Econometrica, Vol. 70, No. 5 (September, 2002), 2075–2091

UNOBSERVED ABILITY AND THE RETURN TO SCHOOLING

BY CHRISTIAN BELZIL AND JÖRGEN HANSEN[1]

1. INTRODUCTION

IN THE EMPIRICAL LITERATURE, it is customary to estimate the return to schooling by ordinary least squares (OLS) or Instrumental Variable (IV) techniques.[2] The choice of OLS is justified only if realized schooling and unobserved labor market ability are uncorrelated. If not, OLS estimates suffer the "Ability Bias" and other estimation methods, such as IV, may be used. The volume of work devoted to the return to schooling is a good indication of the importance of this topic as perceived by economists.[3] Frequently, instrumental variables (IV) techniques are applied in a context where the instrument is only weakly correlated with schooling attainments (Staiger and Stock (1997)). As a consequence, the validity of very high returns to schooling, reported in a simple regression framework, should be seriously questioned; see Manski and Pepper (2000).[4]

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. Indeed, heterogeneity in any component of schooling choices (subjective discount rates, ability or specific taste for schooling) will lead to improper inference if the local returns are erroneously assumed to be constant. As individuals with 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. 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

[1] We would like to thank James Heckman, Thierry Magnac, Jean-Marc Robin, John Rust, a co-editor and three anonymous referees for comments and suggestions on this version or earlier versions. Belzil thanks the Social Sciences and Humanities Research Council of Canada for generous funding. The usual disclaimer applies.

[2] To use the same terminology as in the reduced-form literature, the return (local) to schooling refers to the percentage wage increase per additional year of schooling. In the paper, the terms local and marginal returns may be used interchangeably. The average return refers to the slope of the straight line between the intercept and the expected log wage at a given number of years of schooling.

[3] A World Wide Web survey of the most recent literature indicates that, since 1970, more than 200 published articles or working papers (set in a reduced-form) have been devoted to the estimation of the return to schooling or surrounding issues. Very often, studies based on IV methods conclude that the returns to schooling can be between 20% and 40% above the OLS estimates. Reported estimates around 15% per year (for the US) are not uncommon although standard errors are typically very high. See Card (1999) for a survey.

[4] Manski and Pepper (2000) obtain upper bound estimates of the return to schooling from a sample taken from the NLSY. Their results cast doubts on the high returns reported in the literature.

local returns at schooling attainments that are the most common in the sample data. In the literature, this is referred to as the "Discount Rate Bias".[5]

Estimating the true returns to schooling is therefore not a simple task. As unobserved taste for schooling cannot be identified in reduced-form models, investigating the relationships between school ability, market ability, and the return to schooling requires researchers to use a structural dynamic discrete choice model in which the wage regression is estimated flexibly and in which all heterogeneity components are allowed to be correlated. As far as we know, this has never been done.[6]

In this paper, we remove this oversight. We estimate a structural dynamic programming model of schooling decisions with unobserved heterogeneity in school ability and market ability, in which the wage regression is estimated using spline techniques. The main objective of the present paper is to obtain structural estimates of the local (and average) returns to schooling in a model specification that requires neither orthogonality between schooling attainments and market ability nor linear separability between realized schooling and unobserved taste for schooling. A second objective is to obtain structural estimates of the partial correlation between realized average returns and both school and market ability, which may help understanding of the Discount Rate Bias. A third objective is to investigate the unconditional relationship between market ability and realized schooling attainments and to evaluate the magnitude of the Ability Bias. The model is implemented on a panel of white males taken from the National Longitudinal Survey of Youth (NLSY) covering the years 1979 to 1990.

Our results reported below, cast doubt on the validity of the very high returns reported in the literature. Our estimates of the return to schooling are much below those reported in the literature and, contrary to conventional wisdom, the log wage regression is found to be convex in schooling (Card (1999)). The null hypothesis that the marginal (local) returns to schooling are constant is strongly rejected in favor of a specification where the local returns are estimated using 8 spline segments. The local returns are very low until grade 11 (1% per year or less), increase to 3.7% in grade 12, and exceed 10% only between grade 14 and grade 16. The average return (measured from grade 7) increases smoothly from 0.4% (grade 7) to 4.6% (grade 16). For a representative individual (acquiring between 12 and 13 years of schooling), the average realized return (measured from grade 7) is around 1%. When measured from grade 10, it is between 2% and 3% per year.

The convexity of the log wage regression function implies that, other things equal, those endowed with higher school ability experience higher average returns and those endowed with higher market ability experience lower average returns. Therefore, those who obtain more schooling also experience higher average returns. This indicates that estimates obtained from linear wage regressions may suffer a severe Discount Rate Bias.

[5] The term "Discount Rate Bias" is due to Lang (1993) and is admittedly confusing since it has little to do with the discount rate per se. Card (1999) conjectured that OLS (or IV) estimates will over-estimate the true return to schooling because the local return to schooling decreases with grade level and because there is a positive correlation between individual discount rates and the average return to schooling (i.e. "the disadvantaged will experience higher average returns to schooling").

[6] A relatively small number of authors have estimated structural models of schooling decisions. Keane and Wolpin (1997) have used a structural dynamic programming model of schooling and occupational decisions using a cohort of the NLSY. Eckstein and Wolpin (1999) evaluate the effect of youth employment on academic performance of young Americans. Arcidiacono (2000) has estimated a dynamic model of college choices. For a seminal paper on empirical dynamic programming methods, see Rust (1987).

At the same time, school ability and market ability are strongly correlated. As a consequence, the unconditional relationship between schooling attainment and market ability is ambiguous. Our simulations indicate that there is a positive correlation (0.28) between market ability and realized schooling and orthogonality is strongly rejected. This is compatible with the existence of a positive Ability Bias.[7]

The paper is organized as follows. Section 2 is devoted to the presentation of the empirical dynamic programming model. Section 3 contains a brief description of the sample data. The main empirical results are found in Section 4. The effects of heterogeneity on the estimates of the return to schooling are presented in Section 5 and the reliability of linear wage regression models is discussed in Section 6. The robustness of the results is analyzed in Section 7. Section 8 concludes.

