A Structural Analysis of the Correlated Random Coefficient Wage Regression Model with an Application to the OLS-IV Puzzle

Christian Belzil
Jörgen Hansen

May 2005

DISCUSSION PAPER SERIES

A Structural Analysis of the Correlated Random Coefficient Wage Regression Model with an Application to the OLS-IV Puzzle

Christian Belzil
CNRS-GATE, CIRANO, CIREQ
and IZA Bonn

Jörgen Hansen
Concordia University, CEPR, CIRANO, CIREQ
and IZA Bonn

Discussion Paper No. 1585
May 2005

ABSTRACT

A Structural Analysis of the Correlated Random Coefficient Wage Regression Model with an Application to the OLS-IV Puzzle[*]

We estimate a finite mixture dynamic programming model of schooling decisions in which the log wage regression function is set within a correlated random coefficient model and we use the structural estimates to perform counterfactual experiments. We show that the estimates of the dynamic programming model with a rich heterogeneity specification, along with simulated schooling/wage histories, may be used to obtain estimates of the average treatment effects (ATE), the average treatment effects for the treated and the untreated (ATT/ATU), the marginal treatment effect (MTE) and, finally, the local average treatment effects (LATE). The model is implemented on a panel of white males taken from the National Longitudinal Survey of Youth (NLSY) from 1979 until 1994. We find that the average return to experience upon entering the labor market (0.059) exceeds the average return to schooling in the population (0.043). The importance of selectivity based on individual specific returns to schooling is illustrated by the difference between the average returns for those who have not attended college (0.0321) and those who attended college (0.0645). Our estimate of the MTE (0.0573) lies between the ATU and ATT and exceeds the average return in the population. Interestingly, the low average wage return is compatible with the occurrence of very high returns to schooling in some subpopulation (the highest type specific return is 0.13) and the simulated IV estimates (around 0.10) are comparable to those very high estimates often reported in the literature. The high estimates are explained by the positive correlation between the returns to schooling and the individual specific reactions. Moreover, they are not solely attributable to those individuals who are at the margin, but also to those individuals who would achieve a higher grade level no matter what. The structural dynamic programming model with multi-dimensional heterogeneity is therefore capable of explaining the well known OLS/IV puzzle.

JEL Classification: J2, J3

Keywords: random coefficient, returns to schooling, treatment effects, dynamic programming, dynamic self-selection

Corresponding author:

Christian Belzil
Groupe d'Analyse et de Théorie Economique
CNRS UMR 5824 - University of Lyon 2
93, chemin des Mouilles - B.P.167
69131 ECULLY cedex
France
Email: belzil@gate.cnrs.fr

---

1 Introduction and Objectives

In this paper, we investigate the empirical implications of the correlated random coefficient wage regression model (CRCWRM) using a structural dynamic programming model and we analyze the performance of several instrumental variable (IV) estimates obtained from artificial experiments which have been simulated from the structural parameter estimates. The term "Correlated Random Coefficient Wage Regression Model" refers to the standard Mincerian log wage regression function in which the coefficients, in particular the return to schooling, may be arbitrarily correlated with the regressors.[1] In the context of the wage equation, differences in the wage intercept capture differences in abilities to draw higher wages independently from accumulated human capital whereas a higher slope (for the education parameter) reflects a higher ability to draw high wages in conjunction with schooling. In a world where skills rewarded in the labor market are multi-dimensional, differences in wage intercepts may therefore reflect differences in motivation or differences in nonacademic skills (perhaps blue collar skills) whereas differences in the slopes may reflect differences in academic skills.[2]

Economists have only recently paid particular attention to the specification and the estimation of linear wage regression models set in a random coefficient framework (Imbens and Angrist (1994), Heckman and Vitlacyl (1998), Wooldridge (1997 and 2003), Card (2001), 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 the parameters of the Mincerian wage regression are understood within a skill production function framework, allowing no interaction between ability(ies) and other inputs such as education and experience appears somewhat arbitrary.

As it stands, very little is known about the empirical implications of the CRCWRM and, in particular, about the level of selectivity attributable to

---

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

[2]It should be clear that the model analyzed in this paper does not allow individuals to specialize in a particular skill. For a multiple skills model in a dynamic environment, see Keane and Wolpin (1997).

each coefficient of the CRCWRM. Those estimating the returns to schooling by IV techniques use a framework which can neither disclose the covariances between realized schooling and the individual specific abilities (the wage intercept and the return to schooling) nor the variances of these individual specific components. However, these quantities are important. They may help comprehend the determinants of individual schooling attainments and may help quantify the "Ability Bias" arising in estimating the returns to schooling using regression techniques. Obviously, a random coefficient regression model provides a more realistic framework to evaluate the relative importance of labor markets skills and taste for schooling in explaining educational selectivity than a standard regression framework with individual specific intercepts. Virtually all recent work on empirical earnings functions is directly or indirectly interpreted within a random coefficient framework. For this reason, its properties and implications deserve some attention.[3]

In this paper, 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 return to schooling, a specific wage intercept as well as a specific utility of attending school.[4] The estimation of a mixed likelihood function can capture any arbitrary correlation between 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.[5]

This paper is driven by three main objectives. Our first objective is to investigate the empirical implications of the CRCWRM and characterize the degree of heterogeneity in each component of the wage regression. We focus on three specific components, the utility of attending school, the wage

---

[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]We also allow for unobserved heterogeneity in the yearly employment rate (which is also affected by schooling).

[5]However, we have also estimated a version of the model where all key individual specific components are function of observed characateristics.

regression intercept (the classical measure of unobserved ability) and the returns to schooling. In particular, we measure the relative importance of labor market skills and the individual specific taste for schooling in explaining cross-sectional differences in schooling attainments and we determine the relative importance of the wage intercept versus the return to schooling with respect to selectivity.

A second objective is to illustrate the importance of population heterogeneity and its implications for various measures of treatment effects used in the literature. As is done in most of the literature, we focus on college attendance as the treatment of interest. We show that the estimates of the dynamic programming model with a rich heterogeneity specification, along with simulated schooling/wage histories, may be used to obtain estimates of the average treatment effects (ATE), the average treatment effects for the treated and the untreated (ATT/ATU), the marginal treatment effect (MTE), which is defined as the average return for those who are indifferent between entering college and leaving school after high school graduation, and, finally, the local average treatment effects (LATE).[6]

Finally, the third objective is to reconcile the well known discrepancy between OLS estimates of the returns to schooling and their IV counterparts and explain the well-known OLS/IV puzzle. This is an important issue. The OLS/IV puzzle may be illustrated by a simplified version of the Mincerian wage offer equation,

$$w_i = \beta_0 + \beta_1 \cdot Schooling_i + \eta_i \tag{1}$$

where w_i denotes log wages and η_i plays the role of unobserved market ability. The "Ability Bias" (the sign of the statistical bias that may occur when measuring the causal effect of schooling on wages by a simple correlation or by OLS techniques) has been the subject of much debate over the last 40 years. Ignoring post-schooling labor market experience, the discrepancy between OLS and IV estimates is a reflection of the correlation between schooling and unobserved ability (η_i). A positive (negative) correlation is associated with a positive (negative) ability bias. In the IV literature, the ability bias is only indirectly investigated through the discrepancy between IV and OLS estimates, assuming that the linear model is correct. Card (2001 and 2002)

