A Structural Analysis of the Correlated Random Coefficient Wage Regression Model[1]

Christian Belzil
Concordia University, CIRANO, CIREQ and IZA

Jörgen Hansen
Concordia University, IZA and CEPR

February 13, 2004

Abstract

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. We also analyze the determinants of 3 counterfactual experiments (a college attendance subsidy, a high school graduation subsidy and an overall decrease in the rate of time preference) and examine a proposition often claimed in the "Average Treatment Effects" literature; that the discrepancy between OLS and IV estimates of the returns to schooling may be explained by the relatively higher returns experienced by those affected by exogenous policy changes. We find that the average return to experience upon entering the labor market (0.0863) exceeds the average return to schooling (0.0576) and we find more cross-sectional variability in the returns to experience than in the returns to schooling. Labor market skills (as opposed to taste for schooling) appear to be the prime factor explaining schooling attainments. We find little evidence in favor of a positive correlation between reactions induced by an exogenous experiment and the individual specific returns to schooling.

Key Words: *Random Coefficient, Returns to Schooling, Comparative Advantages, Dynamic Programming, Dynamic Self-Selection.*

JEL Classification: J2-J3.

1 Introduction and Objectives

In this paper, we investigate the empirical properties of the correlated random coefficient wage regression model (CRCWRM) using a structural dynamic programming model.[1] The term "correlated random coefficient wage regression model" refers to the standard Mincerian log wage regression function in which the coefficients may be arbitrarily correlated with the regressors (education and experience). While the comparative advantages representation of the labor market is far from being new (Roy, 1951, Becker and Chiswick, 1966 and Willis and Rosen, 1979), economists have only recently paid particular attention to the specification and the estimation of linear wage regression models set in a random coefficient framework (Heckman and Vitlacyl (1998, 2000), Wooldridge (1997, 2000), Angrist and Imbens (1994), Card (2000) and Meghir and Palme (2001)). In this recent branch of the literature, it is customary to estimate the log wage regression function using Instrumental Variable (IV) techniques and interpret the estimates in a framework where the returns to schooling are individual specific. This surge of new research is understandable. In a context where schooling is understood as the outcome of individual decision making within a dynamic framework, rational individuals base their schooling decisions partly on absolute and comparative advantages in the labor market and partly on their taste for schooling. As a consequence, the random coefficients (the returns to schooling and experience), as opposed to only the individual specific intercept terms, will normally be correlated with individual schooling attainments.

In a linear wage regression, individual differences in the intercept term represent a measure of absolute advantage in the labor market while differences in slopes reflect individual comparative advantages in human capital acquisition via schooling and experience. While it might be tempting to focus solely on heterogeneity in the returns to schooling (and assume homogeneous returns to experience), this approach is likely to be unsatisfactory. If the returns to schooling and experience are truly correlated, ignoring individual differences in the return to labor market experience is likely to affect the estimates of the returns to schooling as well as the causal link between labor market ability and schooling (dynamic self-selection). Modeling wage regressions in a random coefficient framework therefore requires the allowance for

---

[1]The term "correlated random coefficient wage regression model" is also used in Heckman and Vitlacyl (1998).

1

heterogeneity in the returns to experience.[2]

As it stands, very little is known about the empirical properties of the CR-CWRM. For instance, those interested in estimating the returns to schooling by IV techniques usually ignore higher moments such as the variance of the returns to schooling and experience, or use a reduced-form framework which cannot disclose the covariances between realized schooling and the individual specific coefficients. However, these quantities are important. They may shed light on the importance of comparative advantages in the labor market and help comprehend the determinants of individual schooling attainments. Finally, they may help quantify the "Ability Bias" (OLS bias) arising in estimating the returns to schooling using regression techniques. Obviously, a random coefficient regression model provides a realistic framework to evaluate the relative importance of labor markets skills and taste for schooling in explaining cross-sectional differences in schooling attainments. Virtually all recent work on empirical earnings functions is directly or indirectly based on a random coefficient framework. For this reason, it deserves some attention.[3]

Our main objective is to investigate the empirical properties of the CR-CWRM. These include the population average returns to schooling and experience, the relative dispersions in the returns to schooling and experience, and the relative importance of labor market skills and individual specific taste for schooling in explaining cross-sectional differences in schooling attainments. We estimate a finite mixture structural dynamic programming model of schooling decisions with 8 unknown types of individuals, where each type is characterized by a specific log wage regression function (linear) as well as a specific utility of attending school.[4] The estimation of a mixed

---

[2]Individual differences in the return to experience may be explained by comparative advantages in on-the-job training, learning on the job, job search or any other type of post-schooling activities enhancing market wages. Allowing for heterogeneity in the returns to experience is especially important if individual post-schooling human capital investments are unobserved (which is the case in most data sets).

[3]Heterogenity in realized returns to schooling may also arise if the local returns change with the level to schooling. In a recent paper, Belzil and Hansen (2002a) used a structural dynamic programming model to obtain flexible estimates of the wage regression function from the National Longitudinal Survey of Youth (NLSY). They found that the log wage regression is highly convex and found returns to schooling much lower than what is usually reported in the existing literature although the local returns may fluctuate between 1% (or less) and 13% per year.

[4]In this paper, we disregard using available measures like AFQT scores. We do so for two main reasons. First, AFQT scores may often reflect differences in schooling (as

likelihood function has two main advantages. It can capture any arbitrary correlation between any of the heterogeneity components and it obviates the need to incorporate all parents' background variables in each single heterogeneity component or to select, somewhat arbitrarily, which heterogeneity components are correlated with household background variables and which ones are not.

A second objective is to illustrate the importance of population heterogeneity and, more specifically, to analyze the characteristics of the subpopulation (s) most affected by a counterfactual experiment. This is an important issue. In the literature, estimates of the returns to schooling obtained using instrumental variable techniques are often higher than OLS estimates.[5] It is often postulated that these results are explained by the fact that those individuals more likely to react to an exogenous policy change must have higher returns to schooling than average. As far as we know, this claim has neither been proved nor verified empirically in any direct fashion. To illustrate the importance of heterogeneity in the reactions to treatment, we conducted 3 counterfactual experiments. We first increased the level of the utility of attending school so to mimic the effect of a college subsidy. As this experiment is targeted at specific school levels, we also simulated a decrease in the rate of time preference. This experiment is more likely to affect school attendance at all levels, although it is perhaps more difficult to associate an empirical counterpart to it.[6] Finally, we simulated a high school graduation subsidy which, for instance, might be targeted at reducing high school drop out behavior. In all cases, the experiment represents a "truly exogenous" change which may serve as a basis for a natural experiment.

A third and final objective is to investigate the notion of "Ability Bias" in a context where the notion is much deeper than the usual correlation between

tests are not taken at the same age by all young individuals). Second, and perhaps more importantly, we are interested in developing a methodology which may be applicable to a wide range of panel data sets. As far as we know, most panel data sets of labor market histories do not contain observed measures of skill heterogeneity such as AFQT scores and the like.