2. AN EMPIRICAL DYNAMIC PROGRAMMING MODEL

In this section, we introduce the empirical dynamic programming model. Every individual i is initially endowed with family human capital (X_i), innate market and school ability (v_i^w and v_i^ξ), and a rate of time preference (ρ). Young individuals decide sequentially whether it is optimal or not to enter the labor market or to continue accumulating human capital. Individuals maximize discounted expected lifetime utility over a finite (known) horizon T. Both the instantaneous utility of being in school and the utility of work are logarithmic in income. The control variable, d_{it}, summarizes the stopping rule. When $d_{it} = 1$, an individual invests in an additional year of schooling at the beginning of period t. When $d_{it} = 0$, an individual leaves school at the beginning of period t (to enter the labor market). Every decision is made at the beginning of the period and the amount of schooling acquired by the beginning of date t is denoted S_{it}.

2.1. *Household Characteristics and the Utility of Attending School*

When in school, individuals receive income support, denoted ξ_{it}. When an individual leaves school, he loses parental support. The instantaneous utility of attending school, $U^S(.)$, is represented by the following equation:

$$U^S(.) = \ln(\xi_{it}) = X_i'\delta + \psi(S_{it}) + v_i^\xi + \varepsilon_{it}^\xi \tag{1}$$

with $\varepsilon_{it}^\xi \sim \text{i.i.d}\, N(0, \sigma_\xi^2)$, representing a stochastic utility (income) shock. The vector X_i contains the following variables: father's education, mother's education, household income, number of siblings, family composition at age 14, and regional controls. The term v_i^ξ represents individual heterogeneity (ability) affecting the utility of attending school. It is discussed in more details below. The marginal effect of schooling level on parental transfers, $\psi(.)$, is modeled using spline functions.

2.2. *Interruption of Schooling*

We assume that individuals interrupt schooling with exogenous probability $\zeta(S_t)$ and, as a consequence, the possibility to take a decision depends on a state variable I_{it}. When $I_{it} = 1$, the decision problem is frozen for one period. If $I_{it} = 0$, the decision can be made.

[7] The magnitude of the Ability Bias is also central in the empirical literature on the time series properties of wage inequality. For a recent example, see Taber (2000).

The interruption state is meant to capture events such as illness, injury, travel, temporary work, incarceration, or academic failure. When an interruption occurs, the stock of human capital remains constant over the period. The NLSY does not contain data on parental transfers and, in particular, does not allow a distinction in income received according to the interruption status. As a consequence, we ignore the distinction between income support while in school and income support when school is interrupted.[8]

2.3. *The Utility of Work and the Return to Human Capital*

Once the individual has entered the labor market, he receives monetary income w_{it} (an hourly wage rate) and a yearly employment rate, e_{it}. The instantaneous utility of labor market work, U^W, is

$$U^W(.) = \ln(w_{it} \cdot e_{it}).$$

The log wage received at time t is given by

$$\ln(w_{it}) = \varphi_1(S_{it}) + \varphi_2 \cdot Exper_{it} + \varphi_3 \cdot Exper_{it}^2 + v_i^w + \varepsilon_{it}^w \tag{2}$$

where $\varepsilon_{it}^w \sim$ i.i.d $N(0, \sigma_w^2)$ is a stochastic wage shock. $\varphi_1(.)$ is the function capturing the effect of schooling on wages. The return to experience (denoted $Exper_{it}$) is captured in φ_2 and φ_3 while v_i^w is unobserved labor market ability.

The employment rate, e_{it}, is also allowed to depend on accumulated human capital so that

$$\ln e_{it}^* = \ln \frac{1}{e_{it}} = \kappa_0 + \kappa_1 \cdot S_{it} + \kappa_2 \cdot Exper_{it} + \kappa_3 \cdot Exper_{it}^2 + \varepsilon_{it}^e \tag{3}$$

where κ_0 is an intercept term, κ_1 represents the employment security return to schooling, and both κ_2 and κ_3 represent the employment security return to experience. The random shock ε_{it}^e is normally distributed with mean 0 and variance σ_e^2. All random shocks $(\varepsilon_{it}^\xi, \varepsilon_{it}^w, \varepsilon_{it}^e)$ are assumed to be independent.

2.4. *Bellman Equations*

It is convenient to summarize the state variables in a vector (S_{it}, η_{it}) where η_{it} is itself a vector containing the interruption status (I_{it}), the utility shock (ε_{it}^ξ), the wage shock (ε_{it}^w), the employment shock (ε_{it}^e), and accumulated experience $(Exper_{it})$. We only model the decision to acquire schooling beyond 6 years (as virtually every young male in the NLSY has completed at least six years of schooling). We set T to 65 years and the maximum number of years of schooling to 22. Dropping the individual subscript, the decision to remain in school, given state variables S_t and η_t, denoted $V_t^s(S_t, \eta_t)$, can be expressed as

$$\begin{aligned} V_t^s(S_t, \eta_t) = \ln(\xi_t) + \beta\{&\zeta \cdot EV_{t+1}^I(S_{t+1}, \eta_{t+1}) \\ &+ (1-\zeta) \cdot E\,\mathrm{Max}[V_{t+1}^s(S_{t+1}, \eta_{t+1}), V_{t+1}^w(S_{t+1}, \eta_{t+1})]\} \end{aligned} \tag{4}$$

[8] When faced with a high failure probability, some individuals may spend a portion of the year in school and a residual portion out of school. As a result, identifying a real interruption from a true academic failure is tenuous. In the NLSY, we find that more than 85% of the sample has never experienced school interruption.

where $V_{t+1}^{I}(S_{t+1}, \eta_{t+1})$ denotes the value of entering the interruption status. As we cannot distinguish between income support while in school and income support when school is interrupted, $V_{t+1}^{I}(S_{t+1}, \eta_{t+1})$, can be expressed as follows.