---

[6]These different notions are discussed at length in Carneiro, Heckman and Vitlacyl (2003).

reports that a large number of studies find that IV estimates exceed OLS estimates by a wide margin and estimates of the order of 15% per year of schooling are not uncommon.[7] In the context of a random coefficient model, the IV estimator is sometimes referred to as a Local Average Treatment Effect (LATE).[8] The LATE should be understood as a measure of the returns to schooling for the sub-population affected by the experiment. It is often postulated that the high returns are explained by the fact that those individuals more likely to react to an exogenous policy change are those who are at the margin of deciding to enter college before the policy change and that they have higher returns to schooling than average. As far as we know, this claim has neither been proved nor verified empirically in any direct fashion. With structural estimates in hand, we are able to simulate exogenous (artificial) experiments and analyze the characteristics of the sub-population (s) most affected by it. To do so, we simulate three different experiments; an overall decrease in discount rate (experiment 1), a decrease in discount rate from grade 13 onward (experiment 2), and an increase in the utility of attending school from grade 12 onward (experiment 3).[9] In turn, these artificial experiments provide us with a valid instrument which may be used in order to estimate the returns to schooling by standard IV techniques and analyze the discrepancies between IV, OLS (on simulated data) and estimates from our structural dynamic programming model.

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 1994. The main results are as follows. As is common in the structural literature, we find the population average return to schooling to be much below those reported in the existing literature. Our estimates are also lower than those obtained using standard OLS techniques. The return to experience upon entering the labor market (0.059) exceeds the average return to schooling in the population (0.043) and we find that the

---

[7]The validity of very high returns to schooling, reported in a simple regression framework has 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.

[8]See Imbens and Angrist (1994).

[9]In the literature, it is sometimes argued that differences in credit constraints may be captured in the discount rate (see Cameron and Taber (2004) for a recent example).

cross-sectional variability in the returns to schooling is smaller than the cross-sectional variance in the intercept term.

Interestingly, the low wage return in average is compatible with the occurrence of very high returns (the highest type specific return is 0.13) to schooling in some subpopulations. The importance of selectivity based on individual specific returns to schooling is illustrated by the difference between the average returns for those who have not attended college (0.0321) and those who attended college (0.0645). The estimates therefore conform to the intuition. Indeed, the estimate of the average treatment effect (the population average) lies somewhere in between the ATU and the ATT. Our estimate of the MTE (the average return for those who are indifferent between entering college and stopping after high school graduation) is equal to 0.0573. It lies between the ATU and ATT. It also exceeds the average return in the population.

Finally, we show that the degree of dispersion found in the population returns to schooling can be reconciled with the existence of high IV (or LATE) estimates. The importance of individual heterogeneity in the level of reactions to policy changes is well illustrated by the three counterfactual experiments that we implement. The IV estimates resulting from these experiments are 0.1211, 0.0979, and 0.1096, and are they comparable to those very high estimates often reported in the literature. They exceed both the OLS estimate and the average of the type specific returns obtained from the structural dynamic programming model. We show that this follows from the high positive correlation between the returns to schooling and the individual specific reactions and, in particular, that this correlation is not solely attributable to those individuals who are at the margin, but also to those individuals who would achieve a higher grade level in any event. The structural dynamic programming model with multi-dimensional heterogeneity is therefore capable of explaining the coexistence of relatively low returns to education (on average) with very high returns for some identified sub-populations and, in particular, capable of identifying the characteristics of those individuals who would react to a counterfactual experiment. In short, our model is capable to explain to well known OLS/IV puzzle.

The paper is structured as follows. The empirical dynamic programming model is exposed in Section 2. The main results are discussed in Section 3 (the parameter estimates and the model fit). In Section 4, we discuss various measures of treatment effects introduced in the experimental literature and

7

show that all of them can be inferred from the dynamic programming model. In Section 5, we analyze the determinants of the individual specific reactions to a specific counterfactual experiment. In Section 6, we focus on the OLS/IV puzzle and in Section 7, we provide a discussion on out-of-sample fit. Finally, Section 8 concludes the paper.

2 An Empirical Dynamic Programming Model of Schooling Decisions

We assume that young individuals decide sequentially whether it is optimal or not to enter the labor market or continue accumulate schooling. Individuals are further assumed to maximize discounted expected lifetime utility over a finite horizon T and they have identical preferences. Both the instantaneous utility of being in school and the utility of working 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} in the model. The theoretical structure of the problem is similar to the model found in Belzil and Hansen (2002a), but the specification of the wage offer equation is different as it is set in a correlated random coefficient model.

2.1 The Utility of Attending School

The instantaneous utility of attending school, $U^s(.)$, should be interpreted as the monetary equivalent (on a per hour basis) of attending school. It is formulated as the following equation

$$U_{it}^s(.) = \psi \cdot I(S_{it} > 12) + \upsilon_i^\xi + \varepsilon_{it}^\xi \tag{2}$$

in which $I(S_{it} > 12)$ is an indicator function which equals one at grade levels exceeding 12, υ_i^ξ represents individual heterogeneity (ability) affecting the utility of attending school, and ε_{it}^ξ represents a stochastic utility shock. The parameter ψ captures the potential decrease in the net utility of attending

school beyond high school completion which may arise because the direct (or psychic) costs of attending college may be substantially higher than in high school.

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}.[10] 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 schooling remains constant over that period.[11]

2.2 The Utility of Work

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

$$U^w_{it}(.) = \ln(\tilde{w}_{it}) = \ln(e_{it} \cdot w_{it})$$

2.3 The Correlated Random Coefficient Wage Regression Model

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

$$\ln w_{it} = \varphi_{1i} \cdot S_{it} + \varphi_2 \cdot Exper_{it} + \varphi_3 \cdot Exper_{it}^2 + v_i^w + \varepsilon_{it}^w \qquad (3)$$

where φ_{1i} is the individual specific wage return to schooling, while φ_2 and φ_3 measure the effect of experience and experience squared. The term v_i^w represents an individual specific intercept term and ε_{it}^w denotes a pure stochastic shock. We assume that

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

---

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

[11]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.

where $\bar{\varphi}_1$ represents the population average.[12] Following the convention used in the literature, it is convenient to specify the wage regression as a heteroskedastic regression function

$$\ln w_{it} = \bar{\varphi}_1 \cdot S_{it} + \varphi_2 \cdot Exper_{it} + \varphi_3 \cdot Exper_{it}^2 + \varrho_{it} \qquad (4)$$

where

$$\varrho_{it} = \upsilon_i^w + \omega_{1i} \cdot S_{it} + \varepsilon_{it}^w$$

Estimating the population average returns to schooling and experience ($\bar{\varphi}_1$, φ_2, and φ_3) is rendered difficult by the fact that typically

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

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

$$\ln e_{it}^* = \ln \frac{1}{e_{it}} = \kappa_1 \cdot S_{it} + \kappa_2 \cdot Exper_{it} + \kappa_3 \cdot Exper_{it}^2 + \upsilon_i^e + \varepsilon_{it}^e \qquad (5)$$

where κ_1 represents the employment security return to schooling, κ_2 and κ_3 represent the employment security return to experience, and υ_i^e is an individual specific intercept term. The term ε_{it}^e is a random shock.[13]

2.5   The Stochastic Shocks

We assume that $\varepsilon_{it}^s \sim i.i.d\ N(0, \sigma_s^2)$ for $s = \xi, w$ and e. All random shocks ($\varepsilon_{it}^\xi, \varepsilon_{it}^w, \varepsilon_{it}^e$) are assumed to be independent from each other as well as independent of S_{it} and $Exper_{it}$.[14] This simplifying assumption is common in the structural literature and it implies, for instance, that all persistence in wages is explained solely by human capital or omitted abilities.