[5] The validity of very high returns to schooling, reported in a simple regression framework have been seriously questioned (see Manski and Pepper, 2000 and Belzil and Hansen, 2002a). It is also interesting to note that empirical evidence also suggests that standard wage regressions augmented with observable measures of ability (such as test scores and the like) lead to a decrease in the estimated returns to schooling.

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

the individual specific intercept terms of the wage regression and realized schooling attainments. As market ability heterogeneity is multi-dimensional in our model, our estimate of the Ability Bias (OLS bias) is not only explained by the correlation between the individual specific intercept term and realized schooling but also by the simultaneous correlations between schooling and experience and the individual specific deviations from population average returns to schooling and experience.

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. We find population average returns to schooling which are much below those reported in the existing literature. Our estimates are also much lower than those obtained using standard OLS techniques. The average return to experience upon entering the labor market (0.0863) exceeds the average return to schooling (0.0576) and we find more cross-sectional variability in the returns to experience than in the returns to schooling. The returns to schooling and experience are found to be positively correlated. Not surprisingly, the correlated random coefficient wage regression model fits wage data very well. It can explain as much as 78.5% of the variation in realized wages. Overall, the dynamic programming model indicates that labor market skills are the prime factor explaining schooling attainments as 82% of the explained variation is indeed explained by individual comparative and absolute advantages in the labor market and only 18% is explained by individual differences in taste for schooling. Moreover, realized schooling attainments are more strongly correlated with individual differences in returns to experience than in returns to schooling.

The importance of individual heterogeneity in the level of reactions to policy changes is well illustrated by the 3 counterfactual experiments that we conducted. In all cases (the college attendance subsidy, the high school graduation subsidy and the overall decrease in the rate of time preference), the reactions induced by the artificial instrument are a complicated (non-linear) function of individual specific skills and, moreover, the determinants of the reactions will differ according to the nature of the experiment. More specifically, it appears that the determinants of the reactions are very much affected by whether or not the experiment is targeted at a particular school level. However, homogeneity is strongly rejected. Unlike what is often claimed in the average treatment effects literature, the discrepancy between IV and OLS is not necessarily explained by a positive correlation between individual

4

specific schooling reactions and the returns to schooling. While it is possible to reconcile our results with conventional wisdom about IV estimates of the returns to schooling in one of our experiments (the general decrease in the discount rate), a positive correlation between individual specific reactions and the returns to schooling is not a "general consequence" of any exogenous change in the determinant of school attendance.

The paper is structured as follows. The empirical dynamic programming model is exposed in Section 2. The goodness of fit is evaluated in Section 3. A discussion of the estimates of the return to schooling and experience are found in Section 4. In Section 5, we illustrate the links between labor market skills and dynamic self-selection. In Section 6, we analyze the determinants of the individual specific reactions to 3 counterfactual experiments and examine a proposition often claimed in the "Average Treatment Effects" literature; that the discrepancy between OLS and IV estimates of the returns to schooling may be explained by the relatively higher returns experienced by those affected by exogenous policy changes. In Section 7, we discuss the links between our estimates and those reported in the literature and re-examine the notion of Ability Bias in a context where the regression function allows for a rich specification of absolute and comparative advantages. The conclusion is in Section 8.

2 An Empirical Dynamic Programming Model of Schooling Decisions with Comparative Advantages

In this section, we introduce the empirical dynamic programming model. While the theoretical structure of the problem solved by a specific agent is similar to the model found in Belzil and Hansen (2002a), the different stochastic specification and, especially, the allowance for a rich specification of absolute and comparative advantages requires a full presentation.

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 The Utility of Attending School

The instantaneous utility of attending school, $U^s(.)$, is formulated as the following equation[7]

$$U^s(.) = \psi(S_{it}) + \upsilon_i^\xi + \varepsilon_{it}^\xi \tag{1}$$

in which $\varepsilon_{it}^\xi \sim i.i.d \ N(0, \sigma_\xi^2)$ represents a stochastic utility shock, the term υ_i^ξ represents individual heterogeneity (ability) affecting the utility of attending school and $\psi(.)$ captures the co-movement between the utility of attending school and grade level.

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}.[8] When $I_{it} = 1$, the decision problem is frozen for one period. If $I_{it} = 0$, the decision can be made. When an interruption occurs, the stock of human capital remains constant over the period.[9]

2.2 The Utility of Work

Once the individual has entered the labor market, he receives monetary income $\tilde{w}_t$, which is the product of the yearly employment rate, e_t, and the wage rate, w_t. The instantaneous utility of work, $U^w(.)$

$$U^w(.) = \log(\tilde{w}_t) = \log(e_t \cdot w_t)$$

---

[7]The utiliy of school could be interpreted as the monetary equivalent (on a per hour basis) of attending school.

[8]The interruption state is meant to capture events such as illness, injury, travel, temporary work, incarceration or academic failure.

[9]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. In the NLSY, we find that more than 85% of the sample has never experienced school interruption.

2.3 The Correlated Random Coefficient Wage Regression Model

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

$$\log w_{it} = \varphi_{1i} \cdot S_{it} + \lambda_i \cdot (\varphi_2 \cdot Exper_{it} + \varphi_3 \cdot Exper_{it}^2) + v_i^w + \varepsilon_{it}^w \qquad (2)$$

where φ_{1i} is the individual specific wage return to schooling and λ_i is an individual specific factor multiplying the effect of experience (φ_2) and the effect of experience squared (φ_3). The term v_i^w represents an individual specific intercept term. We assume that

$$\varphi_{1i} = \bar{\varphi}_1 + \omega_{1i}$$

$$\lambda_i = \bar{\lambda} + \omega_{2i}$$

where $\bar{\varphi}_1$ and $\bar{\lambda}$ represent population averages. Following the convention used in the literature, it is convenient to specify the wage regression as a heteroskedastic regression function

$$\log w_{it} = \bar{\varphi}_1 \cdot S_{it} + \bar{\varphi}_2^* \cdot \cdot Exper_{it} + \bar{\varphi}_3^* \cdot \cdot Exper_{it}^2 + \varrho_{it} \qquad (3)$$

where

$$\bar{\varphi}_2^* = \bar{\lambda} \cdot \varphi_2$$

$$\bar{\varphi}_3^* = \bar{\lambda} \cdot \varphi_3$$

$$\varrho_{it} = v_i^w + \omega_{1i} \cdot S_{it} + \omega_{2i} \cdot (\varphi_2 \cdot Exper_{it} + \varphi_3 \cdot \cdot Exper_{it}^2) + \varepsilon_{it}^w$$

Estimating the population average returns to schooling and experience ($\bar{\varphi}_1$, $\bar{\varphi}_2^*$ and $\bar{\varphi}_3^*$) is rendered difficult by the fact that typically

$$Corr(\ \varrho_{it}, S_{it}) \neq 0$$

$$Corr(\ \varrho_{it}, Exper_{it}) \neq 0$$

7

2.4   The Employment Rate

The employment rate, e_{it}, is also allowed to depend on accumulated human capital (S_{it} and $Exper_{it}$) so that

$$\ln e_{it}^* = \ln \frac{1}{e_{it}} = \upsilon_i^e + \kappa_1 \cdot S_{it} + \kappa_2 \cdot Exper_{it} + \kappa_3 \cdot Exper_{it}^2 + \varepsilon_{it}^e \qquad (4)$$

where υ_i^e 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.[10] 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.

