2001s-10

Heterogeneous Returns to Human Capital and Dynamic Self-Selection

Christian Belzil, Jörgen Hansen

Série Scientifique
Scientific Series

CIRANO
Centre interuniversitaire de recherche
en analyse des organisations

Montréal
Février 2001

ISSN 1198-8177

Heterogeneous Returns to Human Capital and Dynamic Self-Selection*

Christian Belzil†, Jörgen Hansen‡

Résumé / Abstract

Dans cet article, nous estimons un modèle de programmation dynamique des choix en éducation dans lequel les rendements moyens et marginaux (en éducation et en expérience) sont propres à chaque individu. Nos résultats indiquent une forte corrélation positive entre rendements en éducation et rendements en expérience. Après avoir intégré les effets individuels aléatoires, la fonction de salaire est de forme convexe (les rendements en éducation croissent avec l'éducation). Les effets antagonistes des rendements en éducation et en expérience impliquent une très faible corrélation entre les rendements individuels et l'éducation observée.

We estimate a structural dynamic programming model of schooling decisions and obtain individual specific estimates of the local (and average) returns to schooling as well as the returns to experience. Homogeneity of the returns to human capital is strongly rejected in favor of a discrete distribution version of the random coefficient specification. The results indicate that individuals who have the higher returns to schooling are also those who have the higher returns to experience. There is a 5.9 percentage points difference in the average return to schooling at college graduation between high and low market ability individuals (2.3% vs 8.2%) and a 5.4 percentage points difference in the return to experience upon entrance in the labor market (3.1% vs 8.5%). When averaged over all types, the return to experience in the early phase of the life cycle (6.8%) exceeds the average return to schooling (6.4% at college graduation). After conditioning on a specific type, the log wage regression function remains rather convex in schooling. The conflictual effects of the returns to schooling and experience on schooling decisions imply weak dynamic self-selection; that is educational attainments are only weakly correlated with individual differences in the returns to schooling.

---

* Corresponding Author: Christian Belzil, CIRANO, 2020 University Street, 25th floor, Montréal, Qc, Canada H3A 2A5 Tel.: (514) 985-4000 Fax: (514) 985-4039 email: belzilc@cirano.qc.ca
We would like to thank seminar participants at University of Toronto. Belzil thanks the Social Sciences and Humanities Research Council of Canada for generous support.

† Concordia University and CIRANO

‡ Concordia University

Mots Clés : Rendements locaux en éducation, rendements moyens, programmation dynamique, auto-sélection dynamique

Keywords: Local returns to schooling, average returns, dynamic programming, dynamic self-selection

JEL: J2, J3

1 Introduction

The effect of schooling on wages is one of the most widely studied topic in empirical economics. Whether set in a reduced-form framework or in a structural framework, empirical models are usually based on the ad-hoc assumptions that individual differences in market ability can be captured in the intercept term of the wage regression function and that log wages vary linearly with schooling. The validity of both of these assumptions is however starting to be questioned seriously by empirical labor economists.

First, when individual differences in market ability are reflected in the intercept term of the wage regression equation, those endowed with high market ability have a higher opportunity costs of schooling. In a more general framework, in which market ability can also affects the slope of the wage function, this argument is not necessarily true. As a consequence, it is natural to estimate the returns to schooling in a random coefficient framework, in which potential comparative advantages in schooling can be captured (see Heckman and Vytlacil,1998 and 2000).[1]

The validity of the linearity assumption is also questionable. In a log linear regression model, the local returns to schooling are assumed to be constant and estimates of the return obtained in this framework might be strongly affected by the local returns corresponding to graduation. Belzil and Hansen (2000) use a structural dynamic programming model to obtain flexible estimates of the return to schooling from the National Longitudinal Survey of Youth (NLSY) and find that a model with constant local returns is strongly rejected in favor of a convex log wage regression function composed of 8 segments.[2]

While both the possibility of non-linearities and population heterogeneity are starting to be recognized, as far as we know the returns to schooling have never been investigated in a framework which allows simultaneously for non-constant local returns as well as population heterogeneity in the returns.

[1]Indeed, the need for a random coefficient representation of the log wage equation has been recognized as early as in Becker and Chiswick (1966).

[2]The average return over the entire range (around 4% per year) is found to be much lower than what is usually reported in the reduced-form literature. The model also imply a positive correlation between market ability and realized schooling attainments (the "Ability Bias"). Taber (1999) also investigates the empirical importance of the Ability Bias.

This might be a serious drawback. If the individuals who have higher market ability also have a comparative advantage in schooling (experience higher returns to schooling) and acquire more schooling, the convexity of the wage regression function might only reflect dynamic self-selection (merely a composition effect). That is, as we move toward higher levels of schooling, the local returns to schooling may turn out to be estimated from an increasingly large proportion of high ability workers. However, allowing for individual specific returns to schooling is not sufficient to capture all dimensions of market ability. If more able individuals face higher returns to schooling, they may also face a higher return to experience. For instance, those individuals experiencing high returns to schooling may also have comparative advantages in on-the-job training. If so, a reliable estimation method must allow for ability heterogeneity to affect both the local returns to schooling and experience, while allowing the local returns to change with grade level.