---

[12]In a companion paper, Belzil and Hansen (2002b), we estimate the model with individual specific returns to experience as well.

[13]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 $Var(e_t) = \exp(2\mu_t + \sigma_e^2) \cdot (\exp(\sigma_e^2) - 1)$.

[14]Off course, d_{it} (and therefore S_{it+1}) depends on the $\varepsilon_{it}'s$.

2.6 The Value Functions

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. The value function associated with the decision to remain in school, given accumulated schooling S_{it}, denoted $V_{it}^s(S_{it}, \eta_{it})$, can be expressed as

$$
\begin{aligned}
V_{it}^s(S_{it}, \eta_{it}) \;=\; & U_{it}^s(.) + \beta\{\zeta \cdot EV_{it+1}^I(S_{it+1}, \eta_{it+1}) \\
& + (1-\zeta) \cdot EMax[V_{it+1}^s(S_{it+1}, \eta_{it+1}), V_{it+1}^w(S_{it+1}, \eta_{it+1})]\}
\end{aligned}
\tag{6}
$$

where $V_{it+1}^I(S_{it+1}, \eta_{it+1})$ 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 set equal to the value of attending school. Given this assumption, $V_{it+1}^I(S_{it+1}, \eta_{it+1})$ can be expressed as follows

$$
\begin{aligned}
V_{it+1}^I(S_{it+1}, \eta_{it+1}) \;=\; & U_{it+1}^s(.) + \beta\{\zeta \cdot EV_{it+2}^I(S_{it+2}, \eta_{it+2}) \\
& + (1-\zeta) \cdot EMax[V_{it+2}^s(S_{it+2}, \eta_{it+2}), V_{it+2}^w(S_{it+2}, \eta_{it+2})]\}
\end{aligned}
$$

The value of stopping school (that is entering the labor market), $V_{it}^w(S_{it}, \eta_{it})$, is given by

$$
V_{it}^w(S_{it}, \eta_{it}) = \ln(w_{it} \cdot e_{it}) + \beta E(V_{it+1} \mid d_{it} = 0)
\tag{7}
$$

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

$$
E(V_{it+1} \mid d_{it} = 0) = \sum_{j=t+1}^{T} \beta^{j-(t+1)}\left(-\exp(\mu_{ij}^e + \frac{1}{2}\sigma_e^2) + \mu_{ij}^w\right)
$$

with

$$
\mu_{ij}^e = \kappa_1 \cdot S_{ij} + \kappa_2 \cdot Exper_{ij} + \kappa_3 \cdot Exper_{ij}^2 + \upsilon_i^e
$$

11

and

$$\mu_{ij}^{w} = \bar{\varphi}_1 \cdot S_{ij} + \varphi_2 \cdot Exper_{ij} + \varphi_3 \cdot Exper_{ij}^2 + v_i^w + \omega_1 \cdot S_{ij}$$

Thus, $E(V_{it+1} \mid d_{it} = 0)$ is 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.

2.7 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 $(v_k^\xi, v_k^w, v_k^e, \varphi_{1k})$ for $k = 1, 2...K$. The results reported in this paper are for the case $K = 8$. It is possible to obtain very similar results for all population averages, variances and covariances with 12 types. However, the analysis of heterogeneity presented in Section 4 is rendered substantially easier when the number of types is limited. The results reported in this paper should be seen as resulting from the desire to find a model with robust results which has the smallest number of types. More details are provided in Section 3.

The probability of belonging to type k, p_k, is estimated using a logistic transform

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

with the restriction that $q_8 = 0$.

2.8 Identification

As is the case in all empirical dynamic programming models, identification may only be considered within a particular parametric structure. Structural models require us to specify preferences (or technologies) and are therefore based on both functional forms and parametric assumptions. Moreover, estimating a structural model with unobserved heterogeneity requires to distinguish between the distribution of random shocks and the distribution of

12

unobserved heterogeneity. Typically, and as is generally the case in most complicated non-linear models, structural models are only locally identified.

Since our model is 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 wages and employment rates given a distributional assumption about the stochastic shock. In the present case, where the wage regression function is 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 5% per year (an estimate practically identical to the estimate found in Belzil and Hansen (2002).

2.9 The Likelihood Function

Constructing the likelihood function is relatively straightforward. The likelihood function 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 1994 (denoted L_{3k}). Using the definitions of d_{it} and I_{it} above, it is easy to specify all transition probabilities needed to derive the likelihood function. The transition probabilities that define the choice between interrupting school permanently (start working) and obtaining an additional year of schooling, are given by

$$\Pr(d_{it+1} = 0 \mid d_{it} = 1) = (1 - \zeta) \cdot \Pr(V_{it}^w(S_{it}, \eta_{it}) \geq V_{it}^s(S_{it}, \eta_{it})) \qquad (8)$$

$$\Pr(d_{it+1} = 1 \mid d_{it} = 1) = (1 - \zeta) \cdot \Pr(V_{it}^w(S_{it}, \eta_{it}) < V_{it}^s(S_{it}, \eta_{it})) \qquad (9)$$

$$\Pr(I_{it+1} = 1 \mid d_{it} = 1) = \zeta \qquad (10)$$

where $\Pr(V_{it}^w(S_{it}, \eta_{it}) \geq V_{it}^s(S_{it}, \eta_{it}))$ can be evaluated using equations (6) and (7) above. Equation (8) represents the probability of exercising the right to leave school permanently in $t+1$ (implicitly assuming $I_{it+1} = 0$) while equation (9) represents the probability of staying in school to acquire an additional year of human capital (also implicitly assuming $I_{it+1} = 0$). Equation (10) represents the exogenous probability of entering the interruption state. The likelihood function is constructed from data on the allocation of time between years spent in school ($I_{it} = 0, d_{it} = 1$), years during which school was interrupted ($I_{it+1} = 1, d_{it} = 1$), and information on wages and unemployment episodes.

Ignoring the individual identification subscript, the components of the likelihood function for a given type k are

- the probability of observing a particular sequence of school/interruption histories that can be easily derived from equations (8), (9) and (10) above

$$L_{1,k} = Pr[(d_{0,k} = 1, I_{0,k}), (d_{1,k} = 1, I_{1,k})....(d_{\tau,k} = 1, I_{\tau,k})]$$

- the probability of entering the labor market in year $\tau + 1$, at observed wage $w_{\tau+1}$, which can be factored as the product of a normal conditional probability times a marginal

$$L_{2,k} = \Pr(d_{\tau+1,k} = 0, w_{\tau+1,k}) = \Pr(d_{\tau+1,k} = 0 \mid w_{\tau+1,k}) \cdot \Pr(w_{\tau+1,k})$$

- the density of observed wages and employment rates from $\tau + 2$ until 1994

$$L_{3.k} = Pr\left(\{\tilde{w}_{\tau+2.k}\}..\{\tilde{w}_{1990.k}\}\right) = \Pr\left(\{w_{\tau+2.k} \cdot e_{\tau+2.k}\} \cdot .... \Pr\{w_{1994.k} \cdot e_{1994.k}\}\right)$$

Thus, the likelihood function to be maximized is

$$\ln L = \ln \sum_{k=1}^{8} p_k \cdot L_{1k} \cdot L_{2k} \cdot L_{3k} \tag{11}$$

where each p_k represents the population proportion of type k.