2.5   The Value Functions

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 value function associated with the decision to remain in school, given accumulated schooling S_t, denoted $V_t^s(S_t, \eta_t)$, can be expressed as

$$V_t^s(S_t, \eta_t) = \ln(\xi_t) + \beta\{\zeta \cdot EV_{t+1}^I(S_{t+1}, \eta_{t+1}) \qquad (5)$$

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

where $V_{t+1}^I(S_{t+1}, \eta_{t+1})$ denotes the value of interrupting schooling acquisition. Since 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. $V_{t+1}^I(S_{t+1}, \eta_{t+1})$, can be expressed as follows.

$$V_{t+1}^I(S_{t+1}, \eta_{t+1}) = \log(\xi_{t+1}) + \beta\{\zeta \cdot EV_{t+2}^I(S_{t+2}, \eta_{t+2})$$

$$+(1 - \zeta) \cdot EMax[V_{t+2}^s(S_{t+2}, \eta_{t+2}), V_{t+2}^w(S_{t+2}, \eta_{t+2})]\} \qquad (6)$$

---

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

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

$$V_t^w(S_t, \eta_t) = \ln(w_{it} \cdot e_{it}) + \beta E(V_{t+1} \mid d_t = 0) \tag{7}$$

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

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

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 choices can readily be expressed in closed form.

2.6 Unobserved Ability in School and in the Labor Market

We assume that there are K types of individuals. Each type (k) is endowed with a vector $(v_k^\xi, v_k^w, v_k^e, \varphi_{1k}, \lambda_k)$ for $k = 1, 2...K$. The results reported in this paper are for the case $K = 8$. The probability of belonging to type k, p_k, is estimated using logistic transform

$$p_k = \frac{\exp(q_k)}{\sum_{j=1}^{8} \exp(q_j)}$$

and with the restriction that $q_8 = 0$.[11]

2.7 Identification

As it is the case in all empirical dynamic programming models, identification of the key components of the model will require distributional assumption.

---

[11] As discussed in Belzil and Hansen (2002a), identification of most parameters is relatively straightforward. However, in order to reduce the degree of under-identification (non-parametric), we fixed the discount rate to 3% per year (an estimate practically identical to the estimate found in Belzil and Hansen (2002a). The degree of under-identification arising in estimating structural dynamic programming models is discussed in details in Rust (1994) and Magnac and Thesmar (2001).

9

Our model being structured as a single choice dynamic model, data on both wages and schooling attainments allow us to identify the key parameters the utility of attending school (like in Keane and Wolpin, 1997, Eckstein and Wolpin, 1999 and Belzil and Hansen (2002a). 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 given a distributional assumption about the stochastic shock. In the present case, where the wage regression function may be expressed as a heteroskedastic regression function, identification of the individual specific slopes is facilitated by assuming homoskedasticity of the stochastic shocks. However, in order to reduce the degree of under-identification (non-parametric), we fixed the discount rate to 3% per year (an estimate practically identical to the estimate found in Belzil and Hansen (2002a).[12]

2.8 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 (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 densities and employment 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

$$\log L = \log \sum_{k=1}^{8} p_k \cdot L_k \qquad (8)$$

where each p_k represents the population proportion of type k.

---

[12]The degree of under-identification arising in estimating structural dynamic programming models is discussed in details in Rust (1994) and Magnac and Thesmar (2001).

3   Accuracy of Predicted Schooling and Predicted Wages

Evidence presented in Table 1A shows clearly that the model is capable of fitting the data quite well. A comparison between actual and predicted frequencies reveals that, except for the very low levels of schooling, our model predicts a pattern which is practically identical to the one found in the data. In particular, we are able to predict the large frequencies at 12 years and 16 years. The fit is comparable to what is found in Belzil and Hansen (2001a and 2001b), in which data on household background (parents' education and income, number of siblings and the like) are used explicitly in the utility of attending school as well as in the wage regression.

Using the structural estimates, it is easy to compute a type specific expected schooling attainments. These are reported in Table 1B. The type specific attainments range from 9.4 years (type 4) to 13.7 years (type 3). An in-depth analysis of the links between schooling and individual specific absolute and comparative advantages is delayed to Section 5.

It is also straightforward to use the simulated values of schooling and experience to simulate series of realized lifetime wages. These series can be used to infer the fraction of the variance of realized wages which is explained by the individual specific regression functions. To investigate the goodness of fit, we have simulated wages for a cohort of individuals aged 30 in 1990[13]. The results reported in Table 1C indicate that random coefficient model explains 78.9% of the observed (realized) variation in wages. This is much larger than what is usually reported in the literature, in which standard OLS regressions of wages on schooling and experience typically result in values of R^2 ranging between 0.20 and 0.25 (see Card, 2000).

4   Absolute and Comparative Advantages in the Labor Market: Some Estimates

In this section, we discuss of the estimates of the return to schooling and experience. The model has been fit on a sample of white males taken from the NLSY and we used the exact same sample as in Belzil and Hansen (2002a).

---

[13]We have also simulated wages under various other scenarios and obtained similar results.

More details are found in Appendix. Note that the estimation of a finite mixture dynamic programming model not only allows us to estimate the population average returns to schooling and experience but also the cross-sectional variability in the returns. This is a relatively novel feature. As far as we know, no one has ever been able to obtain estimates of the variances of the returns to schooling and experience in a framework where both are endogenous.[14]

The individual specific estimates of the wage regression function (the returns to schooling and experience as well as the individual specific intercept terms measuring absolute advantages in the labor market) are found in Table 2A. Our estimates of the returns to schooling range from 0.0265 (type 7) to 0.0879 (type 2) while our estimates of the individual specific $\lambda's$ (ranging from 0.1453 to 1.0866) imply that the returns to experience upon entrance in the labor market range from 0.0197 (type 6) to 0.1477 (type 5). Given the estimates for φ_2 (0.1359) and φ_3 (-0.0040), the ordering based on the $\lambda's$ is identical to the ordering based on the product of λ and φ_2 for the most part of the life cycle and, especially, for the early post-schooling period. As an illustration, the individual specific returns to experience measured after 8 years of experience (a level higher than the average level of experience measured in 1990) are 0.0719 (type 1), 0.0222 (type 2), 0.0141 (type 3), 0.0191 (type 4), 0.0781 (type 5), 0.0105 (type 6), 0.0690 (type 7) and 0.0246 (type 8).

Overall, and as reported in Belzil and Hansen (2002a), our estimates of the return to schooling are much lower than those reported in the existing literature.[15] The population average return to schooling (0.0575) is smaller than the population average return to experience upon entrance in the labor market (0.0863). Interestingly, the high degree of dispersion in λ implies a higher standard deviation in the returns to experience (0.0527) than in the returns to education (0.0218). Upon reviewing the estimated $\lambda's$ and the φ'_1s, it is also noticeable, although not surprising, that the returns to schooling and experience are positively correlated. The correlation between φ_{1i} and λ_i

---

[14]Very often, those who focus on the return to schooling use a proxy variable for experience. Rosenzweig and Wolpin (2000) present a critical analysis of empirical work devoted to the estimation of the returns to schooling, which ignores post-schooling human capital investment.