The main objective of this paper is to obtain structural estimates of the local and average returns to schooling within a framework where the log wage regression function is estimated flexibly (the returns may vary with grade level) and is affected by population heterogeneity.[3] A second objective is to investigate the nature of dynamic self-selection; that is the relationship between individual specific returns to human capital and schooling attainments. Finally, a third objective is to compare the returns obtained in a random coefficient framework to those obtained in more standard framework in which the slope coefficients are homogeneous in the population. The model is implemented on a panel of white males taken from the National Longitudinal Survey of Youth (NLSY). The panel covers a period going from 1979 until 1990.

The main results are as follows. Homogeneity of the returns to human capital (schooling and experience) is strongly rejected in favor of a discrete distribution version of the random coefficient model specification. Those individuals who have the higher returns to schooling (comparative advantages in schooling) are also those who have the higher returns to experience. There is a 5.9 percentage points difference in the average return to schooling at college graduation between high and low market ability individuals (2.3% vs 8.2%) and a 5.4 percentage points difference in the return to experience

[3] The structure of the dynamic programming model is identical to Belzil and Hansen (2000).

(3.1% vs 8.5%). When averaged over all types, the return to experience in the early phase of the life cycle (6.8%) exceeds the average return to schooling (6.4% at college graduation). After conditioning on a specific type, the log wage regression function remains rather convex in schooling. The conflictual effects of the returns to schooling and experience on schooling decisions imply weak dynamic self-selection; that is differences in educational attainments are only weakly positively correlated with individual differences in the returns to schooling. We also find that a model with individual specific returns to schooling and a homogeneous return to experience performs poorly. It fails to capture a significant difference in the average return to schooling between high and low market ability individuals. This is easily understood. Those individuals endowed with a high return to schooling are also faced with a high return to experience. If differences in the returns to schooling were the only source of comparative advantages (individuals share the same return to experience), the more able would obtain a substantially higher level of schooling than those who are less able. However, such a positive correlation between market ability and schooling attainments is not born by the data and, as a consequence, the likelihood estimates indicate a minimal level of heterogeneity in the returns to schooling.

The paper is structured as follows. The empirical dynamic programming model is exposed in Section 2. A brief description of the sample data is found in Section 3. The structural parameter estimates are discussed in Section 4 and the goodness of fit is evaluated in Section 5. Section 6 is devoted to the empirical analysis of the role of unobserved labor market ability in explaining dynamic self-selection. In Section 7, we present a statistical test for the random coefficient specification and discuss briefly an alternative model specification which ignores heterogeneity in the return to experience. Section 8 is devoted to a comparison of our estimates with those obtained by OLS as well as structural parameter estimates ignoring population heterogeneity. The conclusion is in Section 9.

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 and a rate of time preference (ρ). Young individuals decide sequentially whether it is optimal or not to enter the labor market or continue accumulate human capital. Individuals maximize discounted expected lifetime utility over a finite horizon T and have identical preferences. Both the instantaneous utility of being in school and the utility of work are logarithmic. 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 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}. These transfers are understood to be net of direct costs (such as books, transportation or other costs). When an individual leaves school, he looses parental support. The instantaneous utility of attending school, $\ln(\xi_{it})$, is represented by the following equation

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

with $\varepsilon_{it}^{\xi} \sim i.i.d\ N(0, \sigma_{\xi}^2)$ and represents a stochastic utility 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 number of siblings is used to control for the fact that, other things equal, the amount of parental resources spent per child decreases with the number of siblings. The household composition variable (Nuclear Family) is equal to 1 for those who lived with both their biological parents (at age 14) and is likely to be correlated with the psychic costs of attending school. The geographical variables are introduced in order to control for the possibility that direct (as well as psychic) costs of schooling may differ between those raised in urban areas and those raised in rural areas, and between those raised in the south and those raised in the north. Yearly family income is measured in units of \$1,000. The term υ_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 ζ 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. 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.[4]

2.3 The Return to Human Capital

The log wage received by individual i, at time t, is given by

$$\log w_{it} = \varphi_{1i}^m(S_{it}) + \varphi_{2i}^m.Exper_{it} + \varphi_3^m.Exper_{it}^2 + \ \upsilon_i^w + \varepsilon_{it}^w \tag{2}$$

where $\varphi_{1i}^m(S_t)$ is the individual specific function representing the wage return to schooling. Both φ_{2i}^m and φ_3^m are parameters to be estimated and υ_i^w is unobserved labor market ability. As we do not observe wage data over the entire lifetime, it is difficult to identify individual specific quadratic terms. As a consequence, only the linear term in experience is allowed to be individual

[4]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.

specific. The non-wage benefit is assumed to be log-linear in schooling, that is

$$\log w_{it}^* = \varphi_0^{nm} + \varphi_1^{nm} \cdot S_{it}$$

where φ_0^{nm} and φ_1^{nm} are parameters to be estimated. The employment rate, e_{it}, is also allowed to depend on accumulated human capital (S_{it} and $Exper_{it}$) so that

$$\ln e_{it}^* = \ln \frac{1}{e_{it}} = \kappa_{0i} + \kappa_1 \cdot S_{it} + \kappa_2 \cdot Exper_{it} \ + \kappa_3 \cdot Exper_{it}^2 + \varepsilon_{it}^e \qquad (3)$$

where κ_{0i} is an individual specific intercept term, κ_1 represents the employment security return to schooling, both κ_2 and κ_3 represent the employment security return to experience.[5] 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 individual in the sample 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

$$V_t^s(S_t, \eta_t) = \ln(\xi_t) + \beta\{\zeta \cdot EV_{t+1}^I(S_{t+1}, \eta_{t+1}) \qquad (4)$$

$$+(1-\zeta) \cdot EMax[V_{t+1}^s(S_{t+1}, \eta_{t+1}), V_{t+1}^w(S_{t+1}, \eta_{t+1})]\}$$