3 An Overview of the Results

The model has been estimated on a sample of white males taken from the 1979 youth cohort of the National Longitudinal Survey of Youth (NLSY). We restrict our sample to white males from the core random sample who were 14 to 16 years old in 1979. We record information on education, wages, and employment rates for each individual from the time the individual is 14 up to December 31, 1994. The NLSY is a widely used data set and the sample used in this work is virtually the same as that analyzed in Eckstein and Wolpin (1999). For this reason, details are only provided in Appendix.

When we estimated the model, we experienced with a relatively large number of types. We considered a maximum of 12 types. However, in order to facilitate presentation of the results, we tried to select a specification which contained the smallest number of types possible while preserving the same fundamental results (making sure that the population averages, the standard deviations and the covariances were comparable). The results reported in the following sections have been obtained under the assumption that the distribution of tastes and abilities is represented by 8 types.

In the literature, it is customary to evaluate in sample fit solely by looking at the predicted frequencies. However, the model estimated here is similar to the theoretical structures found in Belzil and Hansen (2002a), which studied non-linearities in the wage regression, in Belzil and Hansen (2002b), which is a companion paper, and Belzil and Hansen (2003), which focused on the intergenerational education correlation. The capacity of the model to fit the data on schooling attainments is well documented in all of these papers, especially in Belzil and Hansen (2002b), and we therefore do not analyze it here in order to save space. Indeed, this paper focuses on labor market wages and the relationship between the structural dynamic programming estimates and various reduced-form estimates. We shall therefore pay a particular attention to the capacity of the model to explain post schooling wages.

3.1 Some Key Parameter Estimates

The parameters raising most interests are those pertaining to the wage equation as well as the utility of attending school. These are found in Tables 1 and 2. Our estimates of the returns to schooling (found in column 3 of Table 1) range from 0.0032 (type 7) to 0.1278 (type 4). The population average return

to schooling is 0.0428. The cross-sectional dispersion is however found to be important, as the standard deviation (0.0391) is practically as high as the population average. Our estimate of the returns to experience, 0.0586 (Table 2) exceeds the return to schooling and the estimate for φ_3 (-0.0003) implies a relatively weak concavity. Overall, these estimates are much lower than those reported in the experimental (IV) literature but similar to those found in empirical studies based on structural methods (see for instance Keane and Wolpin (1997) and Belzil and Hansen (2002a)).[15]

The dispersion in the individual specific wage intercept terms (column 2 in Table 1) is also found to be quite important. The estimates range between 0.8534 (type 4) and 1.5807 (type 6) and the standard deviation is equal to 0.21. It is also informative to examine the estimates of the intercept terms of the utility of attending school (reported in column 1 in Table 1). 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 -2.9461 (type 5) to -0.4853 (type 2).

The estimate for the effect of reaching grade 13 or more on the net utility of attending school, ψ is found in Table 1B. It is negative (-0.1284) and indicates that the monetary equivalent of the net utility of attending school decreases after grade 12. This is obviously consistent with the view that college attendance may be more costly than high school attendance.

3.2   Heterogeneity and Schooling Attainments

The type specific expected schooling attainments are reported in Table 3. They range from 9.2 years (type 2) to 15.2 years (type 7). Table 4 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, individuals take optimal schooling decisions based on their individual specific taste for schooling and their labor market skills. While some individuals are endowed with a high taste for schooling, schooling decisions are largely affected by the combination of the returns to schooling and the intercept terms of the wage regression. As a

---

[15]Belzil (2005) addresses the discrepancy between structural and experimental estimates of the returns to schooling and experience in his survey of the structural literature.

consequence, it will be impossible to associate a definite set of attributes (say, high or low return to human capital) to each specific type solely on the basis of their expected schooling attainments. Nevertheless, our model is sufficiently rich to capture differences in abilities and tastes among types of individuals that might obtain similar levels of schooling.

To illustrate this, consider the set of individuals (type 2 and type 8) who are predicted to obtain relatively low levels of schooling. Type 2 individuals obtain a low level of schooling (9.2 years) because they have a low return to schooling and a high wage intercept, despite a very high taste for schooling. At the same time, type 8 individuals, who also obtain a low level of schooling (10 years), are endowed with low return to schooling and low wage intercept but a relatively low utility of attending school.

The mechanics of the model can also be illustrated at the higher end of the schooling spectrum. Both type 7 and type 4 individuals are predicted to attain a high level of schooling (15.2 years and 14.3 years, respectively). Clearly, type 4 individuals obtain a high level of schooling because they are endowed with a particularly high return to schooling (the highest in the population) and a relatively low wage intercept. Type 7, on the other hand, reaches a high level of schooling mostly because they are endowed with a higher utility of attending school.

To investigate formally the determinants of individual schooling attainments implied by our estimates, we simulated our model and generated 70,000 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) on schooling attainments. To have a clear picture of the relative importance of each heterogeneity component, we regressed simulated schooling attainments on each heterogeneity component separately (columns 1 to 3 in Table 5), on labor market abilities (column 4 in Table 5) and on labor market ability and taste for schooling (column 5 in Table 5) and report the resulting R^2.

The results indicate that when taken individually, the wage intercept accounts for the largest share of the total variance (0.39). The utility of attending school is the second most important (0.28) and the returns to schooling only account for 7% of the total variation. When taken simultaneously, labor market skills (wage intercept and the return) account for 40% of the total variation. Therefore, selectivity appears to be more strongly based on the intercept term than on the return to schooling itself. We may conclude that,

even though a random coefficient model provides a more realistic representation of the relationship between market wages and abilities, it does not obviate the need to allow for cross-sectional dispersion at the level of the intercept term.

3.3 Predicted Schooling, Wages and in-Sample Fit

As we focus on the ability of the structural estimation to estimate the correlated random coefficient wage regression model, we examine the capacity of our model to replicate the actual distribution of wages. This is our measure of in-sample fit. Out-of-sample fit will be discussed below (in Section 6). Table 6 reports empirical and predicted wage distributions. Despite a certain underprediction at the low end, the model is clearly able to predict wage frequencies over most of the relevant interval.

Perhaps more importantly, it is important to evaluate if the model is capable of generating a correlation between schooling and wages comparable to the correlation found in the actual data. To do so, we also ran OLS on the simulated data. We find the OLS estimate obtained on simulated wages (0.0467) to be reasonably close to the actual OLS estimate computed on pooled data (0.0524). It further indicates the capacity of our model to explain wages relatively accurately. More importantly, the OLS estimate exceeds the population average return (0.0428) by a wide margin. This is not surprising. It reflects the invalidity of the OLS estimate in a context where the wage regression is a correlated random coefficient wage regression model. The degree of over-estimation is consistent with the classical (positive) ability bias hypothesis.