[15]However, in Belzil and Hansen (2002a), the wage regression function is estimated flexibly using spline techniques. There are 8 different local returns which range 0.4% per year to 12.0% per year.

is around 0.11 and is discussed in more details below (Section 6). It may be explained by the fact that labor market skills which enhance wage growth (job training, job search, etc..) are positively correlated with academic skills which are rewarded in the labor market. This result has clear impact for the nature of dynamic self-selection. Those endowed with high returns to education will not necessarily obtain a high level of schooling because they will be facing a higher opportunity cost of attending school.

While it is difficult to evaluate the relative degree of heterogeneity in taste for schooling and in the returns to human capital without performing simulations, it is nevertheless informative to examine the estimates of the intercept terms of the utility of attending school (reported in Table 2B). Clearly, individual differences in the intercept terms of the taste for schooling appear as important as differences in the intercept terms of the wage equation. The intercept terms for the utility of attending school range from -1.7791 (type 2) to -0.6397 (type 7). Interestingly, even after allowing for 8 types, the high degree of variability (as well as the significance level) of the spline estimates shows that the utility of attending school undoubtedly varies with school level.

Table 3 summarizes the type specific rankings according to all heterogeneity dimensions as well as the level of expected schooling. In an empirical model characterized by a rich specification for skill heterogeneity, the self-selection process is intricate. Individuals take optimal schooling decisions based on their individual specific taste for schooling and their absolute and comparative advantages in the labor market. While some individuals are endowed with a high taste for schooling (as can be seen from Table 2B), schooling decisions are largely affected by the combination of comparative advantages (returns to schooling and experience) and absolute advantages (intercept terms of the wage regression). As a consequence, it will be impossible to associate a definite set of attributes (say, high or low return to human capital) to each specific type on the basis of their sole expected schooling attainments. Nevertheless, our model is sufficiently rich to capture differences in comparative advantages among types of individuals that might obtain similar levels of schooling.

To illustrate this, consider the set of individuals (type 1, type 2, type 4 and type 7) who are predicted to obtain a relatively lower level of schooling than the rest of the population. Type 7 individuals obtain a low level of schooling because they have a low return to schooling and a high return to experience, despite a very high taste for schooling. At the same time, type 2

13

individuals, who also obtain a low level of schooling, are endowed with high return to schooling and experience. However, these individuals are endowed with a very high wage intercept (high market ability) and a low utility of attending school.

The mechanics of the model can also be illustrated at the higher end of the schooling spectrum. Both type 3 and type 8 individuals are predicted to attain a high level of schooling (13.7 years and 12.6 respectively). While both types face relatively similar returns to schooling and experience, they differ substantially in terms of the utility of attending school and the wage intercept. Basically, type 8 individuals choose a high level of schooling because they have a high utility of attending school and type 3 individuals choose a high level of schooling because of a very low level of market ability (wage intercept). A more formal analysis of the link between individual specific heterogeneity (comparative advantages) and schooling attainments is performed in the next section.

At this stage, it is informative to examine the estimated correlations between the returns to schooling and other heterogeneity components (taste for schooling, returns to experience and wage intercept). In a standard regression framework where market skill heterogeneity is only intercept based, a positive Ability Bias is easily explained. It arises because the wage intercept term is simultaneously (and positively) correlated with taste for schooling and schooling attainments. However, in the model analyzed therein, self-selection is more complex. The correlation patterns displayed in Table 4 indicate that those who have a high return to schooling also tend to have a high return to experience although the measured correlation (0.1030) is relatively weak. The correlation between the wage intercept and the returns to schooling is also positive (0.2553). This positive correlation indicates that those who tend to have higher wages will also tend to have comparative advantages in schooling and therefore conforms to standard intuition. Interestingly, taste for schooling is found to be positively correlated with the returns to experience (0.2882) but not with the returns to schooling. The link between these correlations and the treatment effects of an increase in the utility of attending school will be discussed later.

5   Explaining Individual Schooling Attainments: Absolute and Comparative Advantages in the Labor Market vs Taste for Schooling

To investigate formally the determinants of individual schooling attainments implied by our estimates, we simulated our model and generated 200,000 observations on schooling attainments. Using standard regression techniques, we estimated the effects of each individual specific components (taste for schooling, wage intercept, return to schooling and return to experience) on schooling attainments. As the exact form of the relationship between realized schooling and the determinants of the model is unknown, we searched for the best specification. We started by including all elements and their squared terms, and gradually removed all those that were found insignificant. We also experimented with log schooling as well as schooling. The resulting regressions are found in Table 5A. To get a better picture of the correlation between schooling and the returns implied by the model, we have plot the returns (average) against each possible schooling levels (Table 5B).

As expected, individual schooling attainments increase with individual specific returns to schooling and taste for schooling but decrease with respect to the wage intercept and the return to experience. In total, individual differences in labor market skills and taste for schooling explain 35% of the total cross-sectional variation in schooling. The remaining 65% is explained by pure random wage and utility shocks. When taste for schooling is excluded from the regression (column 2), labor market skills explain 28% of the total variation in schooling attainments. This is interesting. It means that 82% of the explained variations in schooling attainments are explained by labor market skill endowments and only 18% by individual differences in taste for schooling.

While this does not necessarily contradict results recently reported in the literature, it nevertheless offers a different way of characterizing schooling attainments. For instance, Keane and Wolpin (1997), Eckstein and Wolpin (1999) and Belzil and Hansen (2002a and 2002b) all find that individual schooling attainments are largely explained by differences in individual taste for schooling. These differences are either caused by individual abilities or household human capital. However, in all of these papers, individual differ-

ences in labor market skills are captured in the intercept term of the wage function. The large effects attributed to differences in the utility of attending school may therefore be explained by the restricted level of heterogeneity in labor market skills.[16]

6 Skill Heterogeneity and Treatment Effects: An Analysis of 3 counterfactual Experiments

The importance of type specific endowments can also be used to learn about the individual specific reactions to some "exogenous policy change". In practice, any "natural experiment" creating changes in the incentives to attend school must effect at least one of the basic components of the decision making. Therefore, the effect of any exogenous policy change will depend on the fundamental parameters (v_k^ξ, φ_{1k} ,v_k^w, λ_k and β).

In the literature concerned with standard linear IV estimation, it is implicitly assumed that the induced response is linear in the instrument and that unobserved heterogeneity does not affect treatment. Any departure from these assumptions may have severe consequences. Indeed, the weakness of the IV approach is widely recognized in the treatment effect literature (Card, 2001 and Heckman and Vitlacyl, 1998). As of now, obtaining consistent estimates of the population average parameters, in a heterogenous treatment effect framework, is still a major challenge.[17] Those interested in estimating average treatment effects using standard IV techniques, often claim that their estimates (which are only valid for a sub-population) are higher than OLS estimates simply because the average returns of the sub-population affected by the instrument are higher than the population average returns (Card, 2000). As far as we know, neither the veracity nor the universality of this claim have been proved nor verified empirically. At this stage, we may only conjecture that the patterns of individual specific treatment reactions are

---

[16]However, in Keane and Wolpin (1997), the return to schooling varies across broadly defined occupation types. In Belzil and Hansen (2002b), both the utility of attending school and labor market ability are function of household background variables. The authors decompose schooling attainments into 2 orthogonal sources, parents' human capital and residual school and market abilities. They find that parents' human capital variables are more important than residual ability.