[5]It follows that the expected value and the variance of the log employment rate are given by $E \log e_t = -\exp(\mu_t + \frac{1}{2}\sigma_e^2)$ and $Var(\log e_t) = \exp(2\mu_t + \sigma_e^2) \cdot (\exp(\sigma_e^2) - 1)$ respectively.

where $V^I_{t+1}(S_{t+1}, \eta_{t+1})$ denotes the value of interrupting schooling acquisition. As we cannot distinguish between income support while in school and income support when school is interrupted, the value of interrupting schooling acquisition is identical to the value of attending school. The value of stopping school (that is entering the labor market), $V^w_t(S_t, \eta_t)$, is given by

$$V^w_t(S_t) = \ln(w_{it} \cdot w^*_{it} \cdot e_{it}) + \beta E(V_{t+1} \mid d_t = 0) \tag{5}$$

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

The intercept terms of the utility of attending school (υ^ξ_i), the employment rate equation (κ_{0i}) and of the log wage regression function (υ^w_i) are individual specific. As well, we allow the local returns to schooling $\varphi^m_{1i}(S_{it})$ and the effect of experience φ^m_{2i} to vary across individuals. We assume that there are K types of individuals. Each type is endowed with a vector of intercept terms $(\upsilon^w_k, \upsilon^\xi_k, \kappa_{0k})$ for $k = 1, 2...K$. The results reported in this paper are for the case where $K = 6$. However, it is unrealistic to try to identify 6 different functions representing the local returns to schooling as well as 6 different returns to experience. As a consequence, we assume that the individual specific returns to schooling and experience can be summarized in 2 different functions; one for Group A (types 1, 2 and 3) and one for Group B (types 4, 5 and 6). That is

- $\varphi^m_{1k}(.) = \varphi^m_{1A}(.)$ for $k = 1, 2$ and 3
- $\varphi^m_{1k}(.) = \varphi^m_{1B}(.)$ for $k = 4, 5$ and 6.
- $\varphi^m_{2k}(.) = \varphi^m_{2A}(.)$ for $k = 1, 2$ and 3
- $\varphi^m_{2k}(.) = \varphi^m_{2B}(.)$ for $k = 4, 5$ and 6.

The distribution of unobserved ability is orthogonal to parents' background by construction and, as a consequence, should be understood as a measure of unobserved ability remaining after conditioning on parents human capital. The probability of belonging to type k, p_k, are estimated using logistic transforms

$$p_k = \frac{\exp(q_k^0)}{\sum_{j=1}^{6} \exp(q_j^0)}$$

and with the restriction that $q_6 = 0$.

2.6 Identification

As discussed in Belzil and Hansen (2000), identification of most parameters is relatively straightforward. Nevertheless, estimation of our model will require normalization. Given the absence of data on non-wage benefits, it is impossible to separate the intercept term of the non-wage benefit equation (common to every individual) from the intercept term of the utility of attending school. As a consequence, the intercept term of the non-wage benefit must be absorbed in the utility of attending school and φ_0^{nm} is set to 0. Also, as is well known, identification of the subjective discount rate relies on the standard assumption that preferences are time additive. Finally, it also important to note that, given the relatively modest number of individuals at both very low and very high levels of schooling, it is difficult to identify more than two different regression functions. This is a consequence of our flexible specification of the log wage regression function.

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) and (5), 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 wages densities (2) and employments densities (3). 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

$$\log L = \log \sum_{k=1}^{6} p_k \cdot L_k \qquad (6)$$

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. However, we lacked information on family background variables (such as family income as of 1978 and parents' education). 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.

Descriptive statistics for the sample used in the estimation can be found in Table 1. The education length variable is the reported highest grade completed as of May 1 of the survey year and individuals are also asked if they are currently enrolled in school or not.[6] This question allows us to identify those individuals who are still acquiring schooling and therefore to take into account that education length is right-censored for some individuals. It also helps us to identify those individuals who have interrupted schooling. Overall, the majority of young individuals acquire education without interruption. The low incidence of interruptions (Table 1) explains the low average number of interruptions per individual (0.22) and the very low average interruption duration (0.43 year) . In our sample, only 306 individuals have experienced at least one interruption. This represents only 18% of our sample and it is

[6]This feature of the NLSY implies that there is a relatively low level of measurement error in the education variable.

along the lines of results reported in Keane and Wolpin (1997).[7] 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. Finally, one notes that the number of interruptions is relatively small.

Unlike many reduced-form studies which use proxies for post-schooling labor market experience (see Rosenzweig and Wolpin), we use actual labor market experience. Actual experience accumulated is computed using the fraction of the year worked by a given individual. The availability of data on actual employment rates allows use to estimate the employment security return to schooling.