$$V_{t+1}^{I}(S_{t+1}, \eta_{t+1}) = \log(\xi_{t+1}) + \beta\{\zeta \cdot EV_{t+2}^{I}(S_{t+2}, \eta_{t+2}) + (1-\zeta) \cdot E\mathrm{Max}[V_{t+2}^{s}(S_{t+2}, \eta_{t+2}), V_{t+2}^{w}(S_{t+2}, \eta_{t+2})]\}. \tag{5}$$

The value of stopping school (that is entering the labor market), $V_t^w(S_t, \eta_t)$, is given by

$$V_t^w(S_t) = \ln(w_{it} \cdot e_{it}) + \beta E(V_{t+1} \mid d_t = 0) \tag{6}$$

where $E(V_{t+1} \mid d_t = 0)$ is simply

$$E(V_{t+1} \mid d_t = 0) = \sum_{j=t+1}^{T} \beta^{j-(t+1)}\left(-\exp\left(\mu_j + \frac{1}{2}\sigma_e^2\right) + \varphi_1(S_j) + \varphi_2 \cdot Exper_j + \varphi_3 \cdot Exper_j^2\right) \tag{7}$$

and represents the expected utility of working from $t+1$ until T. Using the terminal value and the distributional assumptions about the stochastic shocks, the probability of choosing a particular sequence of discrete choice can readily be expressed in closed-form.

2.5. *Unobserved Ability in School and in the Market*

In our model, heterogeneity has two dimensions: heterogeneity in school ability (taste for schooling), v_i^{ξ}, and heterogeneity in market ability, v_i^w. We assume that there are K types of individuals and that each type is endowed with a pair of school and market abilities (v_k^w, v_k^{ξ}) for $k = 1, 2, \ldots, K$ and $K = 6$.[9] 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. The probabilities of belonging to type k, p_k, are estimated using logistic transforms

$$p_k = \frac{\exp(q_k)}{\sum_{j=1}^{6} \exp(q_j)}$$

with the restriction that q_6 equals 0.

2.6. *Identification*

Identification of the wage return to schooling, the employment return to schooling, and unobserved market ability is relatively straightforward given panel data on labor market wages and employment rates and, hence, does not require discussion. However, given the very small amount of observations at both very low and very high levels of schooling, it is difficult to identify the local returns below grade 10 and above grade 17. Our model being structured as a single choice dynamic model, data on both wages and schooling attainments allow us to identify the key parameters in the utility of attending school. Finally, as is well known, identification of the subjective discount rate relies on the standard assumption that preferences are time additive.[10]

[9] Belzil and Hansen (2001c) analyze the case where discount rates are also individual specific.

[10] The degree of under-identification arising in estimating structural dynamic programming models is discussed in detail in Rust (1994) and Magnac and Thesmar (2002).

2.7. *The Likelihood Function*

Constructing the likelihood function (for a given type k) is relatively straightforward. It has three components; the probability of having spent at most τ years in school (L_{1k}), the probability of entering the labor market in year $\tau+1$, at observed wage $w_{\tau+1}$ (denoted L_{2k}), and the density of observed wages and employment rates from $\tau+2$ until 1990 (denoted L_{3k}). L_{1k} can easily be evaluated using (4), (5), and (6), while L_{2k} can be factored as the product of a normal conditional probability times the marginal wage density. Finally L_{3k} is just the product of wage and employment rate densities. For a given type k, the likelihood is therefore $L_k = L_{1k} \cdot L_{2k} \cdot L_{3k}$ and the log likelihood function to be maximized is

$$\log L = \log \sum_{k=1}^{6} p_k \cdot L_k \tag{8}$$

where each p_k represents the population proportion of type k.

3. THE DATA

The sample used in the analysis is extracted from the 1979 youth cohort of the *The National Longitudinal Survey of Youth* (NLSY). The NLSY is a nationally representative sample of 12,686 Americans who were 14–21 years old as of January 1, 1979. After the initial survey, re-interviews have been conducted in each subsequent year until 1996. In this paper, we restrict our sample to white males who were age 20 or less as of January 1, 1979. We record information on education, wages, and on employment rates for each individual from the time the individual is age 16 up to December 31, 1990.

The original sample contained 3,790 white males. We lost about 17% of the sample due to missing information regarding family income and about 6% due to missing information regarding parents' education. The age limit and missing information regarding actual work experience further reduced the sample to 1,710. Summary statistics may be found in Table I.

Overall, the majority of young individuals acquire education without interruption. In our sample, only 306 individuals have experienced at least one interruption. This represents only 18% of our sample and it is along the lines of results reported in Keane and Wolpin (1997).[11] Given the age of the individuals in our sample, we assume that those who have already started to work full-time by 1990 (94% of our sample), will never return to school beyond 1990. More details can be found in Belzil and Hansen (2001a).

4. EMPIRICAL RESULTS

This section is devoted to the presentation of the empirical results. The estimates of the parameters capturing the effects of household characteristics on the utility of attending school are found in Table II and discussed in details in Belzil and Hansen (2001a).[12] In

[11] Overall, interruptions tend to be quite short. Almost half of the individuals (45%) who experienced an interruption, returned to school within one year while 73% returned within 3 years. More details may be found in Belzil and Hansen (2001a).