[17]See Florens, Heckman, Meghir and Vytlacil (2002).

likely to change with the nature of the experiment.[18]

To illustrate the importance of heterogeneity, we actually conducted 3 counterfactual experiments. We first increased the level of the utility of attending school so to mimic the effect of a college subsidy. We chose a level of subsidy corresponding to an average of 25% of the per-period utility of attending school so to allow for meaningful changes in schooling.[19] As our objective is only to investigate the level of heterogeneity in the response to treatment, as opposed to obtain a realistic estimate of a college subsidy, the level of the experiment is immaterial. As this experiment is targeted at specific school levels, we also simulated a decrease in the rate of time preference (from 0.05 to 0.04). This experiment may be more likely to affect school attendance at all levels, although it is perhaps more difficult to associate an empirical counterpart to it.[20] Finally, we simulated a high school graduation subsidy which, for instance, might be targeted at reducing high school drop out behavior. Nevertheless, in both cases, the experiment represents a "truly exogenous" change which may serve as a basis for a natural experiment.[21]

6.1 Treatment Effects of a Counterfactual College Attendance Subsidy

The type specific changes in mean schooling following the college attendance subsidy are found in Table 6A. In order to ease the interpretation, we also included the type specific mean schooling attainments (from Table 1B) and the returns to schooling (from Table 2A). The results point out to a substantial level of heterogeneity in the reactions to treatment. The average increase is around 2.7 years and the standard deviation is around 2.2. This relatively high average level of reaction is a consequence of the high level of subsidy which we chose. It is informative to examine the mean schooling

---

[18]Rosenzweig and Wolpin (2000) discuss a similar argument using a simple theoretical model of the probability of continuing in high school (the Angrist-Krueger model).

[19]This subsidy boils down to increase the utility by 0.15.

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

[21]As our model is set in a partial equilibrium framework, this simulations ignore the potential general equilibrium effects of this policy change and may well exaggerate the effects of treatment. However, as our objective is to examine how various types react to an identical change, the relative reactions are most likely unaffected by the magnitude of the treatment.

attainments and the returns to schooling of those types that are reacting the most to treatment. Our finding indicate that the levels of reaction of type 8, type 5 and type 2 individuals are well above the population average. In particular, type 8 individuals (the most affected group) turns out to be composed of individuals who are just at the margin of deciding to enter post high school education (12.6 years on average).[22] Interestingly, type 8 individuals are not endowed with a particularly high returns to schooling. Indeed, their return (0.0400) is even below the population average.

To investigate the determinants of the individual specific reactions and, in particular, the role of the returns to schooling and experience, we computed OLS regressions of the change in schooling on all measures of skill heterogeneity. The specification allows for potential non-linearities of the effects of the returns to schooling and experience. The regressions are in Table 6D (columns 1 and 2). The results reported in column 1 (when all heterogeneity components are included) illustrate clearly that the effectiveness of a college subsidy will vary across types. The counterfactual change in years of schooling decrease with the instantaneous utility of attending school (v_k^ξ), with the level of the wage intercept (v_k^w) and with the individual specific returns to schooling (φ_{1k}) but increase and the returns to experience (λ_k) over the relevant range as the $\lambda's$ vary between 0.20 and 1.08. Obviously, as indicated by the standard errors and the level of R^2, a test for the nullity of all slope parameters (a test for homogeneity of treatment reaction) would be strongly rejected at any reasonable confidence level.

However, it should be noted that the estimates of Table 6D must be interpreted as the marginal effects of each particular heterogeneity component holding other components constant and, as such, are not indicative of the unconditional correlation between the returns to schooling and the individual specific reactions to a policy change. Indeed, the results found in column 2 indicate a negative correlation between treatment reaction and the returns to schooling even if we allow for non-linearities (the estimates are -0.0290 and -0.0070). This result does not support the hypothesis that the returns to schooling of those who are more likely to react to an exogenous change in the utility of attending school are overwhelmingly superior to the population average. On the contrary, the returns should be lower.

---

[22]Overall, there is a negative correlation between schooling and treatment reaction (-0.43)

18

6.2 Treatment Effects of an Exogenous Decrease in the Discount Rate

As argued before, a counterfactual decrease in the discount rate (from 0.05 to 0.04) will, by construction, not be targeted at a specific school level (say college attendance). For this reason, it may be argued that reaction to treatment is less likely to vary across types and more likely to obey the linearity assumption. We can now examine this claim. First, it may be noted from Table 3C that treatment reaction rankings have changed in a non-trivial way. For instance, type 5 individuals are now those who are reacting the most to this counterfactual experiment and they are also endowed with much higher returns to schooling (0.0764) than the population average. As well, the regression found in columns 3 of Table 6D indicates that the signs of most of the parameters associated to the determinant of treatment have changed. That is the reaction to treatment increases with the instantaneous utility of attending school (v_k^ξ) and with the individual specific returns to schooling (φ_{1k}) but decreases with the level of the wage intercept (v_k^w) and the returns to experience (λ_k) over the relevant range. However, as in the college attendance subsidy, homogeneity of treatment reaction across types is still strongly rejected although the lower value of R^2 indicate that the heterogeneity level is somewhat less important.

In this experiment, those reacting the most (type 5) appear to have high returns to education so there is hope that the results are compatible with usual claim found in the literature (Card, 2000). Indeed, we find that the correlation between the returns to schooling and the reactions to treatment is U shaped when non-linearities are considered. In particular, the parameters found in column 4 (-0.44 and 0.04) indicate that at levels of the returns to education above 5.5% per year (virtually the population average return), the positive term will dominate and standard IV estimates will be likely to exceed the population average. In this particular experiment, it is therefore possible to reconcile, at least partially, our results with the common explanation advanced for the incidence of very high IV estimates of the returns to schooling.

6.3 Treatment Effects of a High School Graduation Subsidy

As a final experiment, we simulated the effects of a 25% increase in the per-period utility of graduating in high school (grade 12). As an example, such an increase could follow from a high school graduation subsidy targeted at reducing the number of high school drop-outs. As in all previous experiments, we present the type specific changes in mean schooling following the high school graduation subsidy (Table 6C) as well as the determinants of the individual specific treatment reactions (table 6D, columns 5 and 6).

From Table 6C, it may be inferred that type 8 and type 5 individuals are those who will react the most to this policy. Overall, there is a strong resemblance between the high school graduation subsidy and the college attendance subsidy. While type 5 is composed of high-school drop outs just at the margin of deciding to continue until graduation (their expected level of schooling is 11.5), those types endowed with particularly low level of schooling (type 4 and type 2) do not seem particularly affected by this counterfactual policy. It is also interesting to note that the unconditional relationship between treatment reactions and the returns to schooling has now an inverse U-shape (Table 6D, column 6). Given the parameter values (0.4069 and -0.0496), the negative correlation will however dominate over the range of returns to schooling which exceeds 4% per year. As in the college attendance subsidy, the results of the high school graduation subsidy experiment would not really support the hypothesis that the returns to schooling of those who are more likely to react to are overwhelmingly superior to the population average.