The average schooling completed (by 1990) is 12.8 years. As described in Belzil and Hansen (2000), it is clear that the distribution of schooling attainments is bimodal. There is a large fraction of young individuals who terminate school after 12 years (high school graduation). The next largest frequency is at 16 years and corresponds to college graduation. Altogether, more than half of the sample has obtained either 12 or 16 years of schooling. As a consequence, one might expect that either the wage return to schooling or the parental transfers vary substantially with grade level. This question will be addressed below.

4 Structural Estimates of the Return to Human Capital

The parameter estimates surrounding the utility of attending school, the subjective discount rate and the interruption probability are found in Table 2A. The estimates are very close to those reported in Belzil and Hansen (2000) and we do not discuss them in details.[8] The parameter estimates character-

[7] 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.

[8] The estimates indicate that, other things equal, the utility of attending school increases with parents' education and income. These results are standard in the literature. While the results indicate that mean schooling attainments are increasing with family human capital, they illustrate the relatively weak correlation between parents' human capital and individuals schooling attainments. This is explained by the fact that unobserved school

izing the distribution of all individual specific intercept terms (school ability, employment and wage regression) are found in Table 2B. The estimates of the logistic transforms used to infer the type proportions are also in Table 2B.

The structural estimates of the return to human capital are found in Table 3A. To set the number of splines, we experienced with a larger number of segments (up to 9) and remove the splines that were less significant.[9] As a result, we end up with 6 segments. The local returns are constant from grade 7 to grade 12 and change with grade level between grade 13 and 17. The spline estimates of the local returns to schooling are found in Table 3A. These estimates have been transformed into local returns (after adding up the proper parameters). For each grade level, a corresponding average return has also been computed. The local and average returns are reported in Table 3B. They are analyzed in details below.

An examination of the intercept terms of the wage equation, the employment equation and the utility of attending school (Table 2B) reveals that heterogeneity in employment rates and school ability is relatively more important than heterogeneity in the wage intercept. This is a consequence of allowing both the returns to schooling and experience to be individual specific. Indeed, the dispersion in the wage returns to schooling and experience (Table 3A and Table 3B) should be taken as strong evidence in favor of a random coefficient specification.

The results indicate clearly that those individuals endowed with higher returns to schooling (group B) are also endowed with a higher return to experience. Within groups, the local returns are generally increasing with grade level. For those belonging to group A, the local returns are 0.0048 (grade 7 to 12), 0.0253 (grade 13), 0.0709 (grade 14), 0.0486 (grade 15), 0.0596 (grade 16) and 0.0553 (grade 17-more). The average return to schooling increases smoothly from 0.48% in grade 7 up to 2.33% at college graduation (grade 16). For those belonging to group B, the local returns are 0.0614 between grade 7 to 12, 0.0908 in grade 13, 0.1278 in grade 14, 0.1142 in grade 15, 0.1206

ability plays an important part in explaining individual schooling attainments. Similar results are reported in Belzil and Hansen (2000), Eckstein and Wolpin (1999) and Keane and Wolpin (1997).

[9] As in Belzil and Hansen (2000), we found that the local returns to schooling vary much more beyond high school graduation (from grade 13 onward) than before high school graduation (between grade 7 and 12).

in grade 16 and 0.1210 beyond grade 16. These local estimates also imply a smooth increase in the average return. The average returns range from 6.14% in grade 7 to 8.57% at college graduation. While there is a large difference between the returns to schooling of low ability and high ability workers, each type specific log wage regression function discloses the same tendency for the local returns to increase with grade level. As a consequence, the convexity of the log wage regression function reported in Belzil and Hansen (2000) does not seem to be explained by a composition effect. It appears robust to the allowance for population heterogeneity in the returns to human capital.

Interestingly, the difference in average returns to schooling (around 0.059) is quite close to the difference in the returns to experience. For those individuals endowed with low return to schooling (group A), the return to experience upon entrance in the labor market is 0.0308. The return to experience for individuals belonging to group B is substantially higher; it is found to be 0.0850. Since we are restricting the quadratic terms to be equal (in order to facilitate comparison), individual differences in the return to experience are captured solely in the experience parameter. These estimates imply that the return to every additional year of experience, in the early phase of the life cycle, exceeds the average return to schooling.[10]

5 Accuracy of Predicted Schooling

As indicated earlier, the empirical distribution of schooling attainments discloses an important clustering around grade levels corresponding to high school graduation (grade 12) and college graduation (grade 16). The actual schooling attainment frequencies are reported in the second column of Table 5 while schooling attainments predicted by the structural model are found in column 1. There is clear evidence that the flexible specification of our model allows us to predict schooling attainments accurately. While our predictions are slightly less accurate at very low levels of schooling (grade 6 to grade 8) and high grade levels (grade 18 or more), they are particularly accurate at those grade levels corresponding to high school and college graduation. Overall, our model fits data on schooling attainments very well. While the

[10]This results was also obtained with homogeneous returns to schooling (see Belzil and Hansen, 2000).

model is arguably stylized, it seems to capture the essential features in the data.

6 Labor Market Ability and Dynamic Self-Selection

In standard log wage regression models, where ability heterogeneity is captured in the intercept term and where every individuals face the same return to schooling, higher labor market ability implies a higher opportunity cost of being in school. In a random coefficient framework, the argument no longer follows. Those individuals who are able to transform schooling inputs into a higher level of human capital, will benefit from higher returns to education and are most likely those who will attain high schooling attainments. If so, a sub-population of highly educated workers may tend to be composed of a majority of high market ability workers who may have higher returns to schooling. At the same time, individuals who have a higher return to experience will be impatient to enter the labor market and experience upward sloping wage profiles. If those who face high returns to schooling are also those who face high returns to experience, differences in the returns to schooling and experience may counterbalance each other. The links between market ability and schooling is therefore ambiguous.