[12] The estimates indicate that the utility of attending school increases with the level of human capital in the household. The incidence of schooling interruption is found to be 0.0749 and indicates that, on average, 7.5% of young individuals interrupt school in a given year. The estimate of the

TABLE I
DESCRIPTIVE STATISTICS[a]

	Mean	St. Dev.	Number of Individuals
Family Income/1000	36,904	27.61	1710
Father's educ.	11.69	3.47	1710
Mother's educ.	11.67	2.46	1710
Number of siblings	3.18	2.13	1710
Prop. raised in urban areas	0.73	—	1710
Prop. raised in South	0.27	—	1710
Prop. in nuclear family	0.79	—	1710
Schooling completed (1990)	12.81	2.58	1710
Number of interruptions	0.06	0.51	1710
Duration of interruptions (year)	0.43	1.39	1710
Wage 1979 (hour)	7.36	2.43	217
Wage 1980 (hour)	7.17	2.74	422
Wage 1981 (hour)	7.18	2.75	598
Wage 1982 (hour)	7.43	3.17	819
Wage 1983 (hour)	7.35	3.21	947
Wage 1984 (hour)	7.66	3.60	1071
Wage 1985 (hour)	8.08	3.54	1060
Wage 1986 (hour)	8.75	3.87	1097
Wage 1987 (hour)	9.64	4.44	1147
Wage 1988 (hour)	10.32	4.89	1215
Wage 1989 (hour)	10.47	4.97	1232
Wage 1990 (hour)	10.99	5.23	1230
Experience 1990 (years)	8.05	11.55	1230

[a] Family income and hourly wages are reported in 1990 dollars. Family income is measured as of May, 1978. The increasing number of wage observations is explained by the increase in participation rates.

Section 4.1, we analyze the role of unobserved ability in explaining schooling attainments. We discuss the returns to schooling in Section 4.2 and the internal rates of return in Section 4.3. In Section 4.4, we evaluate the capacity of the model to fit the data on observed schooling attainments.

4.1. *The Role of Unobserved Ability in Explaining Schooling Attainments*

The importance of unobserved ability is illustrated in Table III. The intercept terms of the utility of attending school (v^{ξ}) range from −0.7318 (type 1) to −1.4904 (type 6). The intercept terms of the wage function (v^{w}) range from 2.1345 (type 1) to 1.0816 (type 6). There is clear evidence that those endowed with high school ability are also endowed with high market ability. Both in ascending order of school ability and market ability, the various types can be ranked as follows, type 1, type 3, type 2, type 5, type 4, and type 6. The associated type probabilities are as follows: 0.0541 (type 1), 0.2525 (type 2), 0.1566 (type 3), 0.3022 (type 4), 0.1249 (type 5), and 0.1098 (type 6). Type 1, type 3, and type 2 individuals appear particularly more able than other types. A deeper analysis of the correlation between ability and realized schooling attainments is delayed to Section 5.

subjective discount rate, 0.0299, appears quite reasonable and is close to estimates reported in the financial economics literature; see Kocherlakota (1996).

TABLE II

THE UTILITY OF ATTENDING SCHOOL, SUBJECTIVE DISCOUNT RATES AND INTERRUPTION PROBABILITIES

	Parameter	St. Error
Utility in School		
Father's Education	0.0094	0.0011
Mother's Education	0.0070	0.0012
Family Income/1000	0.0007	0.0001
Nuclear Family	0.0204	0.0056
Siblings	−0.0071	0.0012
Rural	−0.0058	0.0048
South	−0.0176	0.0050
Standard Deviation (σ_ξ)	0.2251	0.0160
Education Splines		
δ_{7-10}	−0.0743	0.0133
δ_{11}	−0.0494	0.0120
δ_{12}	−1.1676	0.0190
δ_{13}	0.2486	0.0293
δ_{14}	1.4286	0.0345
δ_{15}	−0.1151	0.0208
δ_{16}	0.3001	0.0209
$\delta_{17-more}$	−0.7227	0.0168
Interruption Probability	0.0749	0.0031
Discount Rate	0.0299	0.0009
Mean Log Likelihood	−13.9957	

4.2. *The Return to Schooling*

The wage and employment returns to schooling are found in Table IV. Our findings are consistent with what is normally expected; education reduces unemployment and raises wages.[13] We note that the flexibility of our estimation method discloses the weakness of model specifications where the return to schooling is captured in a single parameter. The level of significance of all spline estimates (as indicated by the standard errors) indicates that, at all 8 knots joining the various segments, equality of the local returns across successive segments is strongly rejected.[14] This implies a clear rejection of the model with constant marginal returns to schooling. Our estimates of the local returns, summarized in Table IV, indicate that the marginal returns are generally increasing with the level of schooling up to grade 14. The local returns to college training are substantially higher than the returns to high-school education. Indeed, schooling has practically no value until grade 12. Until grade 10, the local return is below 1% per year (0.4%). It increases to 1.2% in grade 11 and to 3.7% in grade 12. Beyond high school graduation, the local return

[13] The estimate of the effects of education (−0.1331) and actual experience (−0.0156) on the log inverse employment rate indicate that there is a clear negative (positive) significant relationship between individual unemployment (employment) rates and human capital. The effect of experience squared on the log employment rate is found to be very small (and insignificant).

[14] In practice, we kept a spline at each grade level where the equality between successive returns is rejected. In the end, we found that as many as 8 splines were necessary to summarize the log wage regression function accurately.

TABLE III
THE DISTRIBUTION OF UNOBSERVED ABILITY[a]

			Param. (St. Error)	Rankings
Type 1	v_1^ξ	School Ab.	−0.7318 (0.0154)	1
	v_1^w	Market Ab.	2.1395 (0.0199)	1
	q_1	Intercept	−0.7082 (0.0021)	—
Type 2	v_2^ξ	School Ab.	−1.1021 (0.0171)	3
	v_2^w	Market Ab.	1.6787 (0.0133)	3
	q_2	Intercept	0.8329 (0.0449)	—
Type 3	v_3^ξ	School Ab.	−0.8785 (0.0146)	2
	v_3^w	Market Ab.	1.9136 (0.0173)	2
	q_3	Intercept	0.3551 (0.0411)	—
Type 4	v_4^ξ	School Ab.	−1.3206 (0.0202)	5
	v_4^w	Market Ab.	1.3774 (0.0115)	5
	q_4	Intercept	1.0127 (0.0038)	—
Type 5	v_5^ξ	School Ab.	−1.1815 (0.0193)	4
	v_5^w	Market Ab.	1.5488 (0.0082)	4
	q_5	Intercept	0.1291 (0.0230)	—
Type 6	v_6^ξ	School Ab.	−1.4904 (0.0215)	6
	v_6^w	Market Ab.	1.0816 (0.0162)	6
	q_6	Intercept	0.0 (normalized)	—

[a]The type probabilities are estimated using a logistic transform. The resulting probabilities are 0.0541 (Type 1), 0.2525 (Type 2), 0.1566 (Type 3), 0.3022 (Type 4), 0.1249 (Type 5), and 0.1098 (Type 6).