At this stage, it is possible to draw the following 2 conclusions. First, and on a general note, the reactions to policy changes are a complicated non-linear function of individual specific attributes. It would be therefore somewhat unnatural to model treatment in a simple linear regression framework unless there is compelling evidence. As pointed out by Rosenzweig and Wolpin (2000), the access to data characterized by a "natural experiment" will not obviate the need to model individual specific reactions to treatment. This is exactly what our estimates and simulations indicate.

Second, and more specifically, a positive correlation between returns to schooling and the individual specific changes in schooling resulting from a policy experiment is far from a "universal" consequence from any exogenous change in the determinants of school attendance. For instance, we find no evidence that IV estimates of the returns to schooling arising from the use of

20

an exogenous subsidy to college attendance as an instrument should lead to particular high returns to schooling (compared to the population average). To the contrary, they should be lower. A practically similar conclusion may be drawn from a high school graduation subsidy.

7   The Correlated Random Coefficient Wage Regression Model and the Ability Bias

In the existing literature, it is customary to investigate the ability bias indirectly by comparing OLS and IV estimates of the return to schooling. Here, the orthogonality of the cross-sectional error term in the CRCWRM may be investigated directly using simulations. Furthermore, in a context where market ability heterogeneity is multi-dimensional, the notion of ability bias is much deeper than the usual correlation between individual specific intercept terms of the wage regression and realized schooling attainments. Clearly, the asymptotic OLS bias may be expressed as

$$As.\ bias = \text{plim}(\hat{\beta}_{ols} - \beta) = \text{plim}(\frac{W'W}{N})^{-1} \cdot \text{plim}\frac{W'\xi}{N}$$

where

- $\beta = (\bar{\varphi}_1, \bar{\varphi}_{2\cdot}^*, \bar{\varphi}_3^*)'$

- $W = [S_t, Exper_t, Exper_t^2]$

- N=sample size

- $\varrho = v^w + \omega_1' \cdot S_t + \omega_2' \cdot (\varphi_2 \cdot Exper_t + \varphi_3. \cdot Exper_t^2) + \varepsilon_t^w$.

Note that W is a $Nx3$ matrix of endogenous variables measured at t and that the terms $S_t, Exper_t, Exper_t^2, \varrho, \omega_1, \omega_2, v^w$ are all $Nx1$ vectors. Obviously, the asymptotic bias will only be equal to 0 if $\text{plim}\frac{W'\xi}{N}$=0. Furthermore, given that the vector of individual specific error terms ξ is not centered at 0 and that $W'W$ is not, in general, a diagonal matrix, it is impossible to express the asymptotic bias in terms of a simple correlation (as in Card, 2000). The components of the vector $\text{plim}\frac{W'\xi}{N}$ as well as the resulting bias may easily be computed using the sample created in Section 5. The estimates (along

with their p-values) are found in Table 7A. In Table 7B, we also report the correlation matrix of W.

There is clear evidence that accumulated human capital W is not orthogonal to the error term ξ (Table 7A) and that the degree of non-orthogonality between the vectors of W is important (Table 7B).[23] The product of the probability limit of the inverse of the moment matrix and the probability limit of $\frac{W'\xi}{N}$ imply that OLS estimates will over-estimate the return to education and the effect of experience^2 and under-estimate the returns to experience. This may be verified easily by estimating the wage regression by OLS using various cross-sections of the NLSY or applying OLS on the entire panel. Obviously, the OLS estimates for education, experience and experience^2 will fluctuate according to the specific cross-section (year) chosen. To summarize, for the largest cross-sections in the NLSY (88, 89 and 90), the OLS estimate for education will typically fluctuate between 8.5% and 10% per year while the returns to experience will be between 3% and 6% per year.

This illustrates another possible explanation for the difficulties encountered by those interested in estimating these parameters by IV. In absence of data on entry wages, estimates based on regressions that ignore the endogeneity of accumulated experience may suffer serious mis-specification.

8   Conclusion

We have investigated some of the most interesting properties of the correlated random coefficient wage regression model using a structural dynamic programming model. In our model, individuals make schooling decisions according to their individual specific taste for schooling as well as their individual specific labor market skills and heterogeneity in the realized returns to schooling is interpreted as pure cross-sectional heterogeneity.

We find that the average return to experience upon entrance in the labor market (0.0863) exceeds the average return to schooling (0.0576) and we find more variability in the returns to experience than in the returns to schooling. The returns to schooling and experience are found to be positively correlated. Not surprisingly, the correlated random coefficient wage regression model fits

---

[23]Rosenzweig and Wolpin (2000) discuss the non-orthogonality between accumulated experience and ability which may arise when individuals keep optimizing (by choosing the optimal number of hours of work) after having entered the labor market.

wage data very well. It can explain as much as 78.5% of the variation in realized wages.

Interestingly, labor market skills appear to be the prime factor explaining schooling attainments as 82% of the explained variations are indeed explained by individual comparative and absolute advantages in the labor market while 18% only are explained by differences in taste for schooling. Moreover, realized schooling attainments are more strongly correlated with individual differences in returns to experience than in returns to schooling.

Our results indicate that the reactions to counterfactual policy changes differ substantially across types and that, as a consequence, the usage of natural experiments will not obviate the need for modeling reaction to treatment. Furthermore, the characterization of the individual specific reactions differ substantially according to the nature of the experiment. For instance, a college attendance subsidy and a experiment which would result in a decrease in the discount rate would have very different consequences, although in both cases, homogeneity of the reaction is rejected. However, as indicated by our results, a positive correlation between returns to schooling and reactions to treatment is not a "universal" consequence from any exogenous change in the determinants of school attendance. For instance, we find no evidence that IV estimates of the returns to schooling using a subsidy to college attendance (or a high school graduation subsidy) as an instrument should lead to particular high returns to schooling (compared to the population average).

Finally, the evidence presented in this paper is in accordance with the results presented in Belzil and Hansen (2002a). Although, in the current paper, heterogeneity in the returns to schooling are interpreted as purely cross-sectional and the returns do not change with schooling level, our estimates are still much smaller than those reported in the literature. Altogether, the results reported therein, along with those reported in a companion paper, point out to the complexities involved in estimating the returns to schooling. The wage regression function is perhaps a highly non-linear (convex) function and the degree of convexity most likely depends on individual specific comparative advantages. At this stage, it is impossible to say whether skill heterogeneity is more important than non-linearities. Only further work will clarify this rather fundamental issue.

References

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

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

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

[4] Belzil, Christian and Hansen, Jörgen (2002b) "Structural Estimates of the Intergenerational Education Correlation", forthcoming in *Journal of Applied Econometrics.*

[5] 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.

[6] Cameron, Stephen and Chris Taber (2002) "" forthcoming in *Journal of Political Economy*

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

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

[9] Florens, Jean-Pierre, James Heckman, Costas Meghir and Edward Vytlacil (2002) " Instrumental Variables, Local Instrumental Variables, and Control Functions" working Paper.