The type probabilities can be used to compute the correlation between various individual specific intercept terms as well as the correlation between school ability and the return to human capital. These are found in Table 4A. Overall, the correlations are all of the expected sign. The correlation between school ability and the wage intercept ($corr(\upsilon^{\xi}, \upsilon^{w})$), the correlation between school ability and the employment rate intercept ($corr(\upsilon^{\xi}, -\kappa_0)$) and the correlation between the wage intercept and the employment rate intercept ($corr(\upsilon^{w}, -\kappa_0)$) are all found to be positive. They are equal to 0.43, 0.59 and 0.45 respectively. Not surprisingly, there is also a positive correlation (0.27) between school ability and the returns to schooling $corr(\upsilon_k^{\xi}, \varphi_{1k}^{m})$. Obviously, this also implies a positive correlation between school ability and the return to experience.

In order to illustrate dynamic self-selection, we have computed expected schooling attainments and expected wages for each type, along with their respective rank. The results are summarized in Table 4B. Within each group, differences in schooling attainments are explained by differences in school ability (υ^{ξ}) and differences in the intercept term of the wage function (υ^{w}) as well as the employment equation (κ_0). Across groups, differences in type specific expected wages and schooling are also explained by differences in the return to schooling and experience and, in particular, by the correlation between school ability and the return to human capital.

Overall, the type specific predicted schooling attainments vary much less than do expected wages. The average predicted schooling attainments for the 3 types endowed with a low return to human capital (Group A) is 12.40 years and is just below the average for group B (12.90 years). This illustrates the fact that those who have higher return to human capital obtain slightly more schooling. It is partly explained by the positive correlation between school ability (υ_1^{ξ}) and the return to human capital. While individuals belonging to group B (type 4, type 5 and type 6) obtain slightly more education than other types, they obtain much higher wages. This may be explained by the fact that those individuals endowed with high school ability are also endowed with high return to schooling as well as high return to experience. As a consequence, the high return to experience counter balances the willingness to invest in school activities.[11]

7 The Local and Average Returns to Schooling: Testing for Population Heterogeneity

At this stage, it is natural to investigate whether differences in the returns to human capital across groups are statistically significant. A formal approach requires to construct a restricted version of the model. The restricted model has 6 types of individuals and, as for the unrestricted model, each type is endowed with a type specific employment, wage and school ability intercept term. However, each type must share the same return to schooling and expe-

[11]In order to separate the effects of schooling and experience, we have simulated differences in schooling attainments across types when either the return to schooling or the return to experience are set to the population average. Not surprisingly, we find a huge positive correlation between individual specific return to schooling and schooling attainments as well as a huge negative correlation between the returns to experience and schooling attainments.

rience. This amounts to imposing 7 restrictions (6 splines for the return to schooling and the effect of experience). Testing homogeneity can be achieved using a likelihood ratio statistic. The estimation of the restricted model lead to a value of -13.7505 for the average log likelihood which in turn, translated into a p-value below 0.01. We conclude that homogeneity is strongly rejected and that a random coefficient specification of the wage regression function is an accurate representation of the importance of population heterogeneity.

In view of the recent literature on estimating average and local treatment effects, in which the estimates of the return to schooling are often interpreted in a random coefficient framework (Heckman and Vytlacil, 1999 and Imbens and Angrist, 1994), we also estimated a version of the model where individual differences in the return to human capital are captured only in the returns to schooling. Overall, this model specification did not perform really well. First, the average log likelihood was found to be -13.7393 (as opposed to 13.6313 for the unrestricted model) and the likelihood ratio tests strongly rejects the homogeneity of the return to experience at the 0.01 level. Second, and perhaps more importantly, the model is incapable of capturing meaningful differences in the average return to schooling. The average return to schooling at college graduation is 0.0620 for Group A and 0.0676 for Group B. The failure of the model can be explained as follows. If differences in the returns to schooling were the only source of comparative advantages (individuals share the same return to experience), the more able would obtain a substantially higher level of schooling than those who are less able. However, such a positive correlation between market ability and schooling attainments is not born by the data and, as a consequence, the likelihood estimates indicate a minimal level of heterogeneity in the returns to schooling.

8 Comparisons Between Various Estimates of the Average Returns to Schooling and Experience

In the reduced- form literature, the return to schooling is typically estimated within a linear regression framework using OLS estimate or IV methods.

Estimating log wage regression functions by OLS will typically require both schooling and experience to be orthogonal to labor market ability. When using IV techniques, it is customary to ignore actual labor market experience and use approximate measures such as age or potential experience. If actual experience is the appropriate proxy for post-schooling human capital investments, using a different measure may introduce a serious mis-specification in the log wage regression model (see Rosenzweig and Wolpin, for a critical review of the literature).[12]

It is therefore informative to compare our estimates with standard OLS estimates obtained from cross-sectional regressions and with the structural estimates obtained under the maintained hypothesis that both the returns to schooling and experience are homogenous. In order to compare the structural estimates with those obtained by OLS, we report the average return to schooling at grade 12 and at grade 16. The structural estimates obtained within a random coefficient framework are in the first column of Table 6 while those estimates obtained from a restricted version of the model (with homogeneous returns) are in column 2. OLS estimates based on the 1990 cross-section are in column 3 and column 4 (OLS with splines).