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 rises to 12.2%. In subsequent years (corresponding to graduate training), the local returns are estimated to be 8.8% per year. Until college graduation, and except for the lower local return in grade 15, the log wage regression equation is convex in schooling. This is contrary to what is often postulated (see Card (1999)). While it might be tempting to impute the convexity to the maintained assumption that the returns are unaffected by cross-sectional heterogeneity, the results reported in a companion paper (Belzil and Hansen (2001b)) indicate that it is not the case.[15]

The variations in the local returns to schooling illustrate the distortions introduced in a model built on the assumption that the local returns to schooling are constant. As the local and average returns are identical in a linear model, it is also important to compare our estimates of the average return with those obtained in the literature. The average returns are also found in Table IV. Although the average returns are naturally computed from grade 7, they may also be measured from the beginning of the period over which an individual can legally leave school (around grade 10). The low returns at

[15] Another estimation strategy would be to include AFQT scores in the intercept terms of both the utility of attending school and the log wage regression function. However, this approach could lead to an understatement of the effects of schooling on wages, if AFQT scores are themselves explained by schooling. For a discussion of the dynamic programming model set in a random coefficient framework, see Belzil and Hansen (2001b).

TABLE IV

THE RETURNS TO SCHOOLING

	Param.	(St Error)
Employment		
Intercept	−2.8173	(0.0115)
Schooling	−0.1309	(0.0029)
Experience	−0.0158	(0.0024)
Experience^2	0.0001	(0.0000)
σ_e	1.4858	(0.0102)
Wages		
σ_w	0.2881	(0.0023)
Experience	0.0884	(0.0025)
Experience^2	−0.0029	(0.0002)

	Wage Returns to Schooling				
Grade	Spline Param.	(St. Error)	Local	Av. Returns (from Gr. 7)	Av. Returns from Gr. 10
Grade 7–10	0.0040	(0.0012)	0.0040	0.0040	0.0040
Grade 11	0.0080	(0.0019)	0.0120	0.0056	0.0080
Grade 12	0.0252	(0.0025)	0.0372	0.0071	0.0177
Grade 13	0.0227	(0.0019)	0.0599	0.0147	0.0280
Grade 14	0.0670	(0.0023)	0.1269	0.0287	0.0478
Grade 15	−0.0195	(0.0035)	0.1074	0.0374	0.0577
Grade 16	0.0148	(0.0033)	0.1222	0.0459	0.0669
Grade 17+	−0.0345	(0.0027)	0.0877	0.0506	0.0696

lower levels of schooling, along with the sharp increase at grades 12, 13, 14, and 16, imply a relatively smooth increase in the average return to schooling. At college graduation, the average return (measured from grade 7) is around 4.6% per year. In the population, the realized average returns are much smaller as a typical individual would obtain between 12 and 13 years of schooling. In grade 12, the average return is less than 1% per year (0.71%) while, in grade 13, the average return is 1.47%. When measured from grade 10, the average return at high-school graduation is 1.8% per year and it is 6.7% at college graduation.

Finally, the estimates for actual experience (0.0880) and its square (−0.0030) show that our panel is sufficiently long to capture concavity in age-earnings profiles. The high return to experience is not really surprising for young workers. To summarize, our estimates of the return to schooling are much below those reported in the reduced-form literature, and interestingly, they can also be reconciled with the relatively weak correlation between education and growth found in the empirical literature.[16]

4.3. *Internal Rate of Return to Four Year College*

Using a measure of tuition for 4 year colleges, it is also possible to evaluate the internal rate of return; that is the rate of return that equates the discounted sum of direct costs and

[16] Manski and Pepper (2000) also conclude that the true returns are most probably below those reported in the literature. Rosenzweig and Wolpin (2000) argue that estimates of the return to schooling that ignore post-schooling human capital investments are likely to be unreliable. For a review of the empirical growth literature, see Topel (1999).

TABLE V
INTERNAL RATES OF RETURN FOR A 4 YEAR COLLEGE

Types	Yearly Cost ($1990) $5000	Yearly Cost ($1990) $10,000
Type 1	8.4%	6.2%
Type 2	7.0%	4.8%
Type 3	7.7%	5.5%
Type 4	6.0%	3.9%
Type 5	6.6%	4.5%
Type 6	5.1%	3.2%
Average	6.8%	4.7%

foregone wages with the increased wages due to completing college versus high school. It will provide key insights about the mechanics of the model and will enable us to compare the higher local returns to post high-school education with rates of return on other types of investments. Over the period covered by the data, the average tuition in the US was around $3,400 per year (in 1990$).[17] While other direct costs are unknown, the tuition cost is clearly a lower bound for the total cost of attending college. We have computed 2 different internal rates of return; one for a yearly cost of $5,000 per year and one for a cost of $10,000. This difference can account for factors such as college quality. In both cases, we have computed a type specific internal rate of return as well as a population average. These are found in Table V. The internal rates range from 5% to 8% in the $5,000 scenario and from 3% to 6% in the $10,000 scenario. Not surprisingly, we find that internal rates are higher for those with high ability (type 1, type 3, and type 2) than for those with low ability (type 4, type 5, and type 6). Overall, these internal rates are comparable to rates of return on risky investments (Kocherlakota (1996)).

4.4. *Goodness of Fit*

One of the most distinctive features of observed schooling attainments, is the very uneven distribution across various grade levels. Actual frequencies are presented in Table VI. They indicate that around 60% of young males have completed either 12 or 16 years of schooling. The predicted frequencies are also shown in Table VI. There is clear evidence that our model is able to predict large frequencies at grade 12 and grade 16. When compared to the actual frequencies, the predictions appear quite accurate.

