DISCUSSION PAPER SERIES

IZA

IZA DP No. 2650

Testing the Specification of the Mincer Wage Equation

Christian Belzil

March 2007

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Testing the Specification of the Mincer Wage Equation

Christian Belzil
Centre National de Recherche Scientifique,
CIRANO, CIREQ and IZA

Discussion Paper No. 2650
March 2007

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 2650
March 2007

ABSTRACT

Testing the Specification of the Mincer Wage Equation*

I perform the joint estimation of a reduced-form dynamic model of the transition from one grade level to the next, and a Mincer wage equation, using panel data taken from the NLSY. A very high degree of flexibility is achieved by approximating the distributions of idiosyncractic grade transition shocks and wage shocks with high dimensional normal mixtures. The model rejects all simplifying assumptions common in the empirical literature. In particular, the log wage regression is highly convex, even after conditioning on unobserved and observed skills. Skill heterogeneity is also found to be over-estimated when non-linearity is ignored. After conditioning on skill heterogeneity, I also find evidence of non-separability between the effect of schooling and experience (schooling has a positive effect on wage growth). Finally, the variance of the idiosyncratic wage shock is reduced by obtaining higher education.

JEL Classification: J2, J3

Keywords: Mincer regressions, heterogeneity, random coefficient models, returns to schooling, returns to experience

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
E-mail: belzil@gate.cnrs.fr

* I would like to thank James Heckman and two anonymous referees for useful comments and suggestions.

1 Introduction and objectives

The Mincer wage equation is one of the most widely used tools of empirical economics. Mincer wage equations have been applied to numerous areas of Labor Economics such as in the literature on measuring returns to education as well as in the literature on wage inequality. It is also used to investigate statistical discrimination, gender differences in wages and occupation (and sectoral) choices.[1]

This paper is driven by one major objective. It is to obtain estimates of all the components of the Mincer wage equation within an econometric specification in which i) schooling is endogenous, ii) the wage equation is estimated as flexibly as in the structural estimation literature, and (iii) the number of parametric/distributional assumptions is kept to a minimal level. The econometric model is based on two distinct components; a reduced-form dynamic model of schooling attainment based on the hazard specification of the transition from one grade level to the next with observed and unobserved heterogeneity and a non-linear Mincer wage equation model with observed and unobserved skill heterogeneity.

To meet this objective, I perform four main tasks. First, I obtain panel estimates of all the key components of the Mincer wage equation function in a context where i) skill heterogeneity affects the intercept term, the return to schooling and the return to experience, ii) the local return to schooling may vary with grade level (the return to college may be different than the return to grade school or high school), and where iii) returns to experience depend on accumulated schooling. Secondly, I perform statistical tests of these various hypotheses (skill heterogeneity, non-linearity, and separability) in order to shed light on the optimal specification of the celebrated Mincer wage equation function. Thirdly, I perform some variance decompositions of the individual specific intercepts and slopes in order to assess the relative importance of parents background variables, pure individual heterogeneity and accumulated schooling (for the returns to experience) in explaining skill heterogeneity in the labor market. Finally, in order to evaluate the reliability of the most popular model specifications found in the literature (obtained when various dimensions of the most general model specification are removed), I compare the estimates of the first and second moments of returns to schooling and experience obtained under various scenarios.

The main results are as follows. The model rejects all simplifying assumptions common in the empirical literature. I find that the degree of convexity of the wage equation, as measured by the difference in the local returns to schooling before and after high school graduation, is dependent on the allowance for skill heterogeneity. However, the log wage equation remains highly convex, even after conditioning on unobserved and observed skills. The convexity is acute and it is therefore not solely a reflection of omitted skill heterogeneity. Not surprisingly,

[1]For an historical perspective on the Mincer equation, see Heckman, Lochner and Todd (2005) or Belzil (2007).

skill heterogeneity is also found to be quite important, but I also find that ignoring non-linearity inflates the cross-sectional variance in the returns to schooling. After conditioning on skill heterogeneity, there is a positive correlation between accumulated schooling and individual specific returns to experience. This is consistent with the view that accumulated schooling may have a causal effect on wage growth. Finally, I find some evidence that the variance of the idiosyncratic wage shock is reduced by obtaining higher education.

The results reported here are in line with those found in the structural literature. The estimates of the returns to schooling, much lower than point estimates reported in the OLS/IV literature, seem to suggest that the discrepancy between structural estimates and OLS/IV estimates may well be explained by differences in the econometric specification of the wage equations, but not by the parametric assumptions required to achieve structural estimation.

