtained for the US range between 4% and 7% per year (Keane and Wolpin, 1997, and Belzil and Hansen, 2002). On the other hand, IV estimates between 10% and 15% per year are often reported. The discrepancy prevails even when comparable (if not identical) data sets are used. In general, structural estimates are not only smaller than IV estimates but also smaller than OLS estimates. For this reason, structural estimation has revived interest in the notion of “ability bias” and, in particular, 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.

At the outset, and while the focus of this survey is on the estimation of an outcome equation (more specifically, the return to schooling), it should be clear that the contribution of the structural approach to human capital accumulation models is not confined to what is reviewed below. Obviously, the structural approach identifies parameters which are not identifiable in the experimental literature. There is a wide range of issues such as the effects of liquidity constraints on schooling enrollment, risk aversion and 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 return to schooling are found in both structural and experimental papers. A comparison is therefore particularly enlightening.³

The distinction between static and dynamic model specifications is a recurrent theme in the analysis. In what follows, I argue that the distinction between the IV approach and the structural approach may be coined in terms of a trade off between behavioral and statistical assumptions. For this reason, and unless one has a very specific knowledge of the true data generating process, it is neither possible, nor sensible, to claim which approach to estimation is more flexible. More precisely, I show that structural and IV approaches differ mainly at the level of i) the compatibility of the underlying models with truly dynamic behavior,

³Indeed, it is very difficult to find another economic parameter that has been estimated as intensively as the return to schooling and for which the estimation method is also the object of such dispute.

ii) the role of heterogeneity in ability and tastes, iii) the consideration of post-schooling opportunities, and (iv) the specification (and interpretation) of the Mincer wage regression.

The paper is made of two fundamental parts. The first part is devoted to the structural literature itself and comprises the following elements. First, in Section 2, I present a short historical summary of the early literature on schooling and earnings, which dates back to Becker (1964) and Mincer (1958). In Section 3, I turn to the first generation of “structural models” of endogenous schooling decisions which emerged as a special case of the econometric literature on self-selection (Heckman, 1979). This literature is exemplified in a key contribution by Willis and Rosen (1979). After reviewing briefly the structure of empirical dynamic programming models which emerged in the 1980’s in Section 4, I devote Section 5 to the second generation of structural human capital accumulation models which focussed on post-schooling investments and assumed that schooling was pre-determined (Eckstein and Wolpin, 1989 and 1995). Subsequently, in the core of the paper (Section 6), I review a set of papers devoted to the modeling of schooling decisions in a stochastic dynamic programming setting and stress their specificities as well as their estimates of the return to schooling. These include Keane and Wolpin (1997, 2000 and 2001), Eckstein and Wolpin (1999), Belzil and Hansen (2002 and 2005), Magnac and Thesmar (2002b). In section 7, I examine three contributions set in a general equilibrium framework (Heckman, Lochner and Taber, 1998; Lee, 2005; and Lee and Wolpin, 2006).

The second part of this survey is devoted to a comparison of the IV approach (the most popular approach in empirical labor economics) and the structural approach. In Section 8, I discuss and analyze their commonalities as well as their key differences. This leads to Section 9 which is devoted to the potential reasons why structural and reduced form estimates may be different. Finally, some avenues for future research are identified in the Conclusion.

2 The Return to Schooling and the Mincer Equation

The correlation between education and wages has been analyzed in seminal pieces by Becker (1964, 1967) and Mincer (1958, 1974). At the time, the fundamental models underlying schooling choices were all based on a lifecycle income maximization hypothesis. In its simplest form, the early Becker-Mincer model assumes that individuals maximize lifetime earnings, $V(s)$, by choosing the optimal level of schooling s , that is

$$V(s) = w(s) \int_s^T \exp(-r \cdot t) dt \quad (1)$$

where T is the time horizon (it may be finite or infinite), $w(s)$ denotes the earnings of an individual with a schooling level s , and r is the subjective rate of interest. At that time, it was customary to assume that individuals choose the optimal level of schooling without any consideration for post-schooling human capital accumulation opportunities or, alternatively, that post-schooling wage growth is exogenous. It was also common to assume, mostly for expository convenience, that individuals earn nothing while in school. In this framework, the relationship between wages and schooling is understood as a compensating wage differential, emerging from individual differences in discount rates.

The first formal representation of the human capital accumulation process as an intertemporal optimization problem is due to Ben-Porath (1967). Ben-Porath assumed competitive labor markets and, more precisely, that individuals are paid their marginal product according to their individual specific level of skills. In his model, 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

$$W_t = P_t \cdot K_t \quad (2)$$

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).

A quick glance at equation (2) reveals the inherent identification problem that underlies any analysis of human capital theory. Both human capital (K) and skill prices (P) are fundamentally unobservable. The Mincerian approach assumes that human capital accumulation (or skill acquisition) is rendered possible by combining inputs such a time spent in school, time spent in the labor market, and innate ability.

It is the production function representation of the relationship between accumulated skills (as an output) and education, experience and abilities (as inputs) that gave rise to the celebrated Mincer wage regression (Mincer, 1974);

$$\log W_t = \varphi_0 + \varphi_1(S_t) + \varphi_2(Exp_t) + \varepsilon_t \quad (3)$$

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

At this stage, there are two important notes to be made. First, in a competitive framework, the parameters of the Mincer equation should be given a productivity interpretation. Obviously, in markets that have different configurations (search equilibrium, incentive models, etc.), the interpretation must be modified. Second, in more general (and more realistic) models within which post-schooling behavior is endogenous, it will be convenient to rethink $\varphi_2(Exp_t)$ as

containing several endogenous variables.

By the early 1970's, the estimation of the return 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 (or proxies thereof) measuring unobserved ability.⁴

More interestingly, Griliches recognized that the endogeneity of schooling decisions, virtually ignored until then, was a serious problem which might have prevented economists from uncovering the true causal effect of education on earnings. At the econometric level, the main issue may be illustrated by the following simplified version of the Mincerian wage equation,

$$\log W_i = \beta_0 + \beta_1 \cdot \text{Schooling}_i + \eta_i \quad (4)$$

where η_i is unobserved market ability.

Ignoring post-schooling labor market experience, the discrepancy between OLS and IV estimates is a reflection of the correlation between schooling and ability η_i and gives rise to the celebrated notion of ability bias. The ability bias hypothesis was based on the rather intuitive idea that those who are more able (those endowed with a high value of η_i) are also more likely to be more educated. That is when $\text{cov}(\text{Schooling}_i, \eta_i) > 0$, the OLS estimate of the return to schooling is upward bias. However, as pointed out by Griliches, those individuals endowed with higher market ability also face a higher opportunity cost of attending school. For this reason, a negative correlation between schooling and ability (a negative ability bias) is also conceivable. Perhaps at a more fundamental level, (4) does not account for the possible existence of multiple skills and, in particular, ignores the importance of comparative advantages in the labor market. Nevertheless, the points raised by Griliches stimulated a large number of empirical studies aimed at tackling the endogeneity of schooling.⁵

⁴For instance, in the National Longitudinal survey of Youth (NLSY), some individuals were administered the Armed Force Qualification Tests (AFQT), which is a test that measures basic quantitative and analytical skills.

⁵Technically speaking, the term "endogeneity" used in the empirical literature abuses the true meaning of endogeneity. In cross-section data, wages are usually measured much beyond the time when schooling is completed. Because returning to school is rarely observed, schooling may therefore be viewed as pre-determined. It is only the correlation between schooling and unobserved skills that is problematic.

3 Static Models of Schooling and Employment

In the second half of the 1970's, the endogeneity of schooling was widely accepted. As is the case for all econometric models plagued with endogenous variables, there are two alternative approaches to estimating the return to schooling. One branch of the literature, now known as the “Experimental” literature, adopted as its fundamental paradigm the reliance on exogenous variables that are correlated with schooling but uncorrelated with the error term of the wage equation. Ultimately, the analyst “instruments out” schooling so to obtain independent variation.⁶

A second approach, more in line with economic theory, 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”. At that time, economists developed a first vintage of “structural” models of schooling and earnings in which the utility equations were ad hoc approximations of the maximization problem faced by the agent. In modern terminology, it is probably more accurate to refer to this stream of the literature, which I will be discussing now, as “semi-structural”.⁷ Despite its intertemporal aspect, the lifetime earnings maximization paradigm behind most studies was inherently static. That is, economists paid no attention to modeling optimal post-schooling investments and ignored the description of the law of motion that maps current schooling choices into future opportunities.

3.1 The Self-Selection Literature

To a large extent, the new (semi-) structural literature on schooling and earnings was largely an offshoot of the contemporaneous progress in computational facilities and also relied on some important results in the econometrics of self-selection models (Heckman, 1979). The key contribution of the self-selection literature was to provide a statistical framework (based on economic theory) to analyze samples in which the observability of each individual unit was governed by that individual's decision (for example, a sample of college graduates).

The self-selection literature on schooling was also fundamentally static and focused on the decision to attend college. At the time, it was already recognized that measuring the return to education was complicated by self-selection. Obviously, the fact that individuals with very low wages (typically with low levels of schooling) are underrepresented in a standard cross-section survey of the la-

⁶As we will see later, a wide range of empirical papers using instrumental variable (IV) techniques have been published since then.

⁷In the econometrics literature, the term “structural” is widely used, but often designates different modeling strategies. Further, there seems to be no formal definition of the term “semi-structural”.

bor force participants, implies that the correlation between wages and schooling may diverged from the true one.⁸ Furthermore, many researchers questioned the validity of the (positive) ability bias and pointed out its superficial content.

The literature of the late 70's constitutes a first step toward the comprehension of the selectivity process that governs skill acquisition and schooling decisions. Many contributions may be imbedded in the model of comparative advantage developed by Roy (1951). Consider an individual contemplating college attendance. That individual is endowed with the following wage equations:

$$\log W_{i,hs} = X_i\beta_{hs} + \varepsilon_{i,hs} \quad (5)$$

$$\log W_{i,c} = X_i\beta_c + \varepsilon_{i,c} \quad (6)$$

where $W_{i,hs}$ is the wage if the individual stops after high-school graduation and $W_{i,c}$ is the potential wage as a college graduate. The error terms $\varepsilon_{i,hs}$ and $\varepsilon_{i,c}$ represent unobserved abilities in jobs requiring high school training and college training respectively and X_i is a vector of individual characteristics. The optimal decision is represented by the following latent structural equation

$$S_i^* = Z_i\beta_S + \varepsilon_{i,S} \quad (7)$$

where $S_i = 1$ when $S_i^* > 0$ and 0 if not. In most related applications, it is assumed that $\varepsilon_{i,hs}$, $\varepsilon_{i,c}$ and $\varepsilon_{i,S}$ follow a multi-variate normal distribution and the vector Z_i contains all elements of X_i plus other regressors that would affect the discount rate (say parental income).⁹

The self-selection model, as applied to the schooling decision process, was a major advance. In particular, it allowed labor economists to rethink the return to schooling within a framework that no longer views it as a single parameter. This is readily seen upon comparing (5) and (6) with (4). For a randomly selected individual, the return to schooling is the difference between (6) and (5) and it is affected both by observed and unobserved heterogeneity. In a simpler framework such as (4), it is only a single parameter.

3.2 Willis and Rosen (1979): Education and Self-Selection

The semi-structural approach is exemplified by Willis and Rosen (1979), who extend the Roy model to incorporate an exogenous individual/occupation-specific growth rate. Willis and Rosen start with a model of life time earnings, denoted Y_{ij} , where i is the individual subscript and j is the education level achieved by

⁸In this specific case, the OLS estimate of the effect of schooling would under-estimate the true return to schooling and the bias is contrary to the standard "Ability Bias". See the Nobel Prize Public Report on James J. Heckman and Daniel L. McFadden (2000).

⁹Nowadays, the parametric estimation of a system of equations such as (5), (6) and (7), has become highly controversial.

the individual (either high school or college). Willis and Rosen were the first to model schooling decisions in a non-hierarchical structure which allow explicitly for the existence of comparative advantages.