5. EVALUATING THE EFFECTS OF HETEROGENEITY ON THE RETURNS TO SCHOOLING

As a next step, we computed the partial correlations between both forms of ability and realized schooling attainments generated by the parameters of the model. To do this, we generated 200,000 observations. These partial correlations represent an alternative way to illustrate the conflict between school and market ability and are useful in order to understand the Discount Rate Bias and the Ability Bias. In view of the large number of reduced-form estimates based on the assumption that the returns to schooling may be

[17] See Light and Strayer (2000).

TABLE VI
MODEL FIT: ACTUAL VS. PREDICTED SCHOOLING ATTAINMENTS

Grade Level	Predicted (%)	Actual (%)
Grade 6–8	6.0%	3.0%
Grade 9	7.7%	4.7%
Grade 10	7.3%	6.0%
Grade 11	8.5%	7.5%
Grade 12	41.0%	39.6%
Grade 13	6.9%	7.0%
Grade 14	5.8%	7.7%
Grade 15	1.7%	2.9%
Grade 16	10.0%	12.9%
Grade 17+	5.7%	7.9%

estimated using OLS, it is also interesting to investigate the unconditional relationship between market ability and the average returns as well as the correlation between market ability and realized schooling attainments. In the present model, wages are simultaneously affected by endogenous schooling and experience. While there is therefore no obvious way of measuring the bias in the local returns caused by assuming orthogonality between market ability and the regression function in our model, the correlation between unobserved market ability and realized schooling would still provide a good indication of the importance, as well as the source, of the Ability Bias.

The correlations are summarized in Table VII. The correlation between ability in school and ability in the market, $\text{Corr}(v_i^w, v_i^\xi)$, is found to be high (0.95). Holding market ability constant, we find a strong positive correlation between average returns and school ability, $\text{Corr}(v_i^\xi, \varphi(S_i)/S_i \,|_{v_i^w=\bar{v}^w}) = 0.80$. This positive correlation implies that, holding market ability and family background variables constant, those endowed with higher school ability will also experience higher average returns. This is an illustration of the discount rate bias. As expected, and contrary to school ability, the partial correlation between market ability and schooling is negative, $\text{Corr}(v_i^w, \varphi(S_i)/S_i \,|_{v_i^\xi=\bar{v}^\xi}) = -0.84$.

Our estimates of the correlation between unobserved market ability and average returns, $\text{Corr}(v_i^w, \varphi(S_i)/S_i)$, and the correlation between unobserved market ability and

TABLE VII
CORRELATIONS BETWEEN UNOBSERVED ABILITY AND THE AVERAGE RETURNS TO SCHOOLING[a]

Partial Correlations		Correlations (P values)
School Ab. and Av. Returns	$\text{Corr}\left(v_i^\xi \frac{\varphi(S_i)}{S_i} \middle| v_i^w = v^{-w}\right)$	0.80 (0.001)
Market Ab. and Av. Returns	$\text{Corr}\left(v_i^w \frac{\varphi(S_i)}{S_i} \middle| v_i^\xi = v^{-\xi}\right)$	−0.84 (0.001)
Correlations		
School Ab. and Market Ab.	$\text{Corr}(v_i^w, v_i^\xi)$	0.95 (0.001)
Market Ab. and Av. Returns	$\text{Corr}\left(v_i^w, \frac{\varphi(S_i)}{S_i}\right)$	0.23 (0.001)
Market Ab. and Schooling	$\text{Corr}(v_i^w, S_i)$	0.28 (0.001)

[a]The correlations are computed from an artificial sample containing 200,000 simulated schooling attainments.

realized schooling, $\text{Corr}(v_i^w, S_i)$, are both positive. They are respectively equal to 0.23 and 0.28 and are significant at the 1% level. 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. Estimation methods based on the maintained assumption that realized schooling and market ability are orthogonal are therefore questionable. They will clearly lead to an over-estimation of the average return to schooling.

6. CAN A LOG LINEAR WAGE REGRESSION MODEL SUMMARIZE THE AVERAGE RETURNS ACCURATELY?

Estimating the return to schooling and experience in a reduced-form framework is particularly difficult since, in general, schooling and experience cannot be separated linearly from unobserved ability components. To see this, re-write the wage regression function as

$$\ln(w_{it}) = \varphi_1(S_{it}(v_i^\xi, v_i^w)) + \varphi_2 \cdot Exper_{it}(v_i^\xi, v_i^w) + \varphi_3 \cdot Exper_{it}^2(v_i^\xi, v_i^w) + v_i^w + \varepsilon_{it}^w. \tag{9}$$

In the literature, it is usually assumed that $\varphi_1(\cdot)$ is linear in schooling and that S_{it} and v_i^ξ (or its reduced-form equivalent) are linearly separable. For those who seek to estimate the true returns in an instrumental variable framework, these assumptions are particularly crucial. If the linear log wage regression is not supported by the data and the form of the wage regression function is unknown, the estimation method is more complicated.[18] Furthermore, if unobserved ability and schooling cannot be separated linearly, this is even more difficult.[19] However, while a linear regression framework is incapable of capturing changes in the local returns to schooling, it might be argued that it can still estimate the average return accurately.

In Table VIII, we compare the maximum likelihood estimates of the average return to schooling obtained from the structural dynamic programming model to those obtained using OLS estimates. We considered the standard OLS estimate as well as the OLS estimate with splines (with the same spline segments as in the structural dynamic programming model). We report the estimates of the average return to schooling at high school graduation (grade 12) and at college graduation (grade 16). For both of these, we report values measured from grade 7 and values measured from grade 10.