[10] 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.

[11] Heckman, James and E. Vytlacil (2000) "Instrumental Variables Methods for the Correlated Random Coefficient Model", *Journal of Human Resources*, Volume 33, (4), 974-987.

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

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

[14] Magnac, T. and D. Thesmar (2001): "Identifying Dynamic Discrete Decision Processes," forthcoming in *Econometrica.*

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

[16] Meghir, Costas and Marten Palme (2001) "The Effect of a Social Experiment in Education" Working Paper, UCL*.*

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

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

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

[20] Sauer, Robert (2001) "Education Financing and Lifetime Earnings" Working Paper, Brown University.

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

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

[23] Wooldridge, Jeffrey M.(2000) "Instrumental Variable Estimation of the Average Treatment Effect in the Correlated Random Coefficient Model", Working Paper

25

[24] Wooldridge, Jeffrey M. (1997) "On Two-Stage Least Squares estimation of the Average Treatment Effect in a Random Coefficient Model", *Economic Letters (56): 129-133.*

Table 1A
Model Fit: Actual vs Predicted Schooling Attainments

Grade Level	Predicted (%)	Actual (%)
Grade 6	0.0%	0.3 %
Grade 7	1.4%	0.6%
Grade 8	3.4%	2.9%
Grade 9	5.4%	4.7%
Grade10	6.2%	6.0 %
Grade11	7.5%	7.5 %
Grade12	38.4%	39.6 %
Grade13	7.5%	7.0 %
Grade14	5.7%	7.7 %
Grade15	2.7%	2.9 %
Grade16	12.5%	12.9 %
Grade17	2.2%	2.5 %
Grade18	2.7%	2.4%
Grade19	2.0%	1.3%
Grade 20+	1.1%	1.6%

Table 1B
Mean Schooling and Type Probabilities

	Expected Schooling	Type Probabilities (p_k)	q_k (st. error)
type 1	10.81 years	0.1375	-0.0122 (0.0895)
type 2	10.43 years	0.0607	-0.8299 (0.0243)
type 3	13.69 years	0.0951	-0.3808 (0.0314)
type 4	9.41 years	0.0725	-0.6519 (0.0371)
type 5	11.51 years	0.1630	0.1579 (0.0577)
type 6	10.86 years	0.1260	-0.0992 (0.0584)
type 7	10.57 years	0.2059	0.3916 (0.0760)
type 8	12.56 years	0.1392	0.0000 (normalized)

Note: The type probabilities are computed using a logistic transforms;

$$p_k = \frac{\exp(q_k)}{\sum_{j=1}^{8} \exp(q_j)}$$

Table 1C
Model Fit: Actual vs Predicted Wages

Variance of log predicted wages	0.9597
variance of log realized wages	1.2164
Explained Variance (%)	78.9%

Note: Log wages are generated under the assumption that all individuals are aged 30.

28

Table 2A
Absolute and Comparative Advantages in the Labor Market

Parameter (st. error)

Type	Inter. υ_i^w	Educ. φ_{1i}	Wages $\boldsymbol{\lambda}_i$	Experience $\boldsymbol{\lambda}_i \cdot \varphi_2$	$\boldsymbol{\lambda}_i \cdot \varphi_3$	Employment Inter. κ_{0i}
1	1.5325 (0.0308)	0.0858 (0.0052)	1.0000 -	0.1359 (0.0059)	-0.0040 (0.0002)	-3.5753 (0.0363)
2	1.5664 (0.0153)	0.0879 (0.0050)	0.3085 (0.0409)	0.0419	-0.0012	-2.1070 (0.0213)
3	1.3699 (0.0132)	0.0486 (0.0032)	0.1958 (0.0464)	0.0266	-0.0008	-1.5369 (0.0218)
4	1.8741 (0.0321)	0.0595 0.0040)	0.2661 (0.0474)	0.0362	-0.0011	-3.7817 (0.0296)
5	1.2028 (0.0401)	0.0764 0.0037)	1.0866 (0.0472)	0.1477	-0.0043	-3.4752 (0.0286)
6	1.5551 (0.0206)	0.0629 (0.0041)	0.1453 (0.0447)	0.0197	-0.0006	-3.6752 (0.0369)
7	1.3622 (0.0260)	0.0265 (0.0028)	0.9602 (0.0488)	0.1305	-0.0038	-3.4810 (0.0464)
8	1.2539 (0.0156)	0.0400 (0.0031)	0.3417 (0.0352)	0.0464	-0.0014	-3.3763 (0.0400)
ave.	**1.4190**	**0.0576**	**0.6347**	**0.0863**	**-0.0025**	**-3.2559**
S.d.	**0.1810**	**0.0218**	**0.3878**	**0.0527**	**0.0016**	**0.6623**

29

Note: .The estimates of the log inverse employment rate equation are -0.0623 (education), -0.0145 (experience) and 0.0001 (experience squared). The interruption probability is around 7% per year and the log likelihood is -13.7347.

Table 2B
The Utility of Attending School

Param. (st. error)

grade level	Spline $\psi(.)$	Type	Intercept (υ^{ξ})
grade 7-9	0.0164- (0.0080)	**Type 1**	-1.1296 (0.0540)
grade 10	0.3665- (0.0142)	**Type 2**	-1.7791 (0.0922)
grade. 11	-1.0540- (0.0203)	**Type 3**	-1.4172 (0.0384)
grade 12	1.0894 (0.0165)	**Type 4**	-0.8234 (0.0550)
grade 13	-0.5309 (0.0165)	**Type 5**	-1.2595 (0.0487)
grade 14	0.5049 (0.0159)	**Type 6**	-1.1255 (0.0424)
grade 15	-0.8824 (0.0196)	**Type 7**	-0.6397 (0.0326)
grade 16	0.9443 (0.0242)	**Type 8**	-0.9934 (0.0548)
grade 17-	-0.8023 (0.0223)	-	

31

Table 3
Absolute and Comparative Advantages: Type Specific Rankings

| | Schooling | | | Labor Market | | |
| | | | | Wages | | Employment |
	Predicted Schooling	inter. v^ξ	Intercept (abs. adv.) v^w	return to Education φ_1	return to Experience λ	intercept term κ_0
type 1	5	5	4	2	2	3
type 2	7	8	2	1	5	7
type 3	1	7	5	6	7	8
type 4	8	2	1	4	6	1
type 5	3	6	8	3	1	5
type 6	4	4	3	5	8	2
type 7	6	1	6	8	3	4
type 8	2	3	7	7	4	6

Note: To compute the average return to experience, we used the return at 8 years of experience.

Table 4

Correlations between various heterogeneity components

Correlations

	wage intercept	returns to schooling	returns to experience	taste for schooling
wage intercept	1.000	0.2553	-0.4098	0.0272
returns to schooling	-	1.0000	0.1030	0.7175
returns to experience	-	-	1.000	0.2882
taste for schooling	-	-	-	1.000

Note: All correlations are significant at the 1% level.