They assume that

$$Y_{ij} = \bar{Y}_{ij} \cdot \exp(g_{ij} \cdot X_i) \quad (8)$$

where \bar{Y}_{ij} is initial earnings and g_{ij} is individual i 's earnings growth under regime j . The authors assume that both the initial earnings function and the growth rate are represented by a linear equation:

$$Y_{ij} = \beta_j \cdot X_i + U_{ij}^Y \text{ for } j = 1, 2 \quad (9)$$

and

$$g_{ij} = \delta_j \cdot X_i + U_{ij}^g \text{ for } j = 1, 2 \quad (10)$$

where U_{ij}^Y, U_{ij}^g are individual specific error terms representing cross-sectional heterogeneity.

The optimal schooling decision is obtained by assuming that individuals maximize lifetime earnings and that each individual is endowed with a subjective discount rate, r_i , given by the following equation

$$r_i = Z_i \cdot \delta + U_{ir} \quad (11)$$

where the vector Z_i contains regressors that would affect the discount rate (say parental income). Willis and Rosen assumed U_{ij}^Y, U_{ij}^g and U_{ir} to be jointly normal. The schooling decision equation is a latent (single) index equation, like (7), which follows from the lifecycle income maximization hypothesis.

Willis and Rosen estimated their model on a sample of World War II veterans. They found evidence in favor of comparative advantages (positive sorting). That is individuals who go to college have higher lifetime earnings in jobs requiring college than those who did not, and also find that those who did not enroll in college have higher life time earnings in jobs requiring high school training than those who enrolled in college, had they decided to work in high school type occupations. They also find that Mathematics test scores significantly increase the earnings of college educated workers but have no effect on earnings of high school graduates.

Willis and Rosen's contribution was remarkable in many respects, but in particular, it pointed out that in the presence of multiple skills (two skills in this specific instance), the notion of ability bias was much more complex than in a single skill (hierarchical) representation. In particular, when $cov(\varepsilon_{i,hs}, \varepsilon_{i,c}) < 0$, individuals who self-select into low skill jobs may achieve higher lifetime earnings than they would have if they had chosen to attend college. However, as does

all the literature based on self-selection models, their analysis was plagued by the difficulty to infer counterfactual choices. For this reason, they could not obtain an estimate of the covariance between $\varepsilon_{i,hs}$ and $\varepsilon_{i,c}$. Nowadays, identifying correlations between unobservables that generate counterfactual outcomes is still a major challenge.¹⁰

4 Estimable Stochastic Discrete Dynamic Programming Models

In this section, I present the general structure of stochastic dynamic discrete programming models. The models surveyed in Sections 5, 6 and 7 belong to the general class of SSDP models which I will now briefly describe. The reader who is familiar with the details of structural estimation in a dynamic framework, may skip this section and go directly to Section 5.

4.1 The General Structure

To a large extent, the development of dynamic macroeconomic theory that took place in the 1970's stimulated interest in the estimation of dynamic microeconomic models in the 1980's. As more and more economists became familiar with stochastic dynamic programming techniques, and with regular improvements in computer power, the development of empirical methods for the estimation of dynamic discrete choice models was inevitable. The introduction of recursive methods in stochastic economics is typically associated with (Bellman, 1957), although the first known example of modern dynamic programming is found in Massé (1946).¹¹ Nowadays, stochastic dynamic programming belongs to the standard toolkit necessary to the completion of graduate training in economics. This is largely due to the revolution in macroeconomics initiated by the work of Robert Lucas.¹²

At a first stage, econometricians estimated relatively simple discrete dynamic programming models such as optimal stopping models but they gradually moved to models with a relatively high number of choices (Rust, 1987; Pakes, 1986; and Wolpin, 1984).

Structural stochastic dynamic programming models are based on the fundamental idea that agents are forward looking. In the structural approach, the

¹⁰The recent contribution by Carneiro, Hansen and Heckman (2003) constitutes a major breakthrough in the field.

¹¹This was actually pointed out to me by the late Gerard Debreu. I would also like to thank Kenneth Arrow for providing me with some details on Pierre Massé.

¹²Classical references on dynamic programming include Bellman (1957), Whittle (1982) and Stokey, Lucas and Prescott (1989).

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 IV or 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).¹³

Consider an individual facing $k = 1, 2, \dots, K$ possible actions and maximizing the expected discounted lifetime utility over a finite horizon going from $t = s, s+1, \dots, T$. Define the set containing all state variables known by the agent at t as Ω_t . Lifetime utility is assumed to be time separable and the per-period utility of choosing option k , $U_k(t)$, may depend on the choice (subscript k) made by the individual at time t or it may depend on t explicitly. In general, $U_k(t)$ depends on Ω_t but I suppress the dependence to simplify notation. The control variable is $d_{kt} = 1$ when k is chosen and 0 if not and β is the discount factor. The optimal value function (Bellman, 1957) is

$$V_t(\Omega_t) = \text{Max}_{d_{kt}} E \left[\sum_{s=t}^T \beta^{s-t} \cdot \left(\sum_{k=1}^K U_k(s) \cdot d_{ks} \right) \mid \Omega_t \right] \quad (12)$$

As is standard in the literature, assume that Ω_t contains a vector of state specific random shocks (ε_{tk} for $k = 1, 2, \dots, K$) and that these ε_{tk} 's 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). As structural models are typically based on the Rational Expectation hypothesis, there is no distinction to be made between subjective and objective probability distributions.

In many applications, the per-period utility, $U_k(t)$, is additively separable with respect to the random shock. If so, it is the sum of a deterministic component, $\bar{U}_k(t)$, and a state specific utility shock, ε_{tk} , and the state specific value function is then written as

$$V_{kt}(\Omega_t) = \bar{U}_k(t) + \varepsilon_{tk} + \beta E \max \{ V_{1,t+1}(\Omega_{t+1}), \dots, V_{K,t+1}(\Omega_{t+1}) \mid d_{kt} = 1 \}$$

or, more compactly, as

$$V_{kt}(\Omega_t) = \bar{U}_k(t) + \varepsilon_{tk} + \beta E V_{t+1}(\Omega_{t+1} \mid d_{kt} = 1)$$

Relaxing additive separability is possible, but it may entail additional computational problems.

¹³Within 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.

The dynamics are contained in the parametric relationship (denoted $F_t(\cdot)$), which maps current choices (d_{kt}) and current state variables (Ω_t) onto future state variables (Ω_{t+1}), namely

$$\Omega_{t+1} = F_t(\Omega_t, d_{kt}) \quad (13)$$

and where $F_t(\cdot)$ is a distribution function. The Markovian property of (13) may also be relaxed, but only at relatively high computational cost. The expected value in (12) is taken with respect to the future ε'_{kt} s. The order of causation is therefore

$$\Omega_s \rightarrow d_{ks} \rightarrow \Omega(s+1) \rightarrow d_{k,s+1} \rightarrow \dots \Omega(T) \rightarrow d_{kT} \quad (14)$$

In many schooling models where individuals maximize lifetime earnings, the Mincer wage equation is at the same time part of the law of motion and a key element of the per-period utility of working. In words, (13) implies that the wage offered at the beginning of period $t+1$ (which is one element of Ω_{t+1}) is a stochastic function of both the accumulated level of human capital at the beginning of the previous period (schooling and post-schooling human capital accumulated by the beginning of t) and the choice made last period (whether one worked or attended school in t).

The dynamic discrete choice model (when incorporating a large number of options), is general enough to incorporate schooling, on-the-job training, labor supply, or even job search efforts. So, from now on, we think of the Mincer equation as follows:

$$\log W_t = \varphi_0 + \varphi_1 \left(\sum_{j=1}^t d_{k=S,j} \right) + \varphi_2 \left(\sum_{j=1}^t d_{k \neq S,j} \right) + \varepsilon_t \quad (15)$$

where S is the index that identifies schooling as the actual choice exercised in period t .

4.2 Dynamic Self-Selection

Dynamic self-selection may be introduced by assuming that

$$U_{ik}(t) = \bar{U}_k(t, \omega_{ik}) + \varepsilon_{it}^k \quad (16)$$

where ω_{ik} is a time invariant individual-specific taste for choice k , and/or by specifying an individual specific law of motion:

$$\Omega_{i,t+1} = F_{it}(\Omega_{it}, d_{i,k,t}, \eta_i) \quad (17)$$

where η_i is an individual specific effect which influences labor market outcomes. 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 accu-

mulation process. These are permanent and cannot be observed by the econometrician, but are assumed to be known by the agent. This amounts to an extreme form of Rational Expectation.

In the context of a schooling decision problem, ω_{ik} may play the role of unobserved taste for schooling, η_i could be labor market ability, and F_{it} would represent the Mincer wage equation. It will be convenient to assume that $\{\bar{\omega}_i^k, \eta_i\}$ are jointly distributed with distribution function $G(\cdot)$. Note that the introduction of $\{\bar{\omega}_i^k, \eta_i\}$ into the analysis involves some ad-hoc decisions. In the example of the Mincerian wage equation, η_i may affect log wages additively or it may affect slopes as well. The former approach is in line with the recent microeconomic literature on treatment effect.¹⁴

4.3 Identification

In the econometric literature, non-parametric identification results are well known for certain types of reduced-form dynamic models such as Proportional Hazards. In the structural literature, identification is typically considered within a particular parametric structure. As of now, the common view is that structural models require to specify preferences and technology and therefore imply functional forms and parametric assumptions. The degree of under-identification (non-parametric) is analyzed precisely in Hotz and Miller (1993), Rust (1994) and Magnac and Thesmar (2002) although all of them consider the identification of generic models in which data on outcomes are ignored.

To see the issue, it is convenient to examine the representability theorems introduced in Hotz and Miller (1993) and developed further in Magnac and Thesmar (2002).

Dropping the time subscript, define the vector of choice probabilities \bar{P} as

$$\bar{P} = \{\Pr(d_1 = 1), \Pr(d_2 = 1), \dots, \Pr(d_K = 1)\}$$

and the vector \bar{V} containing the non-stochastic parts of the value functions as

$$\bar{V} = \bar{V}_1, \bar{V}_2, \dots, \bar{V}_K$$

where

$$\bar{V}_k = \bar{U}_k(\cdot) + \beta EV$$

It may be shown that, given some normalizations, there exists a one-to-one relationship between value functions and choice probabilities; that is

$$\bar{V}_k = \bar{V}_K + q_k(\bar{P}; G(\cdot)) \tag{18}$$

¹⁴This approach is followed in Belzil and Hansen (2006) as well as in Cohen-Goldner and Eckstein (2002 and 2005).

where $G(\cdot)$ is the distribution of the ε'_k s and where $q_k(\cdot)$ is a function mapping \bar{P} onto \bar{V}_k , and which is determined by $G(\cdot)$.

Basically, (18) states that given a reference choice (\bar{V}_K), the value functions are invertible in choice probabilities. However, invertability requires the specification of the distribution function $G(\cdot)$ and therefore implies that there exists a function $\phi(\cdot)$ such that

$$\bar{V}_1, \bar{V}_2, \dots, \bar{V}_{K-1} = \phi(\bar{P}; \bar{V}_K) \quad (19)$$

Unless some additional restrictions are considered, equation (19) implies that estimation of the parameters of dynamic programming models which use data on discrete choices only requires (i) a specification of the error shock ($G(\cdot)$), (ii) a reference choice and, (iii) setting the discount rate.¹⁵

As the structure of each model is specific, it is impossible to give a general recipe for identification. For instance, in many examples reviewed below, individuals are assumed to maximize earnings (Keane and Wolpin, 1997) or the logarithm of earnings (Belzil and Hansen, 2002). Along with some terminal value conditions, these assumptions imply that the utility of labor market work may be inferred from data on labor market outcomes and that estimations of the parameters for all K options may be performed.

Most empirical papers devoted to human capital use data on choices (schooling and employment) and panel data on labor market outcomes (mostly wages and employment outcomes). The use of panel data is not innocuous. In many applications, the correlation between $\bar{\omega}_i^k$ and η_i becomes a key element of the model. In many papers where identification is discussed informally (Keane and Wolpin, 1997, Eckstein and Wolpin, 1999, and Belzil and Hansen, 2002), data on post-schooling outcomes is typically mentioned as playing an important role in the identification of some of the parameters of the model, and especially, the correlation between unobserved tastes for specific choices and labor market outcomes (wages).

Estimation is usually performed using maximum likelihood techniques or their simulated counterparts. In most of the cases, this requires solving a number of moment conditions equal to the number of parameters. Obviously, estimation requires identifiability of the particular parametric structure. Typically, as with the 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, it is sometimes possible to search a larger parameter space for other local maxima.