The incapacity of OLS estimates to capture the average return over the entire range is well documented. In the 1990 cross-section of the NLSY, the OLS estimate of the return to schooling is 0.0995. When OLS is applied to the entire panel, the estimate is 0.0998. When the OLS regression is composed of 8 splines, the average returns measured from grade 7 are equal to 9.71% in grade 12 and 11.29% in grade 16. When measured from grade 10, the average returns are 8.66% (grade 12) and 9.42% (grade 16). As a comparison, our structural estimates of the average return at high school graduation (0.71%) and at college graduation (4.6%), measured from grade 7, are much below the OLS estimates. This is also true of those average returns measured from grade 10 (1.77% in grade 12 and 6.69% in grade 16). Evidently, estimation methods that do not allow for a flexible estimation of the local returns to schooling will lead to unreliable estimates of both the local and the average returns to schooling.

[18] See Newey, Powell, and Vella (1999) for a discussion of nonparametric methods applied in a context where regressors are potentially endogenous (a static labor supply model) and Blundell and Powell (2000) for an enlightening survey.

[19] This issue is discussed in Heckman and Vitlacyl (2000).

TABLE VIII

ESTIMATES OF THE AVERAGE RETURN TO SCHOOLING[a]

	Average Returns Computed from Gr. 7		Average Returns Computed from Gr. 10	
Estimation Method	(Gr. 12)	(Gr. 16)	(Gr. 12)	(Gr. 16)
Dynamic Prog. (ML)	0.0071	0.0459	0.0177	0.0699
OLS (Linear)	0.0995	0.0995	0.0995	0.0995
OLS (Splines)	0.0970	0.1129	0.0866	0.0942

[a]The OLS estimates are computed on the cross-section of 1990. Both OLS regressions contain experience and experience^2. The OLS regression with splines contains 8 segments (grade 7 to 10, grade 11, grade 12, grade 13, grade 14, grade 15, grade 16, and grade 17 and more).

7. ROBUSTNESS AND ALTERNATIVE SPECIFICATIONS

At this stage, it is natural to investigate the robustness of the results. In our model, the effects of parents' human capital are asymmetric: household background variables affect the utility of attending school but do not affect labor market skills per se. This assumption might be questioned although evidence presented in Belzil and Hansen (2001c) suggests that, after conditioning on schooling, individual differences in family background, variables account for a very small portion of the variation in lifetime wages.[20] To evaluate the sensitivity of the results, we estimated other model specifications that allow for a more symmetric treatment of family background variables. In particular, and consistent with Keane and Wolpin (1997) and Eckstein and Wolpin (1999), we estimated a finite mixture version of the model (with 6 unknown types) that ignores data on household characteristics. In this model specification, asymmetry is obviously not an issue. The flexibility of the specification allows for any arbitrary correlation between the utility of attending school and market ability. The estimates of the wage regression function are found in Table IX.

There is overwhelming evidence that the basic results (the low returns and the convexity of the wage regression) are in no way affected by the treatment of family background variables, as indicated by the local and average returns. Generally speaking, the returns obtained in a mixture model are only slightly higher than those obtained with parents' background, especially until grade 14. As before, the returns to schooling are low until high school graduation (4.8% in grade 12) and increase significantly in grade 14 (to reach 14.5%). The local return in grade 16 (11%) is however lower than in the model using parents' background variables. The high degree of convexity of the wage regression is therefore very robust. Other parameter estimates as well as a more in-depth discussion may be found in Belzil and Hansen (2001a).

We have also estimated another version of the model where all household background variables affect both the utility of attending school and the wage regression equation and where the employment equation is affected by unobserved heterogeneity (6 individual specific intercept terms). The returns to schooling have been found to be generally lower (2% in grade 12) but the wage regression function still displays a high degree of convexity (the local return at college graduation is 11.7%). Not surprisingly, the effect of family background on the utility of attending school has been found to be much stronger than on

[20] Belzil and Hansen (2001c) find that 85% of the explained variation in schooling attainments comes from household human capital and 15% from unobserved abilities. However, around 75% of the variation in entry wages is explained by unobserved labor market ability.

TABLE IX
ESTIMATES OF THE WAGE REGRESSION FUNCTION UNDER AN ALTERNATIVE MODEL SPECIFICATION[a]

Grade	Nonparametric Mixture Model Splines Param. (St. Error)	Local Returns	Av. Returns (from Gr. 7)	Av. Returns (Gr. 10)
Grade 7–10	0.0041 (0.0009)	0.0041	0.0041	0.0041
Grade 11	0.0093 (0.0034)	0.0134	0.0060	0.0088
Grade 12	0.0346 (0.0027)	0.0480	0.0130	0.0218
Grade 13	0.0135 (0.0028)	0.0615	0.0200	0.0318
Grade 14	0.0839 (0.0034)	0.1454	0.0359	0.0545
Grade 15	−0.0365 (0.0034)	0.1089	0.0437	0.0636
Grade 16	0.0034 (0.0040)	0.1123	0.0506	0.0705
Grade 17+	−0.0367 (0.0040)	0.0746	0.0528	0.0710
Exper.	0.0973 (0.0030)		—	—
Exper^2	−0.0039 (0.0003)		—	—
Mean Log Likelihood	−14.1648			

[a]The mixture model is estimated under the assumption that the population is composed of 6 unknown types.

wages. This issue, as well as other matters that are not directly related to the estimation of the returns to schooling, are discussed in detail in a companion paper, Belzil and Hansen (2001c).

8. CONCLUSION

We have estimated a structural dynamic programming model of schooling decisions in which individuals are heterogeneous with respect to ability in school and ability in the labor market. The rich specification of the model has allowed us to estimate the true return to schooling without assuming orthogonality between labor market ability and schooling attainments and to obtain structural estimates of various correlations between school ability, market ability, and realized schooling attainments.

We find that the local returns are very small at low levels of schooling and increase significantly beyond high school graduation. Indeed, schooling has practically no value until high school graduation and the wage regression function displays a high degree of convexity. We also find that a linear wage regression is severely misspecified (the estimates suffer a severe discount rate bias) and that, in general, any regression model (whether linear or nonlinear), estimated under the assumption that realized schooling and market ability are orthogonal, suffers a positive ability bias. The results are very robust. A finite mixture model (with 6 types) ignoring data on parents' background as well as a model specification in which parents' human capital affect both labor market ability and the utility of attending school gave very similar results. The basic findings (the low returns to schooling and the very high degree of convexity of the wage regression) are in no way related to the treatment of household background variables.