The structural estimates of the average return to schooling at high school graduation and at college graduation (4.4% and 6.4%) are both much lower than OLS estimates.[13] Within a standard OLS specification, the average (and local) return is around10.0%. When non-linearities are taken into account using splines, the OLS estimates of the average return are 8.8% at high school graduation and 10.5% at college graduation. This is consistent with the fact that OLS estimates may suffer a strong ability bias. However, it should also be noted the estimates obtained from the restricted version of the structural dynamic programming model (in column 2) are lower than the structural (random coefficient) estimates. In the restricted structural model, the average return is 1.2% at high school graduation and 4.3% at college graduation. While there is evidence that OLS regression lead to an over-statement of the true return to schooling, the converse is true about the return to experience. The structural estimates of the return to experience (6.8% with population heterogeneity and 8.2% in the standard model) are

[12]In our sample, the correlation between schooling attainments and actual experience (as of 1990) is equal to -0.5095.

[13]A similar result was obtained in Belzil and Hansen (2000a).

much higher than those obtained by OLS (between 5.1 and 5.2%).

At this stage, it is possible to draw some conclusions. First, point estimates of the returns to schooling and experience are sensitive to the allowance for population heterogeneity in the returns to human capital. This is not surprising. A random coefficient specification offers a completely different way of interpreting dynamic self-selection and, in particular, the correlation between labor market ability and schooling attainments. Despite the differences in point estimates between a random coefficient specification and the more standard approach, there is overwhelming evidence that estimates of the return to schooling obtained from a structural dynamic programming model are lower than OLS estimates as well as other estimates reported in the literature (see Card, 2000). We also note that setting the empirical analysis of the log wage regression function in a random coefficient framework has not changed the overall shape of the log wage regression function. As in Belzil and Hansen (2000), we find much lower returns to high school education than for post- secondary education. After conditioning on a specific type, the log wage regression function remains rather convex in schooling.[14]

Finally, it is clear that allowing for individual differences in the slopes of age-earnings profile will allow us to fit data on wages much better than models based on homogeneous returns to schooling and experience. In order to evaluate the capacity of the random coefficient model to fit data on wages, we have computed the ratio of the variance of explained wages and actual wages for all three model specifications considered. Overall, the random coefficient framework can explain up to 66% of variances in observed wages. this is much higher than what is observed for OLS estimates (24%) and for a structural dynamic programming model where the returns is estimated using 8 splines (Belzil and Hansen, 2000a).

9 Conclusion

In this paper, we have estimated a structural dynamic programming model of schooling decisions and obtain individual specific estimates of the local

[14]The log wage regression function with homogeneous returns however requires as many as 8 splines (Belzil and Hansen, 2000). This is explained by the fact that equality between successive spline segments at grade 11 and grade 12 fails to be rejected when the returns are individual specific.

(and average) returns to schooling as well as the returns to experience. Homogeneity of the returns to human capital is strongly rejected in favor of a discrete distribution version of the random coefficient specification. The results indicate that individuals who have the higher returns to schooling are also those who have higher returns to experience.

The structural estimates of the average return to schooling at high school graduation and at college graduation (4.4% and 6.4%) are both much lower than estimates reported in the literature. Indeed, when averaged over all types, the return to experience in the early phase of the life cycle (6.8%) exceeds the average return to schooling (6.4% at college graduation). After conditioning on ability, the log wage regression function appears rather convex. As those individuals who have comparative advantages in schooling are also those who are faced with higher returns to experience, the model implies weak dynamic self-selection (weak correlation between market ability heterogeneity and schooling attainments) and strong wage dispersion.

As far as we know, the returns to schooling have never been estimated in such a general framework. There are therefore no benchmark result in the literature. Nevertheless, our estimates cast doubts on the validity of the very high returns usually reported in the literature.

References

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

[2] Belzil, Christian and Hansen, Jörgen (2000) "Unobserved Ability and the Return to Schooling" Working Paper, Concordia University.

[3] 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.

[4] Card, David (1998) "The Causal Effect of Education on Earnings" Working Paper, University of California at Berkeley.

[5] Eckstein, Zvi and Kenneth Wolpin (1999) "Youth Employment and Academic Performance in High School" IZA Discussion Paper no. 18, forthcoming in *Econometrica*

[6] Heckman, James and E. Vytlacil (1999) "Local Instrumental variables and Latent Variable Models for Identifying and Bounding Treatment Effects", Working paper, Department of Economics, University of Chicago.

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

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

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

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

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

[12] Rust, John (1987) ”Optimal Replacement of GMC Bus Engines: An Empirical Analysis of Harold Zurcher” *Econometrica, 55 (5)*, 999-1033.

[13] Taber, Christopher (1999) “The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability”, Mimeo, Northwestern University.

Table 1 - Descriptive Statistics

	Mean	St dev.	# 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
# 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
AFQT/10	49.50	28.47	1710
Schooling completed (1990)	12.81	2.58	1710
# 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

Note: 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.