Finally, it should be noted that the identification results stated earlier ignore conditions that may be imposed on utility functions such as those used in the

¹⁵This argument applies to non-parametric identification. The discount factor may still be estimated, given (parametric) identifiability. Indeed, in many studies, it is estimated (not fixed).

static discrete choice literature. Indeed, in recent work, Heckman and Navarro (2006) have developed a set of assumptions in which it is conceivable to identify a dynamic discrete choice model semi-parametrically. Their results imply that, under certain conditions, estimation of a structural model does not require a parametric distribution for the error shocks.¹⁶ The identification proof is largely influenced by results circulated in the static discrete choice literature (Matzkin, 1994, 2004) and also uses “identification in the limit” arguments. Data on outcomes also play a specific role in the identification proof, which relies heavily on factor structure representations discussed in Carneiro, Hansen and Heckman (2003).

4.4 Estimation

Using recursive techniques, estimation of the parameters of $U_k(\cdot)$, F_t , and potentially β , may be achieved by various method-of-moment estimators (including maximum likelihood). 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

$$\Pr(d_{kt} = 1) = \Pr\{U_{kt}(\cdot) + \beta EV_{t+1}(\cdot \mid d_{kt} = 1) \geq U_{st}(\cdot) + \beta EV_{t+1}(\cdot \mid d_{st} = 1)\} \text{ for all } s \neq k \quad (20)$$

In the case where $U_{kt}(\cdot)$ is linear in the $\varepsilon'_{tk}s$, this expression takes the form of a non-linear discrete choice equation and boils down to the probability that the ε_{kt} exceeds a threshold value.

However, even in such a case, technical difficulties arise in the evaluation of the $EV_{t+1}(\cdot)$. When the model is set in a finite horizon framework, the solution method will be based on recursive techniques (Bellman, 1957). In general, no explicit expression exists for $EV_{t+1}(\cdot)$, and numerical methods need to be used, but, in specific cases, an analytical form may be available.¹⁷

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

¹⁶Their argumentation is developed in the context of an optimal stopping problem.

¹⁷There is indeed a relatively large literature devoted to solution methods. For a discussion, see Rust (1997), Keane and Wolpin (1994), Hotz and Miller (1993) and Eckstein and Wolpin (1989,b). Analytical form solutions may exist when the number of states is limited such as in optimal stopping models or when the random terms follow an extreme value distribution (Rust, 1987).

$$\Pr\{(d_{ik1}, w_{i1}), \dots, (d_{ikT}, w_{iT})\} = \Pr(d_{ik1}, w_{i1}) \cdot \dots \cdot \Pr(d_{ikT}, w_{iT}) \quad (21)$$

where the particular form of $\Pr(d_{ikt}, w_{it})$ will differ according to the specificity of the model considered.

In virtually all models allowing for permanent unobserved heterogeneity, it is assumed that individual unobserved (to the econometrician) skills are fully known by the optimizing agent.¹⁸ This implies forming the following mixed likelihood function:

$$L(\cdot) = \int \prod_{t=1}^T \Pr(d_{kt}, W_t \mid \bar{\omega}_i^k, \eta_i) dG(\bar{\omega}_i^k, \eta_i) \quad (22)$$

where $\theta(\cdot)$ is a set of parameters to be estimated. In practice, (22) is maximized with respect to a discrete approximation of $G(\cdot)$ with a known, fixed, 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;¹⁹

$$L(\cdot) = \sum_{m=1}^M \prod_{t=1}^T \Pr(d_{kt}, W_t \mid \text{type } m) \cdot \Pr(\text{type } m) \quad (23)$$

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

5 Dynamic Labor Supply with Pre-determined Schooling

Until the second half of the 80's, the literature on schooling was mostly confined to static models estimated from cross-sectional data. At the same time, and as mentioned previously, the earlier structural dynamic microeconomic models of the labor market did not focus on human capital per se. In the early 80's, many countries experienced high unemployment rates, and several labor economists paid particular attention to the search behavior of the unemployed within an optimal stopping framework. At a technical level, optimal job search behavior is the simplest dynamic programming model that can be estimated, but it may also be regarded as one of the most relevant forms of post-schooling human capital investment.²⁰

¹⁸However, in the case where post schooling outcomes are modeled within a search-matching framework (such as in Eckstein and Wolpin, 1995), match-specific productivity is not known ex-ante.

¹⁹This approach is in the spirit of Heckman and Singer (1984).

²⁰More precisely, while search activities may not enhance productivity, they allow individuals to maximize post schooling wage growth. For this reason, search activities may be regarded as a particular form of human capital investment. See Flinn and Heckman (1982) for a classic

Because computation time was still a major constraint by the end of the 80's, those who investigated post schooling investments (labor supply, experience accumulation or search behavior) treated schooling as predetermined. In terms of equation (12), this implies that most researchers set s at the time of labor market entrance and include schooling achievements in Ω_s . I now review two of these contributions; Eckstein and Wolpin (1989, 1995), hereafter EW(89) and EW(95).

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

Eckstein and Wolpin (1989) estimate a structural dynamic model of labor force participation of married women which embodies intertemporal substitution through preferences and constraints. Their model analyzes time allocation between 2 options, work and leisure (or home production), from age 45 until retirement. The law of motion is a Mincer equation with endogenous work experience.

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 habit persistence. 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 affects future labor force participation. 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 an intercept term in the instantaneous utility of work). 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.

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.

EW (89) is an important paper in the literature. It is the first structural dynamic programming model of human capital accumulation set in a Mincerian example. The literature is surveyed in Eckstein and van den Berg, 2005.

framework and estimated from micro-data. This is also the first example of the use of structural estimation in order to correct for heterogeneity bias, although it does not appear to be one of the main objectives of the paper. There is therefore a similarity between the paper and the earlier literature on employment and self-selection (Heckman, 1979).

5.2 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 returns to schooling within a search framework.²¹ 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, education is exogenously determined. They assume that firms and workers meet randomly (given a level of effort endogenously determined by both firm and 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.²²

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. It is important to note that the return to schooling is estimated from data on the first job only. 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. Offered wages, on the other hand, are more dependent on market fundamentals (productivity, discrimination, etc.). 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 (the interest rate that equates the present value of the constant streams of offered wages for each incremental schooling level) on both accepted and offered wages. In general, those computed with offered wages (ranging from 6% to 27% for white males) are higher than those computed with accepted wages (ranging from 5 %

²¹Wolpin (1992) 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. As he does not report estimates of the returns to schooling, it is not surveyed here.

²²More precisely, in order to fit the model, they impose a symmetry condition on firm/worker solutions.

to 12%). These differences illustrate the impact of the reservation wage on the return to schooling measured from potential (offered) wages as opposed to the return measured from accepted (observed) wages.

Again, it is possible to draw a parallel between Eckstein and Wolpin's contribution and the early self-selection literature discussed in Section 3.1. Just as the under-representation of low wage workers in a standard cross-section survey of labor force participants is likely to affect estimates of the effect of schooling on market productivity, so the presence of unobserved heterogeneity in search costs (and reservation wages) and the existence of search frictions is likely to distort the correlation between true labor market productivity and schooling, when inference rests solely on accepted wages. The distinction between offered and accepted wages has not received much attention either in the structural or the experimental literature. As we now enter the next section focussing on the recent structural dynamic programming literature devoted to endogenous schooling, we will see that virtually all models are set in a competitive framework. This will imply an automatic correspondence between observed wage growth and human capital enhancement.

6 Dynamic Models of Schooling and Employment Choices

In the 1990's, applied econometricians started to implement dynamic discrete choice models of schooling decisions within a fully structural framework. The development of econometric techniques for the estimation of dynamic programming models allowed economists to re-think the Becker-Mincer paradigm of human capital accumulation in a framework where schooling was treated truly as a discrete choice and where uncertainty played a key role. After all, the dynamic programming representation of the human capital accumulation process obviously embeds the Becker-Mincer model as a special case.

Most of the contributions discussed below allow for a broad set of options which, in all cases, incorporate the decision to attend school, but also other decisions such as employment/occupation choices, savings decisions, the decision to work while in school. In virtually all cases, the law of motion incorporates a standard Mincer equation plus other additional features depending on the structure of the model. Finally, most contributions are characterized by a rich heterogeneity specification and reserve an important role for dynamic self-selection. To a large extent, most of these contributions may be seen as extensions of Willis and Rosen's static schooling decision model.

6.1 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 occupational choices (blue collar, white collar and military occupations). The model is applied to 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 (blue collar, white collar and military occupations). Second, experience accumulation is endogenous; that is, individuals choose to work or not and the type of work experience to accumulate. The authors estimate occupation specific returns to schooling, in a context where both schooling and occupation 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

$$\log W_{imt} = \varphi_{m1} \cdot S_{it} + \varphi_{21} \cdot Exp_{imt} + \varphi_{22} \cdot Exp_{imt}^2 + \eta_{im} \quad (24)$$

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_{im} refers to individual i 's occupation-specific unobserved skills by age 16. The regression function therefore allows for heterogeneity in slopes, although the heterogeneity is only allowed through specific occupations.

Keane and Wolpin postulate that the per-period utility of attending school is additive in the stochastic error term and assume that the per-period utility of work is the wage rate.²³ The utility of attending school is denominated in the monetary equivalent. There exist no closed-form solution to the value functions. Repeated numerical solutions of the value functions must therefore be used. Given the complexity, KW develop an approximation method which is based on interpolation methods which use simple OLS regression techniques.

Basically, their approach is the following. Select a subset of all possible elements of the state space in the last period. Given these elements, it is straightforward to compute the relevant $E_{max}(\cdot)$ function using simulation methods applicable to static discrete choice.²⁴ For those elements at which the $E_{max}(\cdot)$ is not evaluated, use an OLS regression of the simulated $E_{max}(\cdot)$'s on the state space elements, and use a predicted value (an interpolation). As a second step, go to the previous period and select a new set of points in the state space. Using simulation methods compute the $E_{max}(\cdot)$ for the current period. To do this, KW need the future components of the choice specific utilities. These components were either computed by simulation methods at the first step, or they may

²³This means that the utility of work is not additive in the wage error shock.

²⁴They actually use the GHK (Geweke-Hajivassiliou-Keane) simulator.

be approximated using the interpolation technique. The algorithm is repeated until the initial period (age 16).²⁵ Furthermore, in the presence of unobserved heterogeneity, this procedure must be replicated a number of times equal to the number of types considered in the analysis.

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 (22). It is also assumed that the population distribution is approximated by a discrete distribution with a (known) fixed number of types which is set 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.²⁶

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 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 negative, as well as 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.²⁷ Interestingly, individual endowments explain more than 80% of the variance of lifetime wealth.

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 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. Interestingly, KW also considered potential non-linearities (introducing high-school and college graduation dummies), but report those to be unimportant.

The contribution of Keane and Wolpin is remarkable in many respects. Indeed, it is most probably the most important contribution to the empirical schooling literature since Willis and Rosen (1979). It is the first structural paper that stresses that there is no such thing as a single return to schooling and to find relatively low returns to schooling (compared to the IV literature). This is espe-

²⁵This is discussed in detail in Keane and Wolpin (1994).

²⁶This is a way to take into account the endogenous initial conditions.

²⁷This is found in Table 9, page 502.

cially true for blue collar occupations. Perhaps more importantly, KW are the first to estimate the importance of persistent individual endowments in explaining schooling attainment and to provide a formal variance decomposition of lifetime utility (or earnings) in terms of individual skill endowments (at age 16) and accumulated human capital. Indeed, the fact that individual endowments explain so much of the variance of lifetime wealth also implies that college attendance subsidies would be ineffective as a mean to reduce life cycle inequality.

6.2 Keane and Wolpin (2000): Racial Differences in Wages and Schooling Attainments

In their 2000 paper, KW investigate the black/white differential in schooling attainments and show that the schooling/occupation choice model may also be used to fit the behavior of young black males. As the model is structured like KW(97), there is no need to describe it any further. The model is fitted on a sample of black and white males taken from the NLSY 79. Basically, the authors find that by allowing for different skill endowments at age 16 (allowing for different type proportions) and for different skill rental prices for blacks and whites (a racial indicator binary variable in the Mincer wage function), their model is capable of explaining schooling attainments of young blacks. The return to schooling in blue collar occupations (0.020) and white collar occupations (0.071) are quite comparable to those reported in their 1997 paper. They conclude that differences in initial endowments along with racial discrimination can explain the relatively low schooling attainments of blacks and that there is no evidence that young blacks fail to behave as forward looking agents. An interesting result is that when discount rates are estimated separately for blacks and whites, they are quite similar (around 0.93 per year).