4 Measuring Various Treatment Effects

In the treatment effect literature, the schooling decision is typically specified as a static problem or, at most, as a two period problem and it is customary to focus on college attendance as the treatment of interest. In such a case, the schooling decision is represented as a binary choice problem. However, different notions of treatment are often proposed. These include the average treatment effect (ATE), the average treatment for the treated (ATT) or for the un-treated (ATU), the marginal treatment effect (MTE) and the local

average treatment effect (LATE).

Within a binary choice model, the individual decisions are determined solely by individual unobserved (to the econometrician) heterogeneity. This means there is no distinction to be made between the ex-ante choice and the ex-post realization. In a dynamic stochastic environment, the notion of treatment is somewhat different for two main reasons. First, treatment is sequential by nature. Every level of schooling is by definition a potential treatment. Second, the data generating process (DGP) incorporates abilities and tastes (unobserved heterogeneity) as well as stochastic shocks. Even if the effect of schooling on wage does not change with schooling levels (for a given individual), the population (or sub population) of those individuals exercising a choice between stopping school at a given level or continuing to the next is changing as we move up with grade level achievement. However, the importance of this selection effect is inversely related to the importance of stochastic shocks.

In general, the set of individuals contemplating the possibility to stay in school for an extra period (say enter college) is composed of individuals who are endowed with tastes/abilities that make them likely to reach this level ex ante as well as individuals who have experienced a particular history of stochastic shocks. Therefore, while some of these parameters may be recovered directly from the point estimates of the structural dynamic programming model with heterogeneity, others will only be indirectly recovered from related simulations.[16]

The first, and perhaps the most natural, notion of treatment is the ATE. It refers to the population average causal effect of schooling. Obviously, it may be recovered directly from the parameter estimates (the type probabilities and the return to schooling).

The second notion of treatment considered is the ATT (or the ATU). It is defined for a given benchmark level of schooling and, in the case where the focus is on college attendance, it refers to the population average returns to schooling of those who attended (did not attend) college. In the context of our structural dynamic programming model, the ATT and ATU may be computed using simulated schooling/wage histories.

The third notion, the MTE, is also dependent on the benchmark chosen

---

[16]Strictly speaking, these are not parameters but rather functions of the type specific parameters and the realized stochastic shocks.

but it refers to the average return of those who would be indifferent between attending college or not.[17] Within a structural dynamic programming model with unobserved heterogeneity approximated by discrete types, the MTE must be understood as an ex ante notion. It may be estimated by the return of the type of individuals who are expected to complete between 12 and 13 years of schooling. However, this requires that, for at least one of the types, the expected schooling lies in between 12 and 13. If not, then it may be possible to choose the types who have an expected schooling level in this neighborhood (slightly below 12 or slightly above 13) and who are, as a group, expected to complete between 12 and 13 years of schooling on average.

Finally, the fourth notion considered is the LATE. The LATE is different in nature. It is not represented naturally as a function of the fundamental parameters of the model and, moreover, it is defined solely within the context of an instrument (or an experiment). As such, its investigation will also require simulations. Because of its peculiar aspect, we will discuss it separately in Section 6.

In Table 7, we report the average treatment effects for both sub-populations of individuals who have not attended college (those who have completed grade 12 or less) and those who have attended college (those who have completed grade 13 or more). The ATT and the ATU are computed from 70,000 simulated schooling/wage histories. The importance of selectivity based on individual specific returns to schooling is illustrated by the difference between the average returns for those who have not attended college (0.0321) and those who attended college (0.0645). The estimates therefore conform to the intuition. Indeed, as seen earlier, our estimate of the average treatment effect (the population average), which is equal to 0.0428, lies somewhere in between the ATT and the ATU.

As we do not have a specific type of individuals predicted to choose 12.5 years of schooling ex-ante, we use a weighed average of the 3 types who are the closest to 12.5 years. These are type 3, type 4 and type 5. Altogether this sub-population, when weighted appropriately, is expected to complete 12.56 years of schooling. Our estimate of the MTE is equal to 0.0573 and is therefore constructed as an average return over these 3 types of individuals (weighted

---

[17]The notion of marginal treatment effects is more recent. It is mentioned in Björklund and Moffitt (1987) and also appears in Carneiro, Heckman and Vytlacil (2003) and in Heckman and Vytlacil (2004).

by the appropriate type probabilities). Not surprisingly, it lies between the ATT and ATU. It also exceeds the average return in the population.

5 Skill Heterogeneity and the Reactions induced by an Artificial Experiment

Given the relatively rich level of heterogeneity allowed for in our model, it is possible to investigate directly how differences in abilities and tastes relate to individual specific reactions to some "exogenous policy change" and, in particular, how individual differences in the returns to schooling relate to the reactions induced by exposure to an instrument. In practice, any "natural experiment" creating changes in the incentives to attend school must affect at least one of the basic components of the decision making (the fundamental parameters v_k^ξ, φ_{1k}, v_k^w and β). By changing any of these fundamental parameters, it is therefore possible to recover the effects of a corresponding exogenous policy change. In general, the choice of experiment entails choosing which parameter to change and, moreover, over which period(s). For instance, in some particular context, policy makers may be interested in increasing the general level of schooling attainments and introduce a school attendance subsidy at all grade levels. In some other context, the policy of interest may be targeted at higher education and a subsidy may therefore only be paid to those who reach a particular level (say college). In this section, we examine one particular counterfactual experiment in depth; namely a decrease in the rate of time preference (from 0.05 to 0.04).[18] This is an experiment which, as such, is not targeted a specific grade level. Using simulated schooling/wage trajectories, we examine how individual reactions are affected by heterogeneity and, in particular, by the individual specific returns.

Again, we use our model as the data generating process (DGP) and simulate approximately 70,000 outcomes of education and entry wages. With a simple before/after comparison, it is straightforward to compute average type specific reactions. These are found in Table 8A. As a first step, it is informative to proceed with a simple descriptive analysis. First, it may be

---

[18]In the literature, it is sometimes argued that differences in credit constraints may be captured in the discount rate (see Cameron and Taber (2003) for a recent example).

noted that the type specific average reaction rankings have changed in a nontrivial way when compared to the schooling attainment ranking (Table 3). For instance, type 6 and type 4 individuals are now those who are reacting the most to this counterfactual experiment and they are also endowed with much higher returns to schooling (0.0984 and 0.1278) than the population average while type 1 and type 3 individuals (the next groups that react the most) are endowed with lower returns. Notably, some types (type 2 and type 8) have a small negative reaction. This is explained by their low returns to schooling, when compared to the return to experience.[19]

Given the dimensionality of the unobserved heterogeneity components, a more formal analysis of these individual specific reactions may shed light on the model. To do so, we regress the individual specific reactions on all heterogeneity components of the model plus an intercept term. The results are in table 8B. We first regress the reactions on all components (column 1) and remove some gradually until we obtain a regression of reactions on the returns to schooling (column 4). The results indicate that the reaction to treatment decreases with the instantaneous utility of attending school (v_k^ξ) but increase with the individual specific returns to schooling (φ_{1k}), with the level of the wage intercept (v_k^w) and the employment rate intercept $(_k)$ over the relevant range. Clearly, homogeneity of treatment reaction across types is still strongly rejected. Interestingly, as we only include the returns to schooling (and therefore do not control for other heterogeneity components), we find that the reactions to the experiment are increasing with the returns to schooling (column 4).