The paper is structured as follows. In Section 2, I discuss some background literature. The empirical model is exposed in Section 3. Section 4 is devoted to the results of the statistical tests. The structural parameter estimates are discussed in Section 5 and the relative importance of skill heterogeneity and non-linearities is studied in Section 6. In Section 7, I investigate the importance of allowing for non-separability. The conclusion is in Section 8.

2 Background Literature

For a long time, empirical models have been based on the ad-hoc assumptions that individual differences in market skills can be captured in the intercept term of the wage equation function and that log wages vary linearly with schooling. The validity of these assumptions has however been seriously questioned in recent years and many economists have examined the stability of the stylized facts about age earnings profiles reported in Mincer (1974). Consequently, economists have started to pay particular attention to the introduction of heterogeneity in the slopes of the wage equation, to potential non-linearity (the convexity of the wage schooling relationship) and to the separability between education and experience.

With regards to skill heterogeneity, the random coefficient representation of the wage equation function has gained in popularity, along with the literature on estimating treatment effects.[2] At the same time, others have paid a particular attention to potential non-linearities explained by differences in local returns to

[2]The term "correlated random coefficient wage regression model" is often used to refer to the standard Mincerian wage regression model where all coefficients are individual specific. Recent papers devoted to specification and estimation issues surrounding a random coefficient model of the wage regression include Heckman and Vitlacyl (1998, 2005), Wooldridge (1997), and Angrist and Imbens (1994). Belzil and Hansen (2007) present a structural analysis of the correlated random coefficient wage regression model and show that all treatment effect parameters may be obtained within a structural framework.

the schooling across grade levels (Belzil and Hansen, 2002). [3] Furthermore, the recognition that post-schooling human capital investments should be treated as endogenous is likely to translate into new waves of empirical work which, among other things, should question the validity of the separability assumption (Rosenzweig and Wolpin, 2000).

While "skill heterogeneity" and "non-linearity" are not mutually exclusive, they are rarely confronted. This oversight might be a serious drawback. If the individuals who have higher market ability also have a comparative advantage in schooling (experience higher returns to schooling) and acquire more schooling, the convexity of the wage equation function might only reflect dynamic self-selection (merely a composition effect). That is, as we move toward higher levels of schooling, the local returns to schooling may turn out to be estimated from an increasingly large proportion of high ability workers. If so, allowing for cross-sectional heterogeneity in the slope parameter (s) of the wage equation might obviate the need for a flexible (non-linear) specification of the wage equation function and facilitate estimation. Equally, if the wage equation is truly convex (the returns increase with grade level), estimates of the returns to schooling obtained in a standard linear random coefficient framework might over-estimate the importance of cross-sectional heterogeneity.[4]

Knowing the relative importance of the non-linearity and the skill heterogeneity hypotheses is fundamental for those interested in estimating the returns to schooling. In the literature, it is customary to estimate the log wage equation function using Instrumental Variable (IV) techniques and interpret the estimates within a linear random coefficient framework. The linearity assumption is therefore crucial.[5] However, if the linear wage equation is not supported by the data and the form of the wage equation function is unknown, the estimation method is more complicated. Currently, the relative merits of both model specifications are unknown. A casual review of the recent literature would reveal that labor economists tend to favor the skill heterogeneity hypothesis. This preference is the result of ad-hoc assumptions. It is not founded on any empirical evidence.[6]

Similarly, the independence between education and the return to experience, typically illustrated by the fact that age earnings profiles are approximately parallel across broad education groups, is also being questioned (Heckman, Lochner

[3]Belzil and Hansen (2002) used a structural dynamic programming model to obtain flexible estimates of the wage regression function from the National Longitudinal Survey of Youth (NLSY) and found that a model with constant local returns is strongly rejected in favor of a highly convex log wage regression function composed of 8 segments. The average return over the entire range (around 4% per year) is found to be much lower than what is usually reported in the literature.

[4]Carneiro, Hansen and Heckman (2003) adress this issue within a factor structure.

[5]In general, the use of IV techniques requires separability between the instruments and the error term in the treatment equation and it also imposes monotonicity.

[6]For more details on the theoretical foundation of the linearity assumption, see Heckman, Lochner and Todd, 2005.

and Todd, 2005. This suggests that log wages equation may not be separable in education and experience and, in particular, that the return to experience may be affected by schooling. This would be the case, for instance, in post-schooling human capital investment model, as well as in various lifecycle incentive models where wages are upward sloping (Lazear,1999).