6.3 Keane and Wolpin (2001): The Role of Parental Transfers and Liquidity Constraints in Schooling Decisions

In Keane and Wolpin (2001), the authors investigate the extent to which differences in parental transfers account for the positive intergenerational correlation in educational attainment. In their model, young individuals make decisions about school attendance, work participation and consumption. They allow for the possibility of part-time activities in work and college attendance. Estimation is performed using data on schooling, wages, assets, work, parental co-residence and parental education.

In their model, the intergenerational education correlation is potentially explained by several mechanisms. First, individuals endowments at age 16 are potentially correlated with parents' education. Second, parents provide financial transfers to young individuals, according to an exogenous rule which depends on

school attendance. The model also assumes that net assets must exceed a lower bound and that the lower bound is more binding as assets approaches 0. The model therefore allows for the critical importance of parental transfers.

While the focus of the paper is obviously not on the return to schooling, the key findings still have implications for the return to schooling. The return to schooling is found to be 0.075. As we will see later, those individuals estimating the return to schooling by IV techniques often justify their high returns to schooling by the fact that there may be a substantial sub-population of individuals who have relatively high returns to schooling but cannot attend because of liquidity constraints. However, Keane and Wolpin find that the return to schooling is still relatively low when allowing for liquidity constraints; but liquidity constraints appear to play no role. This finding is also consistent with results reported in Cameron and Taber (2004), in which the authors investigate the presence of borrowing constraints using structural methods as well as IV techniques.²⁸

6.4 Belzil and Hansen (2002): The Convexity of the Wage-Schooling Relationship and the Ability Bias

Belzil and Hansen (2002) use a dynamic programming model to investigate the shape of the wage schooling relationship and evaluate ability bias as defined by the correlation between schooling achievement and the individual-specific intercept term of the wage equation. 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, there is no distinction between occupation and sectors. The model may therefore be viewed as a single skill model.

The model is implemented on a panel of white males taken from the 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

$$\log W_{it} = \varphi_1(S_{it}) + \varphi_{21} \cdot Exp_{it} + \varphi_{22} \cdot Exp_{it}^2 + \eta_i \quad (25)$$

where $\varphi_1(\cdot)$ is left unspecified and is estimated flexibly with spline techniques. Local returns are allowed to differ from grade level 8 to 17.²⁹

The utility of attending school is parameterized in a function of parents' background variables (father and mother's education, family income, number of siblings, regional dummies and family status indicating whether the individual was

²⁸One of the tests proposed in Cameron and Taber (2004), is to compare returns to schooling obtained using instruments affecting opportunity costs with returns obtained from instruments affecting direct costs.

²⁹This is the case because in the sample analyzed, virtually everyone has at least 7 years of schooling.

raised by both parents). However, the utility of work (the log wage rate) is not a function of the parents' background variables, after conditioning on schooling.³⁰ 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 $k = 1, 2 \dots K$ 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 schooling choices are modeled 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 potential ability bias, type specific simulations of schooling decisions and wage histories are performed. The correlation between unobserved market ability and realized schooling 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 marginal returns are generally increasing with the level of schooling up to grade 14. 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 increases to 12.2%. In subsequent years (corresponding to graduate training), 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.

These returns to schooling are relatively low. BH explain this result by the presence of large differences in local returns between various grade levels. OLS and linear IV estimates (measuring an average log wage increment per year of schooling) tend to be biased toward the local returns at schooling attainments that are the most common in the sample data. The results illustrate the distortions introduced in a model built on the assumption that local returns to schooling are constant. These distortions co-exist with classical ability bias. It is also noted that structural estimates do not require the form of separability between

³⁰In a companion paper (Belzil and Hansen, 2003), this assumption is relaxed. The authors find that parents' background variables explain most of the cross-sectional variations in schooling attainments but that individual differences in wages are mostly explained by ability endowments orthogonal to family background.

schooling and individual specific endowments which is needed in IV estimation.

It is possible to draw a parallel between these estimates and those reported in Keane and Wolpin (1997). The high degree of convexity reported in BH 2002 may be a reflection of the higher returns to education in white collar occupations (requiring a high level of schooling) than in blue collar jobs (requiring a lower level of schooling).

Aside from the estimates reported, the main contribution of the authors is to show that SSDP models may not only be used for estimating deep structural parameters and simulating counterfactual policy experiments, but may also be viewed as a substitute for linear IV estimation methods which impose a form of separability between the effect of the instrument and the error term in the first stage regression.³¹

6.5 Belzil and Hansen (2006): The Correlated Random Coefficient Model

Belzil and Hansen (2006) investigate the properties of the popular correlated random coefficient wage regression model which has attracted much attention in the treatment effect literature. The starting point of the paper is the following log wage function received by individual i , at time t , which is given by

$$\log W_{it} = \varphi_{i1} \cdot S_{it} + \varphi_{21} \cdot Exp_{it} + \varphi_{22} \cdot Exp_{it}^2 + \eta_i \quad (26)$$

The authors use a model similar to the one found in their 2002 paper and show that the estimates of the dynamic programming model with a rich heterogeneity specification, may be used to obtain all treatment effects proposed in the microeconomic literature; the average treatment effects (ATE), the average treatment effects for the treated and the untreated (ATT/ATU), the marginal treatment effects (MTE) and, finally, the local average treatment effects (LATE). Unlike the standard Late parameters estimated using IV techniques, their estimates are obtained without relying on any form of separability between individual heterogeneity and schooling. To evaluate the LATE, BH simulate a mandatory high-school graduation (grade 12) policy and an increase in the utility of attending school during college. Using ex-ante and ex-post simulated schooling and wage histories, they identify those individuals affected by each specific intervention, and estimate the specific LATE parameter.

BH find support for the validity of the CRCWRM as returns to schooling are subject to a high degree of dispersion across individuals (types). Interestingly, low wage returns are compatible with the occurrence of very high returns to schooling in some subpopulations (the highest type specific return is 0.14 while the lowest is around 0.01). The importance of selectivity based on individual specific returns

³¹This point is developed further in Belzil and Hansen (2006) and is also central to the recent literature on instrumental variables.

to schooling is illustrated by the difference between the average returns for those who are predicted to stop before college (0.051) and those predicted to attend college (0.10). The estimates therefore conform to the intuition. The ATE (the average return in the population) is equal to 0.0573. It therefore lies between the ATU and the ATT.

The LATE estimates also conform to intuition. The estimates are found to be highly instrument dependent. The LATE for the minimum grade attainment experiment (0.051) is found to be lower than the LATE obtained for the college attendance subsidy (0.10). As we show, the variation across experiment is easily explained by the variation in the density of the sub-population affected by each experiment.

The structural dynamic programming model with multi-dimensional heterogeneity is therefore capable of explaining the coexistence of relatively low returns to education (on average) with very high returns for some identified sub-populations and, in particular, capable of identifying the characteristics of those individuals who would react to a counterfactual experiment.

6.6 Eckstein and Wolpin (1999): High-School Drop Out

Eckstein and Wolpin (1999) present a structural model of high-school attendance, work and academic performance. This article does not focus on the return to schooling. Indeed, post high school graduation is not modeled explicitly and there is no reliance 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.

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 ability 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) because the authors only consider young males who were less than 15 on 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 in Keane and Wolpin (1997), all heterogeneity is regarded as unobserved, and it is assumed that the population is represented by four types.

The results indicate that working while in school reduces academic performance, although the effect is small. 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 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). It also sheds light on the reasons behind the ineffectiveness of tuition subsidies for those of low ability in the population, reported in Keane and Wolpin (1997).

6.7 Magnac and Thesmar (2002,b): Cohort Effects in Schooling Attainments

Magnac and Thesmar (2002,b) estimate a dynamic schooling model which does not require the solution (or the approximation) of value functions. Their empirical analysis is based on a set of moment conditions derived from a companion paper in which they analyze the degree of under-identification of dynamic discrete choices (Magnac and Thesmar, 2002,a). They use the fact that there is a one-to-one mapping between value functions and choice probabilities; a fundamental result found in Hotz and Miller (1993). Their constructive identification proofs suggest a simple estimation method.³² In their empirical application, Magnac and Thesmar investigate three competing explanations for the increase in schooling attainment observed between 1980 and 1993 in France. These factors are an 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. An interesting finding is that, in France, the increase in attainment is most likely explained by a decrease in academic selectivity.

³²More precisely, 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 proposed by Hotz and Miller (1993).

6.8 Cohen-Goldner and Eckstein (2002 and 2004): Training and Occupation Choices

In two companion papers, Cohen-Goldner and Eckstein (2002, 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. Although the authors are not concerned with schooling per se (schooling obtained in the Soviet Union is assumed to be pre-determined), government provided training may still be viewed as a form of general skill.

As in Keane and Wolpin (1997), the agent has a choice between working in a blue-collar job or a white collar job. In both papers, the authors assume that the job offer rate and the labor market wage depend on the occupation as well as a participation in training. This constitutes a more general framework to evaluate publicly provided training programs than those which are typically found in the evaluation literature.

As the papers are similarly structured, 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 pre-determined. The offered wage in occupation j follows a Mincerian wage function, that is

$$\log W_{jt} = \varphi_{j1} \cdot S_i + \varphi_{2j} \cdot K_{1,t-1} + \varphi_{3j} \cdot K_{1,t-1}^2 + \varphi_{4j} \cdot K_{2,t-1} + \varphi_{4j} \cdot K_{2,t-1}^2 + \varphi_{6j} \cdot DT_t \quad (27)$$

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 , and 0 if not. The offer probability is also occupation specific and depends on the training indicator.

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, they find that the return to local training is high. More precisely, the wage return to training is almost 20% in white collar occupations but it is not statistically different from zero in blue collar occupations. The authors also 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 findings in the classical evaluation literature, the social and private gain to training is large. This holds despite the virtually null effect that training has on blue collar wages.

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 does 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.

The contribution of Cohen-Goldner and Eckstein goes beyond the structural literature per se. Indeed, φ_{6j} constitutes one of the most investigated parameters in the “treatment effect” literature. This is particularly true when the effect of training is allowed to be individual specific. 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. The key difference is that their model is set in a dynamic setting and considers multiple ways in which training may affect labor market outcomes; namely wages, employment (job search outcomes) and timing of re-employment. 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 advantaged workers is close to zero. However, it has been recognized that training may significantly affect the employment probability (Ham and Lalonde, 1996).

7 General Equilibrium Models of Schooling

7.1 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 equation (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.³³ The change in skill price is typically identified through movements in aggregate measures of some (or all) 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.³⁴

7.2 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.

The human capital production function chosen by HLT is more general than most of those used in the literature. There exists 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.

The wage rate is the product of the skill rental price and the amount of skill 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 depend on how many other people go to college, since the wage rate decreases with school attendance. In partial equilibrium, this is ignored.

HLT pay particular attention to the effect of a simulated change in tuition policies on college enrollment, a topic beyond the scope of this survey.³⁵ 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

³³Multiple skill models are developed in Willis (1986) and Heckman and Sedlacek (1985).

³⁴To 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).

³⁵For a survey of the literature concerned with tuition policies, see Wolpin (2000).

experience is only modeled through the decision to work or 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, wages are not represented by a classical wage regression function with separability. Schooling facilitates on-the-job training and the return to schooling changes with age.

As stated before, the returns to schooling reported in HLT are not directly comparable to those found in a Mincerian wage regression. The returns are age specific and take into account the causal effect of schooling on training opportunities. In addition, 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).³⁶

7.3 Lee (2005): Cohort size, Occupation Choice 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 size, schooling attainment and occupational choice, Lee investigates how cohort size affects wages and how a counterfactual tuition policy change would affect college enrollment. From the individual optimization side, his model is quite close to Keane and Wolpin (1997). Basically, individuals make schooling and occupational choices 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, 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 (1997), the wage regression is linear in schooling and quadratic in accumulated experience and returns to schooling are estimated as 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, using techniques which appear similar to those used in Keane and Wolpin (1997).