Table 5A
Estimates of the Effects of Labor Market Skills
and Taste for Schooling on Schooling Attainments

Parameter
(st. error)

	(1)	(2)	(3)	(4)	(5)
$intercept$	4.6112	3.0602	1.9745	2.4369	2.1898
	(0.0844)	(0.0254)	(0.0038)	(0.0011)	(0.0030)
v_i^w	-0.4317	-0.2286	-	-	
	(0.0630)	(0.0337)			
$(v_i^w)^2$	-0.1611	-0.0726	-	-	
	(0.0173)	(0.0113)			
$\varphi_{1i} * 100$	0.1096	0.0065	-	-0.0049	0.0978
	(0.0024)	(0.0003)		(0.0002)	(0.0012)
$(\varphi_{1i} * 100)^2$					-0.0091
					(0.0001)
λ_i	-0.3422	-0.1130	-	-	
	(0.0029)	(0.0002)			
κ_{0i}	0.2344	0.0395	-	-	
	(0.0016)	(0.0007)			
v_i^ξ	0.7743	-	-0.7035	-	
	(0.0214)		(0.0069)		
$(v_i^\xi)^2$	-0.0569	-	-0.2582	-	
	(0.0031)		(0.0031)		
R^2	**0.3452**	**0.2821**	**0.0929**	**0.0042**	**0.0464**

34

Note: The regressions are performed on 200,000 simulated observations. The dependent variable is log schooling and, for convenience, the returns to schooling and experience are multiplied by 100. Similar results may be obtained using schooling (instead of log schooling).

Table 5B
Average Returns to Schooling by Grade Level Achieved

Grade level	Average Returns	St. Deviation
6	0.0464	0.0164
7	0.0473	0.0180
8	0.0489	0.0202
9	0.0534	0.0230
10	0.0609	0.0239
11	0.0607	0.0225
12	0.0592	0.0211
13	0.0547	0.0189
14	0.0502	0.0160
15	0.0461	0.0102
16	0.0445	0.0059
17	0.0443	0.0048
18	0.0442	0.0045
19	0.0423	0.0040

Note: The average returns are based on 200,000 simulated observations.

Table 6A
A Type Specific Analysis of the Effects of a
Counterfactual College Attendance Subsidy

	Δ in Schooling (per type)	Δ in Schooling (ranking)	returns to schooling	Pred. schooling
Type 1	1.7 years	7	0.0858	10.8 years
Type 2	3.0 years	3	0.0879	10.4 years
Type 3	2.8 year	5	0.0486	13.7 years
Type 4	1.4 years	8	0.0595	9.4 years
Type 5	3.3 years	2	0.0764	11.5 years
Type 6	2.5 years	6	0.0629	10.9 years
Type 7	2.9 years	4	0.0265	10.6 years
Type 8	3.16 years	1	0.0400	12.6 years
average (st. dev)	2.7 years (2.2)	-	0.0576 (0.0218)	11.8 years

Table 6B
A Type Specific Analysis of the Effects of a
counterfactual change in the discount rate

	Δ in Schooling (per type)	Δ in Schooling (ranking)	returns to schooling	Pred. Schooling
Type 1	2.1 years	2	0.0858	10.8 years
Type 2	0.7 years	8	0.0879	10.4 years
Type 3	1.31 year	6	0.0486	13.7 years
Type 4	1.0 years	7	0.0595	9.4 years
Type 5	2.6 years	1	0.0764	11.5 years
Type 6	1.4 years	5	0.0629	10.9 years
Type 7	1.9 years	3	0.0265	10.6 years
Type 8	1.4 years	4	0.0400	12.6 years
average (st. dev)	1.7 years (2.3)	-	0.0576 (0.0218)	11.8 years

Table 6C
A Type Specific Analysis of the Effects of a
High School Graduation Subsidy

	Δ in Schooling (per type)	Δ in Schooling (ranking)	returns to schooling	Pred. Schooling
Type 1	0.0 years	8	0.0858	10.8 years
Type 2	1.3 years	6	0.0879	10.4 years
Type 3	1.7 year	3	0.0486	13.7 years
Type 4	0.0 years	7	0.0595	9.4 years
Type 5	1.9 years	2	0.0764	11.5 years
Type 6	1.4 years	5	0.0629	10.9 years
Type 7	1.5 years	4	0.0265	10.6 years
Type 8	1.4 years	1	0.0400	12.6 years
average (st. dev)	1.33 years (2.2)	-	0.0576 (0.0218)	11.8 years

Table 6D
The determinants of the individual specific reactions
to a college attendance subsidy and a change in discount rate

Parameter (st. error)

	College attendance subsidy		Decrease in discount rate		High school graduation subsidy	
	(1)	(2)	(3)	(4)	(5)	(6)
$intercept$	-8.2350 (0.8536)	3.2170 (0.0420)	7.2156 (0.4835)	2.6480 (0.0450)	-6.6068 (0.4425)	0.8667 (0.0422)
v_i^w	-0.1135 (0.3859)	-	-2.2556 (0.1168)	-	-0.7002 (0.1062)	-
$\varphi_{1i}*100$	-0.1675 (0.0405)	-0.0290 (0.0025)	0.3582 (0.0361)	-0.4430 (0.0180)	-0.3168 (0.0329)	0.4069 (0.0166)
$(\varphi_{1i}*100)^2$	-0.0542 (0.0036)	-0.0070 (0.0010)	-0.0069 (0.0039)	0.0431 (0.0020)	-0.0441 (0.0036)	-0.0496 (0.0015)
λ_i	6.8437 (0.0029)	-	-3.4167 (0.2721)		1.3485 (0.2482)	
λ_i^2	-4.3433 (0.1656)		3.0235 (0.1790)		-0.3044 (0.1633)	
κ_{0i}	-1.4679 (0.0638)	-	0.3349 (0.0348)		-1.4048 (0.0316)	
v_i^ξ	-6.9336 (0.1640)	-	2.3425 (0.1783)		-6.6557 (0.1616)	
R^2	**0.0977**	**0.0125**	**0.0524**	**0.0112**	**0.1307**	**0.0288**

Table 7A
Estimating the Asymptotic Bias

Estimate (P. Value)

	$\mathrm{plim}\frac{W'\xi}{N}$	$\mathrm{plim}(\hat{\beta}_{ols} - \beta)$
Education	6.04	0.04
	(0.01)	(0.01)
Experience	2.09	-0.03
	(0.01)	(0.01)
Experience^2	15.27	0.0013
	(0.01)	(0.01)

Note: The OLS estimates for education, experience and experience^2 will fluctuate according to the specific cross-section (year) chosen. The OLS estimate for education will typically lie between 8% and 10% per year while the returns to experience will be between 3% and 6% per year.

Table 7B
Correlation Matrix of W

$$
\begin{bmatrix}
 & educ & \exp er & \exp er^2 \\
educ & 1.0000 & -0.5158 & -0.5288 \\
\exp er & - & 1.0000 & 0.9553 \\
\exp er^2 & - & - & 1.0000
\end{bmatrix}
$$

Appendix 1
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.[24] 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 along the lines of results reported in Keane and Wolpin (1997).[25] 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

---

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

[25]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.

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.

Table A1 - 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.