Table 2A
The Utility of Attending School, Subjective Discount Rates and Interruption Probabilities

	Parameter	Std error
Utility in School		
Father's Educ	0.0082	0.0010
Mother's Educ	0.0053	0.0011
Family Income/1000	0.0005	0.0001
Nuclear Family	0.0155	0.0050
Siblings	-0.0061	0.0010
Rural	-0.0001	0.0042
South	-0.0149	0.0044
Stand.Dev.(σ_ξ)	0.1940	0.0105
Educ. Splines		
δ_{7-10}	0.0918	0.0103
δ_{11}	0.4559	0.0234
δ_{12}	-1.3735	0.0248
δ_{13}	0.7497	0.0249
δ_{14}	1.6879	0.0072
δ_{15}	-1.1015	0.0190
δ_{16}	1.1700	0.0476
$\delta_{17-more}$	-0.5857	0.0545
Interruption Prob.	0.0749	0.0036
Discount Rate	0.0111	0.0001
mean log Likelihood	-13.6313	

Table 2B
Individual Specific Intercept Terms and Type Probabilities

			Parameter	St Error	Rank
Type 1	υ_1^ξ	**School ab.**	-2.5433	0.0091	4
	υ_1^w	**Wage**	1.4836	0.0094	4
	κ_{01}	**Employment**	-3.3629	0.0301	4
	q_1^0	**Type Prob.**	-0.6301	0.0419	-
Type 2	υ_2^ξ	**School ab.**	-2.2750	0.0200	2
	υ_2^w	**Wage ab.**	2.0051	0.0192	1
	κ_{02}	**Employment**	-2.3251	0.0189	5
	q_2^0	**Type Prob**	-1.4066	0.0378	-
Type 3	υ_3^ξ	**School ab.**	-3.2156	0.0245	6
	υ_3^w	**Wage**	1.6203	0.0121	3
	κ_{03}	**Employment**	-1.5652	0.0241	6
	q_3^0	**Type Prob**	-0.8961	0.0249	
Type 4	υ_4^ξ	**School ab.**	-2.4926	0.0164	3
	υ_4^w	**Wage**	1.4220	0.0112	5
	κ_{04}	**Employment**	-3.6237	0.0211	2
	q_4^0	**Type Prob**	0.1578	0.0074	-
Type 5	υ_5^ξ	**School ab.**	-2.1681	0.0136	1
	υ_5^w	**Wage**	1.7502	0.0121	2
	κ_{05}	**Employment**	-3.6962	0.0102	1
	q_5^0	**Type Prob**	-0.8046	0.0495	
Type 6	υ_6^ξ	**School ab.**	-2.7820	0.0111	5
	υ_6^w	**Wage**	1.1207	0.0106	6
	κ_{06}	**Employment**	-3.5454	0.0255	4
	q_6^0	**Type Prob**	0.0 (normalized)		

Note: The type probabilities are estimated using a logistic transform. The resulting probabilities are 0.14 (type 1), 0.06 (type 2), 0.11 (type 3), 0.3103 (type 4), 0.12 (type 5) and 0.26 (type 6). The correlation between υ^{ξ} and υ^{w} is 0.4228. The correlation between υ^{ξ} and -κ_0 is 0.59. The correlation between υ^{w} and -κ_0 is 0.45.

Table 3A
The Return to Human Capital

	Parameter (asymptotic st. error)	
Non-Wage		
Schooling	0.0081 (0.0005)	
Employment		
Schooling	-0.0586 (0.0024)	
Experience	-0.0147 (0.0023)	
Experience^2	0.0001 (0.0001)	
Wages		
σ_w^2	0.2906 (0.0302)	
	group A	**group B**
educ. 7-12	0.0048 (0.0012)	0.0614 (0.0018)
educ 13	0.0205 (0.0027)	0.0294 (0.0010)
educ 14	0.0456 (0.0019)	0.0370 (0.0013)
educ 15	-0.0223 (0.0023)	-0.0136 (0.0017)
educ 16	0.0110 (0.0051)	0.0064 (0.0019)
educ 17	-0.0043 (0.0027)	0.0014 (0.0014)
experience	0.0308 (0.0009)	0.0850 (0.0010)
Experience^2	-0.0013 (0.0001)	-0.0013 (0.0001)

Table 3B
The Average and Local Returns to Schooling

	Local Returns		**Average Returns**	
	group A	**group B**	**group A**	**group B**
Grade level				
7-12	0.0048	0.0614	0.0048	0.0614
13	0.0253	0.0908	0.0077	0.0656
14	0.0709	0.1278	0.0156	0.0734
15	0.0486	0.1142	0.0193	0.0779
16	0.0596	0.1206	0.0233	0.0822
17-more	0.0553	0.1210	0.0262	0.0857

Note: Group A is composed of type 1, type 2 and type 3. Group B is composed of type 4, type 5 and type 6.