The main results are consistent with those reported above. First, he finds a positive correlation between white collar skill and taste for schooling, but a

³⁶In Belzil and Hansen (2002), the return to college education (4 years) averages 10.4% per year of college.

negative correlation between the blue collar skill and taste for schooling. As in Keane and Wolpin (1997) and Belzil and Hansen (2002), relatively low returns to schooling are found. The return to schooling in white collar occupations is 0.079 while it is equal to 0.048 for blue collar occupations. 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 experience). The returns are quite close to those obtained by Keane and Wolpin (1997), even though they are obtained from a different data set.

Lee uses his model to investigate how the “Baby Bust” generation (born between 1930 and 1940) compared with the “Baby Boom” generation (born between 1950 and 1960), and investigate the sensitivity of college enrollments to college tuition. He finds that those belonging to the “Baby Bust” generation faced higher skill prices, attended college more and worked more than those belonging to the “Baby Boom” generation. He also finds that a one percent increase in college tuition would decrease college enrollments by 1.05%. General equilibrium effects are found to be not so important.

7.4 Lee and Wolpin (2006): Intersectoral Mobility

Lee and Wolpin (2006) extend both Keane and Wolpin (1997) and Lee (2005) to estimate a two-sector (goods and service) equilibrium model of endogenous schooling, occupational and sectoral choices. As do Lee (2005) and Heckman, Lochner and Taber (1998), they assume a competitive equilibrium model. The model embeds both supply and demand features that may explain the secular increase in service sector employment in the US.

The model includes an aggregate Constant Elasticity of Substitution (CES) production function for the goods and service sector with three skill types (white-, pink-, and blue-collar occupations) and capital. The model is quite comprehensive. It also includes both time-varying neutral and non-neutral technological change and accounts for aggregate productivity shocks. Agents choose among eight discrete alternatives: schooling, home production and any of the six sector-occupations. In order to solve for the six different skill prices (which must equate the marginal revenue product), the authors develop a reduced-form approximation of the rational expectation equilibrium.

To estimate the model, a method-of-moment estimation technique, as well as data moments on employment (occupations and sectoral shares), wages and schooling from the Current Population Survey (CPS) from 1968 to 2001 and the NLSY79, are used. As the model has a large number of parameters, it is difficult to summarize. However, given the objective of this survey, it is important to recall that both sectoral and occupation mobility costs are found to be very high. For instance, the cost of moving from any occupation in the goods sector to the same occupation in the service sector is close to \$10,000 per year. Given the structure of the model, the return to schooling is sector and occupation specific. Schooling

increases white collar wages the most in both sectors, and given occupation, the return to schooling is higher in the service sector. Overall, the returns are low and comparable with those reported before.

7.5 A Brief Summary

After having reviewed a relatively large number of important contributions, it is now informative to present a brief synthesis of the results. Table 1 summarizes structural estimates of returns to schooling obtained in all models in which schooling is endogenously determined. As seen earlier, in a linear setting, the marginal effect of schooling on log wages lies between 4% and 7% per year of schooling. Returns are also low when non-linearities are allowed, although local returns during 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, and are therefore compatible with the existence of a positive ability bias.

As mentioned in the introduction, structural estimates are also much lower than those reported in the experimental literature. To see this, a set of estimates (along with their standard errors) reported in the experimental literature is summarized in Table 2.³⁷ 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 2 exceed OLS estimates and are close to (or exceed) 10% per year. Except for Lemieux and Card (obtained with Canadian data), these estimates have been obtained with US data.³⁸

³⁷I report standard errors for IV estimates because they are known to be relatively imprecise. Structural estimates, on the other hand, are typically very precise and are characterized by very small standard errors.

³⁸Card (1999 and 2001) provides an in-depth discussion of several of these papers and covers other contributions as well.

8 A Comparison of the IV and the Structural Approaches

This section is devoted to a comparison of the structural and the IV (experimental) literature. In order to perform this comparison, it is useful to review those aspects that are in common to both, but is also important to point out the main attributes of the structural literature that distinguish it from the IV literature.

8.1 The Experimental Literature

As mentioned earlier, a large number of papers based on IV techniques have been devoted to the estimation of returns to schooling. For a large part, the literature is based on institutional features of the education system, or on “Natural Experiments”. Card (1999 and 2001) presents an in-depth survey of the IV literature devoted to returns to schooling and discusses the main conceptual issues within a unifying theoretical structure (closely related to the Becker model) in which individuals compare the benefits of schooling with the costs of schooling born early in the life cycle.

At the outset, it must be understood that the Becker/Card model is fundamentally different from the dynamic models found in Keane and Wolpin (1997, 2001), or in many other contributions reviewed above. More precisely, the theoretical framework found in Card (2001) is fundamentally static in that it does not consider the effect that schooling may have on post-schooling human capital accumulation opportunities (post-schooling wage growth is exogenous). Ultimately, the relationship between wages and schooling is understood as a compensating wage differential, emerging from individual differences in abilities and/or discount rates.

In the dynamic (finite horizon) models of Keane and Wolpin (1997) and Belzil and Hansen (2002), the opportunity cost of schooling also incorporates lost labor market experience and forward looking agents may substitute experience for schooling in order to magnify life cycle income. As we will see in more details later, this means that exogenous events occurring before individual school completion may constitute useful statistical information, but may rarely be sufficient to uncover the key economic parameters that characterize human capital accumulation over the entire life cycle.

At the empirical level, Card (2001) reports that a large number of studies find that IV estimates exceed OLS estimates by a wide margin and notes that the tendency to obtain IV estimates, which exceed their corresponding OLS estimates, has been observed in many different data sets and for many different countries. Rosenzweig and Wolpin (2000) present a survey of the economic literature using natural experiments and focus their discussion on the implicit assumptions made in the IV literature. They discuss the return to schooling and ability bias as a

special case.

It is common to explain the OLS/IV discrepancy by the existence of potential heterogeneity in returns to schooling.³⁹ In the presence of heterogeneity in the returns, IV estimates are inconsistent for the population average. Indeed, there is a large econometric literature concerned with the interpretation of IV estimates when slopes are individual specific. In the context of a correlated random coefficient model, the IV estimator is sometimes referred to as a Local Average Treatment Effect (LATE).⁴⁰ Under certain conditions, the LATE parameter may be viewed as an average return to schooling for the sub-population affected by the experiment (the instrument), although the density of this sub-population is not identifiable. It is often postulated that high returns are explained by the fact that those individuals who are 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 the same individuals have higher returns to schooling than average. The difference between OLS and IV estimates is therefore explained by the non-representativity of this particular sub-population. Ultimately, the OLS-IV discrepancy is not incompatible with a positive ability bias.

8.2 Behavioral vs. Statistical Assumptions

Despite the existing split in the econometrics literature between those advocating the IV approach and those who prefer structural analyses, it is useful to compare the foundations underlying each approach and to recognize their commonalities.

Obviously, empirical work is impossible without some assumptions. For the sake of the presentation, it is important to distinguish between two types of assumptions: namely behavioral (primary) assumptions and statistical (secondary) assumptions. Behavioral assumptions are those which link a particular empirical model to economic theory. For example, some specific assumptions will dictate whether the underlying model is consistent with static or dynamic (forward looking) behavior, or whether an agent is behaving according to a rational expectation hypothesis, or whether an agent faces uncertainty or not. Secondary assumptions, on the other hand, belong to a later stage of econometric modeling and consist of auxiliary statistical assumptions. In the contemporaneous econometric jargon, these assumptions are often referred to as “functional form” assumptions.

In the IV literature, behavioral assumptions are explicitly stated by the designation of endogenous variables but are also implied by orthogonality conditions. The orthogonality conditions indicate which variables, not directly part of the model but part of the error term, are implicitly assumed to be exogenous. In the structural literature, the behavioral assumptions are explicitly stated because

³⁹Measurement error will be discussed briefly in Section 9. Recent surveys of the IV literature do not seem to retain the measurement error explanation as an important one (see Heckman, Lochner and Todd, 2005).

⁴⁰See Imbens and Angrist (1994).

the optimization problem, assumed to be solved by the agent, is also an input to estimation.

Statistical assumptions are, for the most part, dictated by convenience and rarely originate from economic theory. In the structural literature, such assumptions may include the parametric form of preferences and technology or the distribution of stochastic shocks.⁴¹ In the IV literature, the statistical assumptions are those that characterize the outcome equation as well as the change in the endogenous variable induced by the instrument.

As we will see below, the distinction between behavioral and statistical assumptions may be a useful way to try to comprehend the distinctions between both approaches.

8.3 What Distinguishes the Structural Literature from the IV Literature?

As neither estimation method can be nested in the other, it is important to understand the key dimensions of the structural approach that distinguish it from the more mainstream IV approach. These differences may be the key to understanding the reasons behind the diverging estimates of the return to schooling. When trying to comprehend these discrepancies (in Section 9), it will be useful to organize the discussion around these points.

To fix ideas, suppose one wishes to analyze the effect of a policy experiment, which consists in implementing a college attendance subsidy. To grasp the fundamental differences between the two approaches, let's consider the basic inputs used by IV advocates and those used by structuralists, in this particular context.

The inputs of the structural analysis are i) the individual (per-period) preferences, $U_{ik}(t, \omega_{ik}, \varepsilon_{it}^k)$, and ii) the law of motion, $\Omega_{i,t+1} = F_{it}(\Omega_{it}, d_{ikt}, \eta_i)$. To refresh the reader's memory, ω_{ik} and η_i are time invariant individual-specific taste and ability, ε_{it}^k is a state specific utility shock, $d_{kt} = 1$ when k is chosen and 0 if not, and Ω_t defines the set containing all state variables known by the agent at t . Most of the time it is assumed that $U_{ik}(t, \omega_{ik}, \varepsilon_{it}^k)$ is linearly separable in ε_{it}^k . Under some conditions that will ensure regularity of the problem, the sequential probability of exercising choice k at date t is representable by a function $\Psi_{itk}^s(\cdot)$, such that

$$\Pr(d_{ikt} = 1) = \Psi_{itk}^s(\Omega_{it}, \omega_{i1}, \dots, \omega_{iK}, \eta_i, \beta; \dots) \quad (28)$$

The superscript s stands for "structural" since an equivalent function will also

⁴¹Technically speaking, it is sometimes possible to use economic theory as an input to preference specification. For instance, a per-period utility of income has to be concave for a risk averse agent.

be introduced for the experimental analysis.⁴² Because the structural approach usually allows for a large set of endogenous variables, the state specific duration outcomes, $D_1(t), D_2(t), \dots, D_K(t)$, measuring accumulated periods spent in each state as of date t , and defined as

$$D_k(t) = \sum_{j=1}^t d_{kj} \quad (29)$$

depend also on a large number of endogenous variables since Ω_{it} depends on all previous choices (the $d_{ik1}, d_{ik2}, \dots, d_{ikt-1}$) and on the realizations of the random shocks $\varepsilon_{i1k}, \varepsilon_{i2k}, \dots, \varepsilon_{it-1k}$.

To link structural analysis with a policy experiment, one has to recognize that an “experiment” is representable by a change in one component of the determinants of the individual choices (a state variable or a parameter). This parameter (or variable) may belong either to $U(\cdot)$, to $F_t(\cdot)$, or it may be one element of Ω_t . In a sequential setting, the change may affect a full sequence of parameters. In the college attendance subsidy, it is natural to view the policy as creating an increase in the net utility of attending college, or, alternatively as a decrease in the cost of attending college between grade 13 and grade 16⁴³ Let’s denote the cost of schooling at by ω^1 and suppress the calendar time for simplicity. Suppose that the experiment boils down to changing ω^1 by an infinitesimal amount, and that $\Psi^s(\cdot)$ is differentiable in ω^1 , then $\frac{\delta \Pr(d_{ik}=1)}{\delta \omega^1}$ (like $\Psi^s(\cdot)$) is not separable in the heterogeneity components (ω_{ik}, η_i) . It is also important to note that the recursive structure of the model imply that a change in ω^1 entails a changes in choice made prior to college enrollment through the option value (because of the Rational expectation hypothesis) as well as changes in post-college choices through the dynamic recursive structure.