Overall, our results cast serious doubts on the validity of most results reported in the empirical labor economics literature. First, OLS estimates of the return to schooling which depend on the assumption that endogenous schooling attainment is orthogonal to unobserved determinants of wages are therefore most likely unreliable. Clearly, they can seriously over-estimate the average return to schooling. This has often been postulated but rarely verified. Second, the convexity of the log wage regression function, along with the nonseparability of realized schooling and abilities, imply that standard IV techniques are also ill-equipped to tackle the estimation of the return to schooling. As it stands now, the development of nonparametric (or semiparametric) econometric models with endogenous regressors is a major topic of ongoing research (see Blundell and Powell (2000)). These techniques are certainly not used widely by labor economists and, as a consequence, the validity of the reduced-form results already reported in the literature can seriously be questioned.

As is done in most of the literature, we have assumed that differences in market ability can be captured in the intercept of the log wage regression function and that individual differences in realized returns to schooling are explained by nonlinearities in the wage regression function. This is an adhoc assumption. Heterogeneity in the returns to schooling may also arise in a model characterized by the allowance for absolute and comparative advantages. Indeed, in a companion paper (Belzil and Hansen (2001b)), we estimate a finite mixture dynamic programming model of schooling decisions in which the log wage regression function is set in a random coefficient framework and in which the population is composed of 8 unknown types.[21] However, a more challenging question would be to determine which heterogeneity component (cross-sectional skill heterogeneity or nonlinearities) is more important. This is likely to be difficult. Confronting a nonlinear wage regression with a random coefficient is not simple. Cross-sectional heterogeneity and nonlinearities are difficult to identify separately, unless observed measures of ability are available. As a consequence, the empirical investigation of wage regression models in which unobserved market skill heterogeneity affects both the returns to schooling and the return to experience and in which the local returns may vary with grade level, is a promising avenue for future research.

Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. W., Montreal, QC, H3G 1M8, Canada, CIRANO, CIREQ, and IZA; belzilc@vax2.concordia.ca; http://vax2.concordia/~belzilc/

and

Department of Economics, Concordia University, 1455 de Maisonneuve Blvd. W., Montreal, QC, H3G 1M8, Canada, CEPR, CIRANO, CIREQ, and IZA; hansen@vax2.concordia.ca; http://vax2.concordia/~hansen/

Manuscript received May, 1999; final revision received November, 2001.

REFERENCES

ARCIDIACONO, P. (2000): "Option Values, College Quality, and Earnings: A Dynamic Model of College and Major Choice," Working Paper, Duke University.

BELZIL, C., AND J. HANSEN (2001a): "Unobserved Ability and the Return to Schooling," Working Paper, Concordia University.

[21] In the paper, we also investigate how individual specific reactions to a generous counterfactual college attendance subsidy are correlated with individual absolute and comparative advantages and examine the veracity of a popular claim often found in the treatment effects literature: in order to reconcile the discrepancy between OLS and IV estimates of the return to schooling, those individuals more likely to react to an exogenous policy change must have much higher returns to schooling than average.

——— (2001b): "A Structural Analysis of the Correlated Random Coefficient Wage Regression Model," Working Paper, Concordia University.

——— (2001c): "The Intergenerational Education Correlation and the Rate of Time Preference," Working Paper, Concordia University and Cirano.

BLUNDELL, R., AND J. POWELL (2000): "Endogeneity in Non-Parametric and Semi-Parametric Regression Models," Paper presented at the 2000 World Congress of the Econometric Society.

CARD, D. (1999): "The Causal Effect of Education on Earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card. Amsterdam: Elsevier Science, North-Holland Publishers, 1801–1863.

ECKSTEIN, Z., AND K. WOLPIN (1999): "Youth Employment and Academic Performance in High School," *Econometrica*, 67, 1295–1339.

HECKMAN, J., AND E. VYTLACIL (1998): "Instrumental Variables Methods for the Correlated Random Coefficient Model," *Journal of Human Resources*, 33, 974–987.

——— (2000): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," Working Paper, Department of Economics, University of Chicago.

KEANE, M. P., AND K. WOLPIN (1997): "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473–522.

KOCHERLAKOTA, N. (1996): "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature*, 24, 42–71.

LANG, K. (1993): "Ability Bias, Discount Rate Bias and the Return to Education," Working Paper, Department of Economics, Boston University.

LIGHT, A., AND W. STRAYER (2000): "Determinants of College Completion: School Quality or Student Ability," *Journal of Human Resources*, 35, 299–332.

MAGNAC, T., AND D. THESMAR (2002): "Identifying Dynamic Discrete Decision Process," *Econometrica*, 70, 801–816.

MANSKI, C., AND J. PEPPER (2000): "Monotone Instrumental Variables: with an Application to the Returns to Schooling," *Econometrica*, 68, 997–1013.

NEWEY, W., J. L. POWELL, AND E. VELLA (1999): "Nonparametric Estimation of a Triangular Simultaneous Equation System," *Econometrica*, 67, 565–604.

ROSENZWEIG, M., AND K. WOLPIN (2000): "Natural Natural Experiments in Economics," *Journal of Economic Literature*, 38, 827–874.

RUST, J. (1987): "Optimal Replacement of GMC Bus Engines: An Empirical Analysis of Harold Zurcher," *Econometrica*, 55, 999–1033.

——— (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.

STAIGER, D., AND J. H. STOCK (1997): "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

TABER, C. (2000): "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability," forthcoming in *Review of Economic Studies*.

TOPEL, R. (1999): "Labor Markets and Economic Growth," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card. Amsterdam: Elsevier Science, North-Holland Publishers, 2943–2984.