At this stage, we may only conjecture that the positive correlation between individual specific reactions and returns to schooling suggests that it may be possible to reconcile our results with the common explanation advanced for the incidence of very high IV estimates. That is, IV estimates capture the returns to schooling of those who are induced to change their school attendance decision and that those specific individuals are indeed endowed with returns that are higher than the population average. However, we have only investigated one specific experiment. In theory, every potential experiment may entail a different selectivity pattern. To push the analysis

---

[19]Of course, this negative recation is only possible within a framework where forward looking agent base their schooling decisions on the return to schooling as well as the return to experience. Within a simpler framework, the relationship between schooling and the discount rate is unambiguous (Card 2001).

further, we now turn to an analysis of the IV estimates performed on a set of different counterfactual experiments.

6   Investigating the IV/OLS Puzzle

In the literature using 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 (see Imbens and Angrist (1994), Heckman (1997), Heckman and Vitlacyl (1998), and Card (2001)). In particular, it is well known that the IV estimate obtained in a context where slopes are heterogenous is only valid for a sub-population. In the treatment effect literature, the resulting estimate is sometimes interpreted as a Local Average Treatment Effect (LATE).[20]

As of now, obtaining consistent estimates of the population average return to schooling and a measure of dispersion, in a heterogenous treatment effect framework, is still a major challenge. Those estimating treatment effects using standard IV techniques often claim that their estimates are higher than the corresponding OLS estimates because the average returns of the sub-population affected by the instrument are higher than the population average returns. This may be true if, for instance, some individuals are credit constrained or have a particular high cost of attending school (see Card (2001)).[21] As far as we know, the veracity of this claim has never been verified empirically in a direct fashion. We now turn to this issue.

With an artificial experiment already performed (see previous section), our strategy is to compute an IV estimate on the simulated data and examine if the LATE differ substantially from the structural estimate of the population average and the OLS estimate. As different experiments may produce different reactions for a single individual, it is informative to perform additional experiments. To obtain a more diverse set of experiments, we simulated two additional experiments targeting college attendance or high

---

[20]More details may be found in Imbens and Angrist (1994).

[21]Note however that in the present context, it may also be explained by the possibility that some individuals are endowed with a particular high wage intercept.

school graduation; namely a decrease in the discount rate from grade 13 onward, as well as an increase in the utility of attending from grade 12 onward.

There is a key distinction between the interpretation that must be given to IV estimates set in a discrete (two-period) framework and estimates obtained within a dynamic model. In a simple two-period model, the individuals who would have decided to attend college without exposure to the instrument are obviously not affected by policy changes. In the literature, the discussion of the IV interpretation is often pitched within such a framework and, for this reason, the high estimates are imputed solely to those who would not have attended college without exposure to the instrument and, more precisely, those "who are at the margin" of choosing college enrolment. In a fully dynamic setting, the issue is more complicated. If one considers introducing a college attendance subsidy (paid over 4 years of college), the subsidy will not only affect those who would not have attended college but also some of those who would have entered college even without being exposed to the experiment by increasing their continuation (graduating) probabilities. It is therefore important to understand that the effects of a policy change are therefore not solely located within a sub-population of individuals who are close to be at the margin ex-ante.

The estimates resulting from the IV estimation exercises, along with their standard errors are found in Table 9. The IV estimates resulting from all three experiments are compatible with the usual findings reported in the literature. They all exceed the OLS estimate and the structural dynamic programming estimate (the population average). The IV estimate resulting from the first experiment (the overall decrease in the discount rate) is equal to 0.1211. For the second experiment (a decrease in the discount rate from grade 13 onward) and the third experiment (an increase in the utility of attending school from higher grade 12 onward) the estimates are equal to 0.0979 and 0.1096, respectively. It is also important to note that the IV estimates exceed the MTE substantially. This is another illustration of the fact that the IV (LATE) estimates are identified from a population which encompasses more than only those who are indifferent between college and high school graduation.

As shown in Section 5, these results are explained by the positive correlation between the returns to schooling and the individual specific reactions and are not solely attributable to those individuals who are at the margin, but also to those individuals who would achieve a higher grade level no mat-

24

ter what. This latter group of individuals is endowed with particularly high returns to schooling. As mentioned above, the IV estimates should be understood as a measure of the returns to schooling within the sub-population affected by the experiment. However, the density and the characteristics of this sub-population are not identified in this framework while they are identified within a structural dynamic programming framework.

As a final step, it is interesting to apply IV techniques to the actual data and compare these estimates to our structural estimate. To do so, we selected a set of variables available in the NLSY and that are sometimes used as an instrument in the literature. These variables are parents' education (father's and mother's education measured in years), number of siblings and an indicator recording whether or not the individual was raised in a nuclear family at age 14. Similar to most of the instruments used in the literature, each of them can be questioned. Indeed, a judgment on the validity of each possible instrument is left to the reader. However, we would like to stress that our objective is not to advocate IV as an estimation method but rather evaluate the discrepancy between OLS and IV estimates. For this, we need a set of variables that are representative of those typically chosen by empirical researchers.[22]

In Table 10, we report 4 different IV estimates along with their respective standard errors. Each of these estimate must be compared to the corresponding OLS estimate (0.0534). Clearly, all estimates (0.0787, 0.0973, 0.1015 and 0.0803) are above the corresponding OLS estimate. As such, they are compatible with the existence of heterogeneity in the returns. In the empirical literature, it is the prevalence of such high estimates that lead to serious questioning about the meaning of IV estimates.

These results are encouraging. They indicate that the structural dynamic programming model with multi-dimensional heterogeneity is capable of explaining the coexistence of relatively low returns to education (on average) with very high returns for some identified sub-populations. Our model is therefore capable to explain to well known OLS/IV puzzle.

---

[22]Carneiro and Heckman (2002) present a critical survey of several instruments used in the literature. They argue that most of the instruments used are invalid either because they are uncorrelated with education and/or because they are correlated with unobserved ability. Consequently, the findings that IV estimates generally exceed OLS may simply be due to the use of poor instruments.

7   Out-of-Sample Fit

In the structural literature, it is common to analyze the model performance using predicted frequencies and comparing these to actual choices. This type of "in-sample fit" is rarely accompanied by a discussion of out-of-sample fit. In this paper, and as indicated in the previous section, we showed that the structural dynamic programing model with a rich level of heterogeneity is capable of generating the very high returns found in the experimental literature, while being consistent with the very low returns to schooling found in the structural literature.[23] Indeed, Card ((2001) and (2002)) surveys more than 20 studies set in the experimental framework in which the return to schooling are found to exceed 10%. Most of them are actually obtained from US data sets other than the NLSY. Moreover, many studies cover periods which do not overlap with the period over which we draw our inference. Therefore, we view this large set of empirical results as overwhelming evidence in favor of the capacity of our model to obtain realistic estimates of the population distribution of the returns to schooling. To a large extent, the prevalence of high IV estimates reported in the literature provides us with a very high level of out-of-sample fit.