In the IV literature, there are also two fundamental inputs; the outcome equation and the first-stage regression model. However, the distinction between parameters, state variables or other primitive quantities is not possible. The starting point of the analysis is usually an ad hoc model of the desired level of schooling ($D_{ik}(t) = S_i^*$). The econometrician typically assumes that there exists some exogenous variable, say z_1 , which is one element of a potentially larger set of variables Z , which captures the changes in the environment of the individual. In most cases, it is assumed that the binary variable (z_1) is sufficient to characterize the change in the economic environment faced by the agent. To perform the analysis, the researcher assumes that it is possible to decompose the

⁴²The reader interested in the technical aspect of discrete dynamic programming may consult Rust (1994), or Whittle (1982).

⁴³In some very specific cases, for instance in a model with liquidity constraints, this policy experiment may also entail a change in the discount factor. Note also that I am abstracting from the very special experiment which would consist of imposing a dictatorial choice d_{kt} . Mandatory school attendance would fall in this category.

relevant labor market history (schooling attainment) into the sum of a regression component that typically depends only on the binary indicator (z_1) and on an orthogonal residual component. In practice, the underlying schooling decision model may be represented by a single index model such as

$$D_{ik}(t) = S_i^* = \Psi^E(Z_i, \varepsilon_i) \quad (30)$$

where $\Psi^E(\cdot)$ is weakly separable in Z and ε (the error term), and where a form of orthogonality between Z and ε must be maintained. The first stage of the IV approach may therefore be seen as an approximation of $\Psi^s(\cdot)$. In practice, it is also common to simplify the analysis by assuming that the relevant choice is a single discrete decision (say, the decision to enter college or stop after high school), in which case the equation for S^* serves as a latent variable model that generates the observed decision. Obviously, in such a case, the schooling decision is reduced to a static (one shot) problem and the lack of allowance for multiple schooling levels removes the sequential nature of the schooling decision.⁴⁴

To obtain identification, z_1 (and possibly other elements of Z_i) must be excluded from the outcome equation. In the case where the outcome (wage) equation is not linear in the error term, a form of “monotonicity” must be assumed.⁴⁵ Strictly speaking, the monotonicity condition implies that a change in the relevant element of Z_i must affect all individuals in the same direction. In the example of the payment of a college attendance subsidy, z_1 may be a binary indicator which records exposure to the new policy. Effectively, when Z is a scalar, the analyst must assume that it is sufficient to summarize all changes in the environment, even if the changes affect a vector of parameters.⁴⁶

At this stage, two important points should be emphasized. First, the regression representation of the outcome equation used in the IV approach (30) is only valid if there are no endogenous state variables that are allowed to vary until school completion. In a general multi-state dynamic model, the conditional mean schooling outcome would depend on other endogenous variables (perhaps all of the $D_k(t)$'s) and an exogenous change in a state variable (or a parameter) would automatically create a change in some (or all) of these endogenous variables. The IV approach would then break down.⁴⁷ In the context of the college tuition example, this could arise if the decision to stay in school is affected by accumulated periods of employment, or by participation in part-time employment while in school. Obviously, in a more restrictive theoretical setting, this condition could be met. For instance, this may be the case in an optimal stopping model

⁴⁴For internal consistency, the outcome equation (the wage function) has also to be modified so to be consistent with a discrete outcome.

⁴⁵The term “uniformity” is also occasionally used.

⁴⁶The issue becomes even more complicated if one considers potential general equilibrium effects.

⁴⁷Some more specific examples are provided in Section 8.2.4.

of schooling where endogenous work experience (or non market time) is never accumulated before schooling is completed.⁴⁸

Second, because the IV approach does not allow for a distinction between unobserved heterogeneity (from the perspective of the econometrician) and ex-ante risk (from the perspective of the agent), there is no correspondence between the multi-dimensional random terms of the DP model and the scalar error term of the IV approach.

To summarize, the distinction between the IV approach and the structural approach may be coined in terms of a trade off between behavioral and statistical assumptions. Generally speaking, the structuralists design dynamic economic models which allow for a relatively large set of endogenous variables, and relatively general law of motions. To the extent that structural models allow for more endogenous variables than do natural experiment analyses, the structural approach is characterized by a smaller number of behavioral assumptions. However, this comes at the expense of a typically large set of statistical (secondary) assumptions, which are needed in order to estimate any model in which subjective probability distributions play a central part.⁴⁹ On the other hand, most experimentalists are concerned with only the first moment of both the outcome equation and the endogenous (first stage) variable, and therefore make a smaller number of parametric assumptions. However, because “natural experiments” are scarce, most papers set in an IV framework consider only one endogenous variable, and must therefore make a large number of behavioral assumptions. As seen earlier, these assumptions take the form of orthogonality conditions and restrict the class of models that are amenable to IV estimation to either static models or to a very narrow class of dynamic models.

It is important to understand that, unless one has very specific knowledge of the true data generating process, it is neither possible, nor sensible, to determine which approach to estimation is more flexible. The class of models compatible with each specific estimation approach is typically limited by behavioral and statistical assumptions. For one thing, the economist does not really know the dimensionality of the vector of endogenous variables (especially over the life cycle). Further, the relative impact of the statistical and behavioral assumptions made in each specific paper is usually the object of speculation, but not the object of formal work. Put differently, it is virtually impossible to distinguish the structural literature from the IV literature on the basis of their propensity to make assumptions. It is rather by understanding the nature of their specific assumptions that economists will move toward comprehension of the fundamental

⁴⁸The exact conditions under which the regression function is well behaved would have to be worked out on a case by case basis. It is not my intention to discuss general rules herein.

⁴⁹I am abstracting from recent results found in Heckman and Navarro (2006) which show how subjective probability distributions may be non-parametrically identified in a finite horizon optimal stopping model.

differences between the two approaches.⁵⁰

8.3.1 Selectivity

By imposing an econometric structure in which young individuals make schooling decisions based on a set of parameters which includes their individual specific market abilities (ω_{ik}, η_i) , 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. 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” and the structural estimates allows the econometrician to quantify directly the OLS bias.

In the IV literature, the selectivity process is viewed as being driven by nuisance parameters in which the analyst has no explicit interest.⁵¹ Indeed, estimation does not specify anything about the information set of the agent when schooling decisions are made. This, however, does not imply an absence of assumption. As mentioned earlier, the separability/monotonicity assumption implies that a change in the instrument must affect everyone in the same direction. Heterogeneity in the effect of the instrument on schooling is therefore largely conditioned by this property. In the extreme case where $\Psi^E(\cdot)$ is linear in the instrument and every individual shares the same slope, the marginal effect of the instrument is independent of all heterogeneity components and the implied marginal effect of the policy experiment must be identical for everyone. In the tuition policy experiment, this is automatically violated. Indeed, in virtually all experiments (mandatory schooling requirement, high school graduation subsidy,..etc), homogeneity of the marginal effect would be violated. For instance, in a linear probability first stage regression model, a common parameter for the effect of z_1 is equivalent to the assumption that $\frac{\delta \Pr(d_{ik}=1)}{\delta \omega^1}$ is independent from ω_{ik} and η_i . In such a case, and unless the econometrician has access to variables that measure all relevant skills, selectivity is ruled out.

8.3.2 The Specification of the Mincer Equation

In many papers surveyed above, estimates of the returns to schooling are obtained from a more flexible specification of the Mincerian wage regression than the one

⁵⁰Obviously, from a purely logistic (computational) perspective, the IV approach is much easier to implement. It requires neither high-level econometric skills, nor advanced knowledge of dynamic economic methods.

⁵¹This is why, at a somewhat anecdotal level, it is common to hear that the IV approach does not require any “functional form” assumption.

typically found in the experimental literature. In some cases, this flexibility refers to the possibility that local returns to schooling vary with grade level. Structural analysis allows the uncovering of all local returns [the functional form of $\varphi_1(\cdot)$ in equation (3)]. In other cases, it translates into an allowance for individual (or occupation) specific returns. If so, the researcher has also interest in characterizing higher moments of returns to schooling.

In the experimental literature based on IV techniques, the effect of schooling is linear by choice, or is implicitly linearized, and it is ultimately summarized in a single parameter estimate. However, the linearity assumption is virtually always rejected (Belzil and Hansen, 2002, 2003).⁵²

8.3.3 Identification of the population Affected by Exogenous Changes

In the IV literature, the schooling decision is often specified as a static problem or, at most, as a two period problem and it is customary to focus on college attendance as the treatment of interest. If so, the individuals who would have decided to attend college without exposure to the instrument (as well as those who are not attending even after the implementation of the change) are identified as those not affected by policy changes. In a fully dynamic setting, the issue is more complicated. If one considers the 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 would have entered college even without being exposed to the experiment by increasing their continuation (graduating) probabilities. Equally, the subsidy may attract potential high school drop out to graduate from high school. This implies that those affected by an exogenous change in schooling costs are not only those who are at the margin of taking a decision (say to attend college) but may also be those who would have gone to college even in absence of the subsidy, but were affected through continuation (or graduation) probabilities. This is also a major difference between the structural approach and the IV approach.

8.3.4 Post Schooling Optimization and Orthogonality Conditions

In Section 8.2, I already mentioned the limitations of an IV model based on the decomposition the endogenous variable into the sum of a regression component that depends only on the experiment indicator and on an orthogonal residual. I now focus on the outcome (log wage) equation. Although similar, the line of argumentation will now be more precise. Obviously, in a static single-period optimization 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 such a case, orthogonality requires a form of non-anticipation. When the

⁵²This is more likely a consequence of the scarcity of true experiments.

wage outcomes are measured after the effect of the instrument has set in, the exclusion restrictions typically imply that a variable observed at one point in time affects the choice that an individual made at that time, but does not affect future outcomes. These conditions act as identifying conditions. However, in a dynamic setting, the orthogonality between the instrument and the endogenous variable(s), realized after the effect of the instrument is generally violated because the event (the instrument) is part of the state variables observed by the individual.⁵³ If so, the exclusion restrictions are incompatible with a dynamic model specification.

This may be well illustrated by examining the treatment of labor market experience as a substitute for time spent in school, in order to achieve higher life cycle wages. While most authors using IV techniques base 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 because 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 because inferences are seldom drawn from entry wages. A survey of the IV literature reveals that practically no paper presents a joint estimation of 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.

To see this, consider the following simple framework (a slightly modified version of (15)), where the wage equation is

$$w_{it} = \beta_0 + \beta_1 \cdot S_{it} + \delta \cdot PSHC_{it}(\{d_{k \neq s, j}\}_{j=1}^t) + \eta_i + \varepsilon_{it} \quad (31)$$

where $PSHC_{it}$ denotes an aggregate measure of all cumulated post schooling human capital investment activities and where $\{d_{k \neq s, j}\}_{j=1}^t$ is a vector representing the sequence of discrete choices other than schooling (work, training, search or 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 (31) using IV methods⁵⁴, namely

$$w_{it} = \beta_0 + \beta_1 \cdot S_{it} + \eta_{it}^* \quad (32)$$

where

$$\eta_{it}^* = \delta \cdot PSHC_{it} + \eta_{it}. \quad (33)$$

⁵³Presumably, the period between the realization of the instrument and the measurement of the wage outcome (namely s) may also be relevant in assessing the lack of orthogonality.

⁵⁴The argument will work as well if you add an additional control for potential experience or age (as is sometimes done) .

Assume that you are given access to an instrument Z_i . Typically, Z_i refers to the occurrence of an event that took place before t , say at t' , or, equivalently, 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

$$Corr(Z_i, \eta_{it}^*) = 0 \quad (34)$$

which is a much stronger condition than $Corr(Z_i, \eta_i) = 0$. However, in general, (34) cannot be true since Z_i is typically correlated with $PSHC_{it}$ by construction. This may easily be demonstrated using some technicalities but it can be also illustrated with more intuitive arguments.