Table 4A
Some Correlations between School and Market Ability

	Param (p value)
$Corr(\upsilon_i^{\xi}, \upsilon_i^{w})$	0.4321 (0.01)
$Corr(\upsilon_i^{\xi}, -\kappa_{0i})$	0.5939 (0.01)
$Corr(\upsilon_i^{w}, -\kappa_{0i})$	0.4493 (0.01)
$Corr(\upsilon_i^{\xi}, \varphi_{1i}^{m})$	0.2711 (0.01)

Table 4B
Unobserved Heterogeneity, Mean Schooling Attainments and Predicted Wages

	(1) **Mean Schooling**	(2) **Schooling Ranking**	(3) **Mean Wage**	(4) **Wage Ranking**	(5) **Type Probability**
Type 1	13.15	2	\$6.03	6	0.14
Type 2	12.99	3	\$10.05	4	0.06
Type 3	11.30	6	\$6.19	5	0.11
Group A	12.40				0.31
Type 4	12.77	4	\$15.85	2	0.31
Type 5	12.61	5	\$21.76	1	0.12
Type 6	13.19	1	\$12.09	3	0.26
Group B	12.90				0.69

Note: Group A is composed of type 1 ,2 and 3. Group B is composed of types 4, 5 and 6.

Table 5
Model Fit: Actual vs Predicted Schooling Attainments

Grade Level	**Predicted (%)**	**Actual** %
Grade 6	0.0%	0.3 %
Grade 7	0.9%	0.6%
Grade 8	2.4%	2.9%
Grade 9	4.8%	4.7%
Grade10	7.1%	6.0 %
Grade11	7.7%	7.5 %
Grade12	40.1%	39.6 %
Grade13	7.1%	7.0 %
Grade14	7.1%	7.7 %
Grade15	2.0%	2.9 %
Grade16	12.9%	12.9 %
Grade17	2.1%	2.5 %
Grade18	2.5%	2.4%
Grade19	1.8%	1.3%
Grade 20+	1.2%	1.6%

Table 6
Average Return to Schooling and Experience in the Population: Structural Dynamic Programming vs OLS Estimates

	(1) **DP/ML**	(2) **DP/ML**	(3) **OLS**	(4) **OLS/SPLINES**
Specification	random coeff.	Homo. returns	Homo. returns	Homo. returns
Population Average return to schooling				
grade 12	0.0438	0.0122	0.0997	0.0879
grade 16	0.0639	0.0430	0.0997	0.1050
Population Average return to experience				
Experience	0.0682	0.0817	0.0516	0.0514
Experience^2	-0.0013	-0.0027	-0.0014	-0.0012
Model Fit				
$\frac{var.pred.wages}{var.observedwages}$	66 %	25%	24 %	25%

Note: The average returns in column 1 (DP/ML) are obtained from the structural dynamic programming maximum likelihood estimates reported in Table 4B. The estimates obtained from a restricted version of the model (with homogeneous returns) are in column 2. The OLS estimates (in column 3 and column 4) are computed on the cross-section of 1990. The OLS regression with splines (column 4) has the same number of splines as the structural model (column 1) and both OLS regressions contain experience and experience squared.

Liste des publications au CIRANO *

Cahiers CIRANO / *CIRANO Papers* (ISSN 1198-8169)

99c-1 Les Expos, l'OSM, les universités, les hôpitaux : Le coût d'un déficit de 400 000 emplois au Québec — Expos, Montréal Symphony Orchestra, Universities, Hospitals: The Cost of a 400,000-Job Shortfall in Québec / Marcel Boyer

96c-1 Peut-on créer des emplois en réglementant le temps de travail? / Robert Lacroix

95c-2 Anomalies de marché et sélection des titres au Canada / Richard Guay, Jean-François L'Her et Jean-Marc Suret

95c-1 La réglementation incitative / Marcel Boyer

94c-3 L'importance relative des gouvernements : causes, conséquences et organisations alternative / Claude Montmarquette

94c-2 Commercial Bankruptcy and Financial Reorganization in Canada / Jocelyn Martel

94c-1 Faire ou faire faire : La perspective de l'économie des organisations / Michel Patry

Série Scientifique / *Scientific Series* (ISSN 1198-8177)

2001s-09 Return to a High School Diploma and the Decision to Drop Out: New Evidence from Canada / Daniel Parent

2001s-08 Leader and Follower: A Differential Game Model / Hassan Benchekroun et Ngo Van Long

2001s-07 Emission Taxes and Standards for an Asymmetric Oligopoly / Ngo Van Long et Antoine Soubeyran

2001s-06 Risque de modèle de volatilité / Ali Alami et Éric Renault

2001s-05 The Effect of Pay-for-Performance Contracts on Wages /Daniel Parent

2001s-04 Incentive Pay in the United States: Its Determinants and Its Effects / Daniel Parent

2001s-03 Testing and Comparing Value-at-Risk Measures / Peter Christoffersen, Jinyong Hahn et Atsushi Inoue

2001s-02 Empirical Assessment of an Intertemporal Option Pricing Model with Latent Variables / René Garcia, Richard Luger et Éric Renault

2001s-01 Asymmetric Smiles, Leverage Effects and Structural Parameters / René Garcia, Richard Luger et Éric Renault

2000s-60 Technological Paradigms and the Measurement of Innovation / Bernard Sinclair-Desgagné

2000s-59 Taxpayers' Response to Tax Rate Changes: A Canadian Panel Study / Robert Gagné, Jean-François Nadeau et François Vaillancourt

2000s-58 Competition and the Reform of Incentive Schemes in the Regulated Sector / Marcel Boyer et Jean-Jacques Laffont

---

* Vous pouvez consulter la liste complète des publications du CIRANO et les publications elles-mêmes sur notre site Internet à l'adresse suivante :

http://www.cirano.umontreal.ca/publication/documents.html