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 show that the estimates of the dynamic programming model with a rich heterogeneity specification, may be used to obtain estimates of most different measures of treatment effects proposed in the literature; the average treatment effects (ATE), the average treatment effects for the treated and the untreated (ATT/ATU), the marginal treatment effect (MTE) and, finally, the local average treatment effects (LATE).

---

[23]See Belzil (2005).

We find population average returns to schooling which are much below those reported in the existing literature. Our estimates are also lower than those obtained using standard OLS techniques. The return to experience upon entering the labor market (0.059) exceeds the average return to schooling in the population (0.043). The importance of selectivity based on individual specific returns to schooling has been illustrated by the difference between the average returns for those who have not attended college (0.0321) and those who attended college (0.0645). The structural parameters imply that those who are at the margin of entering college ex-ante are endowed with an average return equal to 0.0573.

Our results indicate that the reactions to counterfactual policy changes differ substantially across types and that, as a general rule, these reactions increase with the individual specific returns to schooling. This is true after conditioning on all other heterogeneity components as well as unconditionally. The importance of individual heterogeneity in the level of reactions to policy changes is well illustrated by the three counterfactual experiments that we implement. The IV (LATE) estimates are equal to 0.1211, 0.0979, 0.1096, and are comparable to those very high estimates often reported in the literature. They exceed both the OLS estimate and the population average of structural dynamic programming type specific estimates. The IV estimates resulting from all experiments are compatible with the usual findings reported in the literature. They all exceed the OLS and the structural dynamic programming estimate (the population average). These results are explained by the positive correlation between the returns to schooling and the individual specific reactions. Interestingly, the high IV estimates are not solely attributable to those individuals who are at the margin, but also to those individuals who would achieve a higher grade level in any event.

These results are encouraging. They indicate that the structural dynamic programming model with multi-dimensional heterogeneity is capable of explaining the coexistence of relatively low returns to education (on average) with very high returns for some identified sub-populations. As well, our model is capable to explain to well known OLS/IV puzzle.

References

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

[2] Belzil, Christian (2005) "The Return to Schooling and Experience and the Ability Bias in Structural Models: A Survey" forthcoming in *The European Economic Review.*

[3] Belzil, Christian and Hansen, Jörgen (2002a) "Unobserved Ability and the Return to Schooling" *Econometrica, 70, 575-91.*

[4] Belzil, Christian and Hansen, Jörgen (2002b) "A Structural Analysis of the Correlated Random Coefficient Wage Regression Model" IZA Working paper No. 512, IZA, Bonn, Germany.

[5] Belzil, Christian and Hansen, Jörgen (2003) "Structural Estimates of the Intergenerational Education Correlation" *Journal of Applied Econometrics, 18, 679-96.*

[6] Björklund, Anders and Moffitt, Robert (1987) "The Estimation of Wage Gains and Welfare Gains from Self-Selection Models" *Review of Economics and Statistics, 69, 42-49.*

[7] 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*, 262-333.

[8] Cameron, Stephen and Taber, Chris (2004) "Estimation of Educational Borrowing Constraints Using Returns to Schooling" *Journal of Political Economy, 112, 132-82.*

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

[10] Card, David (2002) "Estimating the Returns to Schooling: Some Progress on Problems" *Econometrica, 70, 575-591.*

[11] Carneiro, Pedro and Heckman, James "The Evidence on Credit Constraints in Post–secondary Schooling" *Economic Journal, 112, 705-34.*

[12] Carneiro, Pedro, Heckman, James and Vytlacil, Edward (2003) "Understanding what Instrumental Variable Estimate: Estimating Marginal and Average Returns to Education, Working Paper, University of Chicago.

[13] Eckstein, Zvi and Wolpin, Kenneth (1999) "Youth Employment and Academic Performance in High School" *Econometrica 67, 1295-1339.*

[14] Heckman, James (1997) "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources, 32 (3), 441-62.*

[15] Heckman, James and Vytlacil, Edward (1998) "Instrumental Variables Methods for the Correlated Random Coefficient Model", *Journal of Human Resources, 33, 974-87.*

[16] Heckman, James and Vytlacil, Edward (2004) "Structural Equations, Treatment Effects and Econometric Policy Evaluations" forthcoming in *Econometrica.*

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

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

[19] Magnac, Thierry and Thesmar, David (2001) "Identifying Dynamic Discrete Decision Processes" *Econometrica, 70, 801-16.*

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

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

[22] Roy, Andrew (1951) "Some thoughts on the Distribution of Earnings" *Oxford Economic Papers, 3, 135-46.*

[23] 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.

[24] Willis, Robert and Rosen, Sherwin (1979) "Education and Self-Selection", *Journal of Political Economy, 87, S7-S36.*

[25] Wooldridge, Jeffrey M. (1997) "On Two-Stage Least Squares estimation of the Average Treatment Effect in a Random Coefficient Model" *Economic Letters, 56, 129-33.*

[26] Wooldridge, Jeffrey M.(2003) "Further Results on Instrumental Variables Estimation of Average Treatment Effects in the Correlated Random Coefficient Model" *Economics Letters, 79, 185-91.*

Table 1
Structural Estimates: The Utility of Attending School and Labor Market Outcomes

Type	Utility of School Inter. υ_i^ξ	Wages Inter. υ_i^w	Educ. φ_{1i}	Employment Inter. κ_{0i}
1	-1.1511	1.4771	0.0547	0.2099
	(0.0659)	(0.0197)	(0.0052)	(0.0474)
2	-0.4853	1.4746	0.0210	-0.3872
	(0.0756)	(0.0214)	(0.0050)	(0.0468)
3	-1.4590	1.3306	0.0397	1.7001
	0.0802)	(0.0181)	(0.0032)	(0.0696)
4	-2.6933	0.8535	0.1278	1.6631
	(0.0835)	(0.0331)	0.0040)	(0.0658)
5	-2.9461	1.4746	0.0071	1.0965
	(0.0711)	(0.0170)	(0.0037)	(0.0484)
6	-1.4751	1.5807	0.0984	-0.5096
	(0.1117)	(0.0326)	(0.0041)	(0.0638)
7	-1.2973	1.3746	0.0032	2.1587
	(0.0413)	(0.0250)	(0.0028)	(0.0811)
8	-1.8215	1.2398	0.0072	0.4923
	(0.0367)	(0.0181)	(0.0031)	(0.0612)
Average	-1.5545	**1.3633**	**0.0428**	**0.6219**
Std. dev.	0.8466	**0.2050**	**0.0391**	**0.8562**

Note: Asymptotic standard errors are shown in parantheses. The interruption probability is around 7% per year and the log likelihood is -20,080.