In technical terms and in the context of the model presented in Section 4, 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 by $\omega_{t'+s}^w$. The non-orthogonality is explained by the fact that all elements of $\Omega(t' + s)$, including $\omega_{t'+s}^w$, are affected by Z_i since the choices made at t' (the $d_{kt'}$ s) are affected by Z_i (because Z_i is an instrument). Therefore, this also means that actions taken between t' and $t' + s$ are also in principle affected by Z_i . From the recursive structure illustrated in (14), we get that

$$\begin{aligned} & \omega_{t'+s}^w(d_{k,t'+s-1}, \Omega_{t'+s-1}) \\ & d_{k,t'+s-1}(\Omega_{t'+s-1}) \\ & \Omega_{t'+s-1}(d_{k,t'+s-2}, \Omega_{t'+s-2}) \\ & \dots \\ & d_{k,t'+1}(\Omega_{t'+1}) \\ & \Omega_{t'+1}(d_{k,t'}, \Omega_{t'} \setminus Z_i, Z_i) \end{aligned} \quad (35)$$

This illustrates again that any element $\omega_{t'+s}^w$ may only be independent of Z_i after conditioning on $\Omega_{t'+s-1}$.

In most empirical applications, the researcher assumes that the effect of Z_i is limited to $d_{kt'}$ only. However, realistically, η_{it}^* is affected by the entire history of $d'_{kt'}$ s from t' to $t' + s$. This illustrates the differences between an IV strategy set in an intertemporal context, where individuals optimize sequentially in every periods, and one set in a standard static context.

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

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 and discuss the season of birth example. They also point out that 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 its estimation.

Obviously, this is not to say that structural models always take into account specific post-schooling investment activities. Indeed, many models reviewed above do not. However, to the extent that they all take into account that schooling entails an opportunity cost in terms of work experience, they automatically capture some post-schooling endogeneity.

IV techniques, which have been introduced in econometrics more than sixty years ago, were developed in a context where the immense majority of economists could only envisage economic problems within a static (single period) framework. Nowadays, it is difficult to say if the popularity of IV techniques is due to a lack of interest of empirical labor economists in dynamic models, or is instead due to the technical facility with which IV techniques may be implemented. It is nevertheless interesting to note that the conceptual difficulty of introducing instruments in a dynamic framework have been recognized for a long time in the macro-econometric literature.⁵⁵

To a large extent, it is possible to view the structural literature on the return 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 when the experimental approach may not be conclusive. It is the key differences between the structural and the experimental approaches that allow the researchers to shed new perspectives on the estimation of returns to schooling and the sign of the ability bias.

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

This is, of course, the natural question to ask at this stage. Explaining the divergence between structural and IV estimates is largely complicated by the fact that neither estimation method is directly nested within the other and because IV estimates may not always constitute central location parameters. Finally, IV methods typically provide a single point estimate for the return to schooling

⁵⁵It should be pointed out that this particular issue arising in a dynamic setting has already been noted in the dynamic rational expectation literature in the early 80's. For a classic example, see Hansen and Sargent (1982).

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 fundamental differences between the structural and the IV approaches (identified in Section 8) may now serve as a basis for understanding the differences in corresponding estimates. Ultimately, I have grouped these potential reasons into seven distinct sets.

9.1 The Possible Inconsistency of IV Estimates

As indicated in the previous section, the reliability of the experimental approach may be questioned by virtue of the assumptions behind IV strategies. This is the topic of much on-going research in theoretical econometrics, but it is certainly not the focus of this paper. However, in order for this paper to be self-contained, I present two main arguments; the consequences of using weak instruments and the functional form restriction of the first stage regression.

9.1.1 Weak Instruments

First, IV techniques may be applied when 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 inconsistent. The large inconsistency is explained by the magnifying effect that the weak correlation between the instrument and the endogenous variable may have on the possible correlation 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.⁵⁶

9.1.2 The Specification of the First Stage Model

As discussed in Heckman and Vytlacil (2005), in the presence of heterogeneity in slopes of the outcome equation, the specification of $\Psi^E(\cdot)$ is not innocuous. The interpretation of the IV estimates depends crucially on a-priori assumptions regarding $\Psi^E(\cdot)$. Intuitively, this may be seen by taking the probability limit of the IV estimator obtained in a correlated random coefficient model. That

⁵⁶Carneiro and Heckman (2003) present an in-depth analysis of various instruments used in the literature.

probability limit is function of the instrument (z), of the conditioning variables (Z) and of the model that generates the discrete choice, namely $\Psi^E(\cdot)$.⁵⁷

Basically, the IV approach may be deficient for three main reasons. First, as illustrated in a multi-state dynamic (recursive) structure such as found in (33)-(35), it is generally impossible to decompose schooling decisions into a regression component that contains only a binary variable and an orthogonal component.

Second, when the true model generating the endogenous variable is marked with multi-dimensional heterogeneity and contains a large number of state variables, the vector Z_i must be sufficiently large to capture all the relevant variables that may proxy unobserved skills and market ability, as well as all the relevant state variables affected by the policy experiment. If not, the model is mis-specified. This also implies that the existence of the instrument is not sufficient to decompose the first stage equation into a regression component that depends solely on exogenous regressors and on an orthogonal error term.

Finally, as shown in Heckman and Vytlacil (2005), when the separability (or monotonicity) condition is violated, IV/LATE methods provide even inconsistent estimates of the effect of the change in policy for those affected. In the cross-section econometric literature, the violation of the monotonicity property is usually illustrated by the existence of “two-way flows” caused by the instrument.⁵⁸ . For instance, the monotonicity property may be violated when the econometrician is not controlling for a sufficiently large number of elements in the Z vector, or when policy interventions entail general equilibrium effects. IV therefore relies heavily on the first stage model specification. This is contrary to conventional wisdom about the robustness of classical IV estimation to model mis-specification. The separability (or monotonicity) restrictions are also at odds with a framework where the outcome equation is parametrized so as to allow enough heterogeneity (a feature that is typically the motivation behind most empirical work based on IV). The separability condition may be viewed as a statistical (or functional form) assumption. It is in no way related to economic theory.⁵⁹

Obviously, the fact that structural models depend on a-priori assumptions is widely known. As noted earlier, once utility functions and distributions are explicitly stated, the link between outcomes and state variables $\Psi_i^s(\cdot)$ is automatically specified. The key difference is that structural models allow to quantify the importance of various sources of heterogeneity. Obviously, as for any non-linear extremum estimator, structural estimates are generally inconsistent if the model

⁵⁷This implies that changing the functional form of $\Psi^E(\cdot)$ (for a given Z_i), may change the interpretation of the IV estimate.

⁵⁸Belzil and Hansen (2006) show that, in specific cases where an intervention may affect the discount rate, this same intervention may also affect the schooling decisions of some individuals in opposite directions.

⁵⁹As far as I know, the representability result is due to Vytlacil (2002) and the monotonicity condition (applied to IV) originates with Imbens and Angrist (1994). Several implications of the monotonicity assumptions are discussed in Heckman and Vytlacil (2005).

is mis-specified. In order to minimize the impact of distributional assumptions, the structuralist has no other option than to use more flexible distributions.⁶⁰

However, the fact that IV estimates are based on a a-priori assumptions is not always recognized in the empirical literature. It is nevertheless important. IV methods are often preferred to structural analyses on the ground that they are said “to provide statistical analyses that are less prone to mis-specification”.⁶¹ As pointed out in the recent IV literature, this is only true in very specific cases, where both the first stage and the outcome equations are linear (additive) in the error term.⁶²

9.2 Systematic Bias and Measurement Error in Structural Methods

It is also natural to search the intrinsic nature of the structural approach for a systematic bias which could, for instance, lead to low returns to schooling. Such a bias would arise, for instance, if one specifies schooling choices in a context where the correlation between school and market ability is forced to be positive. Assuming homogeneity in taste for schooling and/or in market ability would most likely also have severe implications for the estimated return to schooling. A survey of the structural papers in which an estimate of the ability bias is either directly or indirectly available, indicates that all of them allow for the more able to have a higher opportunity cost of schooling (conditional on taste for schooling), and also allow for a free (flexible) correlation between the relevant heterogeneity components. 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 a sufficient explanation. The measurement error argument often advanced to explain the OLS-IV difference 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. Structural models are intrinsically non-linear and the estimates are obtained from the solution of a large number of moment conditions. Moreover, structural estimates are not only lower than IV estimates, they are also much below OLS estimates. Indeed, the distance between structural and the corresponding OLS estimates is typically

⁶⁰See Heckman and Navarro (2006).

⁶¹This is discussed in details in Keane (2006).

⁶²Heckman, Urzua and Vytlačil (2005) present a good synthesis of the literature on IV and explain several intrinsic weaknesses of the IV estimation strategies. Belzil and Hansen (2006) present a characterization of several counterfactual experiments and show how the reactions are correlated with individual specific returns to schooling. They also show that, in specific cases where an intervention may affect the discount rate, this same intervention may also affect the schooling decisions of some individuals in opposite directions.

as large as the OLS/IV difference. Without real advances in the econometric analysis of the impact of measurement error (non-standard) within non-linear models, it is difficult to say more.⁶³

9.3 The Rational Expectation Hypothesis

As structural estimates are obtained from panel data on choices and labor market wages, they allow a distinction between individual persistent factors and random shocks. However, all papers surveyed above assume that these factors are known with certainty. This rather extreme form of Rational Expectation may not be an innocuous assumption. It means, for instance, that high-wage workers (those endowed with high market ability) automatically perceive a high opportunity cost of attending school and this implies that high-wage workers attending school must also be endowed with high academic ability. The degree of correlation between abilities and schooling (ultimately driving the ability bias) may therefore be affected by the assumption that abilities are known with certainty.

On the other hand, IV estimates, which are typically obtained from cross-section data, ignore the distinction between permanent abilities and purely random shocks, and do not require to assume a formal information set.⁶⁴ This may be an advantage of the IV approach, although the implication of the Rational Expectation hypothesis remains unknown.

9.4 Liquidity Constraints

In his survey of the IV literature, Card (2001) postulates that one reason for the existence of high IV estimates is the existence of liquidity constraints for a subset of the population endowed with high returns to schooling. However, as appealing as this interpretation may be, the incidence of liquidity constraints is not identifiable unless one models borrowing constraints explicitly. Until now, the structural literature has not found much evidence in favor of the existence of liquidity constraints. This is the case in Keane and Wolpin (2001) as well as in Cameron and Taber (2004). There may be several reasons for that, but one explanation may be that many individuals have the option to attend college by working while in school, or simply, choose colleges or universities that impose lower tuition costs. Undoubtedly, this is a topic of research that will be popular in a near future.

⁶³The reader should remember that human capital literature is fundamentally plagued by the curse of measurement error in that, human capital itself, cannot be observed. This is true regardless of whether or not schooling is measured accurately.

⁶⁴This form of Rational Expectations may also be avoided in a semi-structural framework, where the schooling decision process is modeled as a reduced-form approximation to the sequential choices made by the individuals, and in which heterogeneity is allowed to be correlated with market ability (wage fixed-effects). See Belzil (2006).

9.5 Post-Schooling Dynamics and the Endogeneity of Post-Schooling Human Capital Accumulation

Another explanation may simply be that the Mincerian wage regression incorporates too 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 of different estimates. 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 a difficult task to predict the sign of the correlation between omitted post schooling wage growth and both schooling and specific instruments.

To see the issue, it is convenient to consider the dynamic (recursive) structure laid down in equations (31) to (35). Obviously, in the presence of post-schooling dynamics, the sign of the OLS bias depends on the covariance between post-schooling human capital investments ($PSHC_{it}$) and schooling, as much as on the covariance between schooling and unobserved ability. A related argument applies to IV estimates. If economic policies that raise schooling also affect the costs and/or the benefits of post-schooling human capital investments, the inconsistency of IV becomes clear. For instance, when the covariance between $PSHC_{it}$ and the instrument (Z) is positive, it is obvious that, for a given level of unobserved market ability, IV over-estimates the true effect of schooling. As of now, the importance of this type of bias is virtually unknown. Only further work will clarify this issue.

9.6 The Functional Form of the Mincer Equation

When the effect of schooling on wages depends on the level of schooling itself, the discrepancy is much easier to explain. When IV techniques are chosen, the log wage regression is usually assumed to be linear in schooling. 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.⁶⁵ 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 which are most common in the sample data. As shown in Belzil and Hansen (2002), this type of mis-specification is much more important than the ability bias itself. In this case, the difference between the average return to

⁶⁵This issue is sometimes referred to as the Discount Rate bias.

schooling obtained from structural estimates and IV estimates is easily explained.

9.7 The Consideration of Psychic Costs

Finally, the simultaneous consideration of individual specific psychic costs and wage returns, present in structural models, may also be a key reason for diverging estimates. In the structural approach, observed wage differentials across schooling levels are matched to actual choices by adjusting psychic costs estimates (for a given rate of impatience). For example, high returns are only compatible with decisions to leave school in the presence of non-trivial psychic costs. However, and as seen earlier, structural estimates, which are low compare to IV estimates, already imply a relatively important high-school/college premium (a relatively high degree of convexity) or, perhaps equivalently, a high blue collar-white collar premium. Assuming that psychic costs are prevalent in late high-school or in college, matching optimal schooling decisions with actual choices, in the presence of returns to schooling as high as 15% per year, would most likely require a much more important role for psychic costs than what is currently found. In return, this higher level of psychic costs would probably have to translate into a much larger high-school/college premium than is seen in the data.⁶⁶

9.8 Summary

In summary, it is probably fair to say the discrepancy between structural and IV estimates is primarily a reflection of the differences in the objects estimated by the researchers. 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 when agents continually optimize in a dynamic environment.

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. 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 interest in the notion of “ability bias” and has brought new credibility to the classical hypothesis, that the ob-

⁶⁶A similar issue arises when considering risk aversion. Given schooling costs (tuition plus psychic costs), a rate of return of 15% would require either a much higher degree of risk aversion than is found in the data, or equivalently, a huge level of risk associated with schooling. See Belzil and Leonardi (2006) for a discussion.

served correlation between wages and schooling is an over-estimate of the true causal effect of schooling on wages.

10 Conclusion and Avenues for Future Research

In the applied labor economics profession, it is common to hear the claim that structural models are only useful for estimating “deep structural parameters”. This common belief is at best only partially true, at worst completely misguided. As indicated by the content of the papers reviewed in this survey, the structural approach is also a powerful instrument for estimating outcome equations when the econometrician feels that the maintained hypotheses required for IV estimation are undesirable.⁶⁷

Obviously, there are several areas in which the structural literature is likely to make progress in a near future. First, and as stated earlier, structural models are typically formulated within a Rational Expectation framework, and in particular, 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. Whether this is a valid assumption or not is debatable. For instance, if heterogeneity identified from post-schooling data is partly explained by firm heterogeneity, this assumption becomes quite unrealistic. It would be interesting to develop an estimation framework which allows for gradual learning about academic and, in particular, labor market skills.⁶⁸

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 as a simple discrete choice. There are no compelling reasons to do so. Little is known about the relationship between schooling and the intensity of post-schooling human capital accumulation (including on-the-job training decisions as well as work intensity). This may also help understanding the accuracy of IV estimates that entail orthogonality conditions between an instrument and error terms realized at a later stage.

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 attainment and the effect of wage dispersion and schooling decisions has been completely ignored. Indeed, it is well known that the distinction between offered and accepted wages is relevant and may affect both the estimate of the return to schooling and experience.

⁶⁷Interestingly, very few structural papers appear to focus on the estimation of parameters such as the degree of risk aversion or the rate of time preference.

⁶⁸This issue is discussed in Cunha, Heckman and Navarro (2005).

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 when promotions and human capital accumulation are alternative methods to enhance life cycle income.

Off course, there are no reasons why these topics should be investigated solely by structuralists. Each of these questions is interesting in its own right and, indeed, structural estimation is often viewed as controversial in empirical labor economics.⁶⁹ 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 cycle. This means that exogenous events occurring 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.

In conclusion, the measurement of the return to schooling remains an open question. The split between those advocating structural analyses and those advocating the use of IV techniques is a sheer indication of the lack of agreement on the best way to proceed. However, and in light of the recent literature, which has documented a clear increase in the OLS estimate of the return to schooling in conjunction with recent technological changes, it is more than ever important to obtain a reliable measure of this parameter.

References

- [1] Bellman, Richard (1957) “*Dynamic Programming*” Princeton, New-Jersey, Princeton University Press.
- [2] Becker, Gary (1964), *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*, New-York: National Bureau of Economic Research
- [3] Belzil, Christian (2006) “Testing the Specification of the Mincer Wage Equation” IZA Working Paper (in Press)

⁶⁹Interestingly, structural estimation is “mainstream” in other areas of economics such as industrial organization, finance and macro economic theory. It is also growing in the field of Marketing.

- [4] Belzil, Christian and Hansen, Jörgen (2006) “A Structural Analysis of the Correlated Random Coefficient Wage Regression, forthcoming in *The Journal of Econometrics*.
- [5] Belzil, Christian and Hansen, Jörgen (2002) “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 (2004) “Risk Aversion, Earnings Dispersion and Education” *Research in Labor Economics*, volume 23.
- [8] Belzil, Christian and Marco Leonardi (2006) “Can Risk Aversion Explain Schooling Attainments?” IZA Working Paper.
- [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] Cameron, Stephen and Christopher Taber (2004) “Estimation of Educational Borrowing Constraints Using Returns to Schooling” *Journal of Political Economy*, 112 (1), 132-182
- [12] Card, David (2001) “The Causal Effect of Education on Earnings” Handbook of Labor Economics, edited by David Card and Orley Ashenfelter, North-Holland Publishers.
- [13] 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.
- [14] Carneiro, P., K. Hansen and J. Heckman (2003) “Estimating Distributions of Counterfactuals with an Application to the returns to Schooling and measurement of the Effects of uncertainty on Schooling Choice” *International Economic Review*, 44, 361-422.
- [15] Cunha, Flavio, James J. Heckman and Salvador Navarro (2005) “Separating Uncertainty from Heterogeneity in Life Cycle Earnings” Hicks Lecture, Oxford University, April 2004; *Oxford Economics Papers*, 57, 191-261 .

- [16] 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.
- [17] Cohen-Goldner, Sarit and Zvi Eckstein (2002) “Labor mobility of Immigrants: training, Experience, Language and opportunities” IZA Working Paper no 519.
- [18] Eckstein, Zvi and Gerard van den Berg (2003) “Empirical Labor Search: A Survey, IZA Working Paper # 929.
- [19] Eckstein, Zvi and Kenneth Wolpin (1999) “Youth Employment and Academic Performance in High School”, *Econometrica* 67 (6,)
- [20] Eckstein, Zvi and Kenneth Wolpin (1989a) “The Specification and Estimation of Dynamic Stochastic Discrete Choice Models”. *Journal of Human Resources*, 24, 562{598.
- [21] Eckstein, Zvi and Kenneth Wolpin (1989b) “Dynamic Labour Force Participation of Married Women and Endogenous Wage Growth”. *Review of Economic Studies*, 56, 375{390.
- [22] 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.
- [23] 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.
- [24] Hansen, L.P. and T.J. Sargent (1981) “Linear Rational Expectations Models for Dynamically Interrelated Variables” in Rational expectation and Economic Practice, edited by R.E. Lucas and T.J. Sargent, Minneapolis, University of Minnesota Press
- [25] 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.
- [26] Heckman, James (1979) “Sample Selection as a Specification Error” *Econometrica* 47 (1), 153-162.
- [27] Heckman, James, Lochner, Lance and Petra Todd (2005) “Earnings Functions, Rates of Return, and Treatment Effects: The Mincer Equation and Beyond” NBER Working Paper 11544

- [28] Heckman, James and Salvador Navarro (2006) “Dynamic Discrete Choice and Dynamic Treatment Effects” forthcoming in *Journal of Econometrics*.
- [29] 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.
- [30] Heckman, James and E. Vytlacil (2005) “Structural Equations, Treatment Effects and Econometric Policy Evaluations” *Econometrica*.
- [31] 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.
- [32] Imbens, Guido and J. Angrist (1994) “Identification and Estimation of Local Average Treatment Effects”, *Econometrica*, 62, 4,467-76.
- [33] Keane, Michael (2006) “Structural vs. Atheoretic Approaches to Econometrics, forthcoming in *Journal of Econometrics*
- [34] Keane, Michael P. and Wolpin, Kenneth (1997) ”The Career Decisions of Young Men” *Journal of Political Economy*, 105 (3), 473-522.
- [35] Lazear, Edward (1997) “Personnel Economics: Past Lessons and Future Directions” *Journal of Labor Economics*” vol 17, no 2, pp 199-236
- [36] 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
- [37] Lemieux, Thomas and David Card (2000) “Education, earnings and the Canadian GI Bill”, *Canadian Journal of Economics*
- [38] Magnac, T. and D. Thesmar (2002,a): “Identifying Dynamic Discrete Decision Processes,” *Econometrica*, vol. 70, 2002, p. 801-816.
- [39] 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.
- [40] Massé, Pierre (1946) “Les Réserves et la Régulation de l’avenir dans la vie économique”, Paris, Hermann, 1946; vol 1, Avenir déterminé, 148 pp; Vol II, Avenir aléatoire, 229 p.
- [41] Mincer, Jacob (1974) *Schooling, Experience and Earnings*, New-York: Columbia University Press,

- [42] Rosen, Sherwin (1977) "Human Capital: A Survey of Empirical research" Research in Labor Economics, Edited by R. Ehrenberg, JAI Press, vol 1.
- [43] Rosenzweig Mark and K.Wolpin (2000) "Natural Natural Experiments in Economics" *Journal of Economic Literature*, December, 827-74.
- [44] Roy, Andrew "Some thoughts on the Distribution of Earnings" *Oxford Economic Papers*, vol 3 (June), 135-146.
- [45] 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.
- [46] Sauer, Robert (2004) "Education Financing and Lifetime Earnings" *Review of Economic Studies*.
- [47] Staiger, Douglas and James H. Stock (1997), "Instrumental Variables Regression with Weak Instruments" *Econometrica*, 65: 557-586.
- [48] Stokey, N., Lucas, R.E. (with Ed Prescott) (1989). Recursive Methods in Economic Dynamics. Harvard University Press. Massachusetts.
- [49] Vytlačil, Edward (2002), "Independence, Monotonicity and latent Index Models: An Equivalence Result", *Econometrica* 70, 331-341.
- [50] Whittle, Peter (1982) Optimization over Time: Dynamic Programming and Stochastic Control, New-York, John Wiley and Sons
- [51] Willis, R. and S. Rosen (1979) "Education and Self-Selection", *Journal of Political Economy*, 87, S-7-S36.
- [52] 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
- [53] Wolpin, Ken (1984) "An Estimable Dynamic Stochastic Model of Fertility and Child Mortality", *Journal of Political Economy* 92:852-74
- [54] 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.
- [55] Wolpin, Ken (2003) "Wage Equations and Education Policy" in Dewatripont and L.P. Hansen, *Advances in Economics and Econometrics*, Cambridge University Press.

Table 1
Returns to schooling in Structural Models
with Endogenous Schooling

	Sample	return to schooling
Keane and Wolpin (1997)	blue collars (white males)	0.024
	white collars(white males)	0.070
	military (white males)	0.058
Keane and Wolpin (2000)	blue collars (white males)	0.071
	white collars(white males)	0.020
Keane and Wolpin (2001)	white males	0.075
Belzil and Hansen (2002)	in high school (males)	0.007
	in college(males)	0.104
	average (males)*	0.046
Lee (2005)	blue collar (males/Females)	0.048
	white collar(males/Females)	0.079
Belzil and Hansen (2006)	(type 1, males)	0.0189
	(type 2, males)	0.1387
	(type 3, males)	0.0732
	(type 4, males)	0.0732
	(type 5, males)	0.0154
	(type 6, males)	0.0264
Lee and Wolpin (2006)	blue collars/goods	0.027
	blue collars/service	0.044
	white collars/goods	0.054
	white collars/service	0.076
	Pink collars/goods	0.046
	pink collars/service	0.047

*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 2
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 (0.000)	0.101 (0.033)
	1930-39 cohort	0.063 (0.000)	0.060 (0.030)
	1940-49 cohort	0.052 (0.000)	0.078 (0.039)
Card (1995)	NLS young Men,1966 cohort	0.073 (0.004)	0.132 (0.05)
Lemiex and Card (2000)	1971/81 Canadian Census	0.070 (0.002)	0.164 (0.053)
Stock and Staiger (1997)	1980 US Census men		
	1930-39 cohort	0.063 (0.000)	0.098 (0.015)
	1940-49 cohort	0.052 (0.000)	0.088 (0.015)