31

Table 2
Remaining parameter estimates

	Estimate	std. err.
utility of school		
σ_ξ^2	0.8525	0.0500
ψ	-0.1284	0.0496
employment rate		
κ_1 (S_{it})	-0.3406	0.0073
κ_2 (exp)	-0.1776	0.0052
σ_e^2	1.7560	0.0127
Wages		
σ_w^2	0.3176	0.0025
φ_2 (exp)	0.0586	0.0028
φ_3(exp^2)	-0.0003	0.0001

Table 3
Mean Schooling and Type Probabilities

	Expected Schooling	Type Probabilities (p_k)	q_k (std. err.)
type 1	10.7	0.2181	0.7879 (0.0657)
type 2	9.2	0.2217	0.8043 (0.0755)
type 3	14.2	0.1053	0.0594 (0.0301)
type 4	14.3	0.1186	0.1783 (0.0610)
type 5	10.9	0.1348	0.3063 (0.0513)
type 6	9.2	0.0515	-0.6547 (0.0889)
type 7	15.2	0.0508	-0.6697 (0.0368)
type 8	10.0	0.0992	0.0000 (normalized)

Note: The type probabilities are computed using logistic transforms

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

Table 4
Ability and Taste Heterogeneity:
Type Specific Rankings

Rankings

	Schooling		Wages		Employment
	Predicted Schooling $E(S_i)$	Taste for Schooling v^ξ	Intercept term v^w	Return to Education φ_1	Intercept term κ_0
type 1	5	2	2	3	6
type 2	7	1	3	5	7
type 3	3	4	6	4	2
type 4	2	7	8	1	3
type 5	4	8	3	7	4
type 6	7	5	1	2	8
type 7	1	3	5	8	1
type 8	6	6	7	6	5

Table 5
The contributions of individual endowments
to the differences in schooling attainments

Heterogeneity Terms	v_i^ξ	v_i^w	φ_{1i}	v_i^w/φ_{1i}	$v_i^\xi/\varphi_{1i}/v_i^w$
R^2	0.28	0.39	0.07	0.40	0.46

Note: All $R^{2\prime}s$ are computed from OLS regressions of 70,000 simulated trajectories on the individual specific heterogeneity components.

Table 6
Model Fit: Actual vs Predicted Wages

log wages	Predicted (%)	Actual (%)
-∞-1.5	1.4%	12.9 %
1.5-2.0	14.0%	16.8%
2.0-2.5	37.5%	35.6%
2.5-3.0	33.7%	27.5%
3.0-3.5	12.2%	6.6 %
3.5-∞	1.3%	1.0%

Note: We also ran OLS on the simulated data and found the estimate obtained on simulated wages (0.0467) to be reasonably close to the actual OLS estimate computed on pooled data (0.0524).

Table 7
Population heterogeneity, Average Treatment Effects
and Marginal Treatment Effects

	ATU	ATT	MTE	ATE
Estimate	0.0321	0.0645	0.0563	0.0428
std. err.	(0.027)	(0.051)	(0.012)	(0.002)

The treatment effects are defined as

1. ATU=$E(\varphi_{1i} \mid S_i \leq 12)$

2. ATT=$E(\varphi_{1i} \mid S_i \geq 13)$

3. MTE=$E(\varphi_{1i} \mid 12 < E(S_i) < 13)$

4. ATE=$E(\varphi_{1i})$

Note: The ATT and the ATU are computed from 70,000 simulated schooling/wage histories. The MTE is constructed as an average return over these 3 types of individuals (type 3, type 4 and type 5) and is weighted by the appropriate type probabilities.

Table 8A
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	Predicted Schooling
Type 1	0.3 year	3	0.0547	10.7 years
Type 2	0.0 year	6	0.0210	9.2 year
Type 3	0.2 year	4	0.0397	14.2 year
Type 4	0.7 year	2	0.1278	14.3 years
Type 5	0.1 year	5	0.0071	10.9 year
Type 6	0.8 year	1	0.0984	10.9 year
Type 7	-0.1 year	6	0.0032	9.2 year
Type 8	-0.1 year	6	0.0072	15.2 year
Average	0.3 year	-	0.0428	10.0 year

Table 8B
The determinants of the individual specific reactions
to a counterfactual change in discount rate

Parameter
(st. error)

	(1)	(2)	(3)	(4)
$intercept$	-0.9517	-0.8276	-0.3145	-0.0700
	(0.0048)	(0.0034)	(0.0287)	(0.0045)
v_i^w	0.4978	0.4111	0.1629	-
	(0.0012)	(0.0212)	(0.0189)	
$\varphi_{1i} * 100$	0.0739	0.0717	0.0695	0.0643
	(0.0014)	(0.0009)	(0.0009)	(0.0001)
κ_{0i}	0.0397	-	-	-
	(0.0054)			
v_i^ξ	-0.0883	-0.1065	-	-
	(0.0045)	(0.0042)		
R^2	**0.0994**	**0.0987**	**0.0902**	**0.0893**

Note: The OLS regressions are computed from 70,000 simulated reactions to the experiment.

Table 9
Population heterogeneity and the OLS/IV Discrepancy

	Experiment 1	Experiment 2	Experiment 3
Method	IV (LATE)	IV (LATE)	IV(LATE)
Data	simulated	simulated	simulated
parameter	0.1211	0.0979	0.1096
st.error	(0.0133)	(0.0223)	(0.0071)
Sample	70000	700000	70000

Note: The experiments are defined as follows,

1. An overall decrease in discount rate (experiment 1).

2. A decrease in discount rate from grade 13 onward (experiment 2).

3. An increase in the utility of attending school in grade 12 (experiment 3).

Table 10
IV estimates computed from Actual instruments

Instruments	Parents' Education	Number of Siblings	Nuclear Family	All
Parameter	0.0782	0.0973	0.1015	0.0803
std. err.	0.0158	0.0421	0.0679	0.0153

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 1994 and bi-annually after 1994. In this paper, we restrict our sample to white males from the core random sample who were 14 to 16 years old in 1979. We record information on education, wages and on employment rates for each individual from the time the individual is 14 up to December 31, 1994.

Before any additional sample selections, we have 874 white males. We lost about 8% of the sample due to missing information regarding family income, about 10% due to missing information regarding parents' education, and finally about 3% because of missing information on AFQT scores. After these additional sample selections, the final sample consist of 682 white males.

Descriptive statistics on family background variables can be found in Table A1. The measure of family income is obtained using information on family income from the years 1978 and 1980. When data for both years are available, the average is used. The average value, measured in 1980 dollars, is $28,812. Parent's education, measured as highest grade completed, averages around 12 years for both the father and the mother. The average number of siblings is 2.8 and a majority of our sample members were living in an intact family with both biological parents present at the age of 14. Finally, the average value of AFQT scores is 49 while the average number of time periods we have information on the sample members is 14.4 (out of a maximum of 17). Overall, the sample averages of the observed characteristics are similar to those reported in previous papers relying on sample from the NLSY.

Educational attainment is measured as 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 may be right-censored for some individuals. Unlike many reduced-form studies which use proxies for post-schooling labor

---

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

market experience, we use actual labor market experience. Accumulated actual work experience is computed using the fraction of the year worked by a given individual. The availability of data on actual employment rates also allows use to estimate the employment security return to schooling.

The average schooling completed (by age 29) is 12.8 years. As described in Belzil and Hansen (2002a), 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.
family income	28,812	14,992
father's educ	12.4	3.2
mother's educ	12.1	2.3
# of siblings	2.8	1.8
prop. raised in urban areas	0.74	-
prop. raised in south	0.26	-
prop in nuclear family	0.81	-
AFQT	49.11	27.08
Numer of time periods	14.42	3.65