Finally, it should be noted that the literature is not only characterized by the diversity of applications and by differences at the level of the functional form and the stochastic specification, but also by a variety of estimation methods. While the vast majority of econometric estimates of the returns to schooling or experience are obtained in an OLS or an IV framework, estimates have also been obtained using structural dynamic programming techniques based on maximum likelihood methods (or their simulated counterparts). There is a surprising discrepancy between estimates obtained in a structural framework and those obtained in a standard OLS/IV framework. While OLS and IV estimates are typically high (estimates lying between 10% and 15% per year are often reported for the US), structural estimates (such as those reported in Keane and Wolpin, 1997 and Belzil and Hansen, 2002) are much lower.[7] These results are difficult to reconcile, as each estimation method is based on a large number of assumptions.[8]

The model is estimated using data from the National longitudinal survey of Youth (79-90). I restrict myself to this period because the resulting sample is virtually the same sample used by Belzil and Hansen (2002) and Keane and Wolpin (1997). Because of this, I can then compare returns to schooling and experience obtained from structural models with those obtained from a reduced-form approximation of the dynamic discrete choice. A brief description of the sample data is found in Appendix. The empirical likelihood function maximizes the joint probability of the observed schooling attainment and a particular wage history observed between 1979 and 1990. The estimation method is flexible. It is semi-parametric in spirit and allows for observed and unobserved heterogeneity in all dimensions. Each component of the wage equation (intercept term, returns to schooling and returns to experience) require 11 parameters (4 support points and 7 observable regressors). It also allows for a flexible estimation of the error shock of the grade transition model and the post schooling wage distribution by assuming that the errors are drawn from a mixture of 5 normal distributions. As far as I know, this is one of the most general Mincer wage equation function ever estimated.

[7]The reader will note that, strictly speaking, there is no such thing as "structural estimates of the returns to schooling". Structural estimation does not identify new parameters of the Mincer wage equation. However, I use the term "structural estimates" to refer to economic models where endogenous schooling is modeled through the solution of an intertemporal model, in which the return to schooling is in the information set of the agent.

[8]The difference between IV and structural estimates is surveyed in Belzil (2007).

3 A Reduced-form Dynamic Model of Schooling and Wages

The model is based on two items; a hazard function of grade completion and a wage equation model flexibly specified.

3.1 Schooling attainments

The econometric model used to deal with the endogeneity of schooling attainment is a hazard function model of grade transition. I generate the hazard function from an individual/grade specific index γ_{iS}^*, expressed as

$$\gamma_{iS}^* = \gamma_{0S} + X_i'\gamma_{1S} + \gamma_{2S} \cdot \theta_i^G + \varepsilon_{it}^S \tag{1}$$

where γ_{0S}, γ_{1S} and γ_{2S} are vectors of grade specific intercept and slopes to be estimated. Without loss of generality, I define the index, γ_{iS}^*, as the difference between utility of leaving school after completing grade S minus the utility of continuing in school beyond grade level S. The decision to stop is recorded in a variable $\gamma_{iS} = 1$ when $\gamma_{iS}^* > 0$ and $\gamma_{iS} = 0$ if not. The conditional probability of stopping school with grade level S (the hazard rate) is given by $F_{S=s}(\gamma_{i,S}^*)$ where $F_{S=s}(.)$ is a cumulative distribution function of ε_{it}. There are as many $F_{S=s}(.)'s$ as there are possible grade levels. The continuation probability is equal to one minus the hazard rate. The term θ_i^G represents an individual specific unobserved term affecting the propensity to acquire schooling. The vector X_i is composed of observable family characteristics; father's education, mother's education, an interaction term between father's and mother's schooling, household income, Armed Forces qualification tests (AFQT) scores, number of siblings and an indicator equal to one if the individual has been raised by both biological parents and 0 if not. Yearly household income is reported as of 1978 and measured in units of \$1,000. AFQT scores are corrected for the level of schooling at the time when the test was taken.[9] Note that θ_i^G is assumed to be orthogonal to X_i. This approach amounts to the estimation of a vector of grade level specific intercept terms for each type, along with the restriction that the distance between each type specific intercept (at one particular grade level) is the same at all different grade levels.

3.2 The Mincer Wage equation

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

[9]To do so, I regressed AFQT scores on schooling and kept the residual.

$$\log w_{it} = \varphi_i^w + \varphi_i(S_i, Exper_{it}) + \varepsilon_{it}^w \tag{2}$$

where S_i denotes schooling and $Exper_{it}$ is accumulated experience at date t. I use actual experience as opposed to potential experience and assume that it is exogenous (see Appendix 1). I assume that ε_{it}^w has density $f_{S=s}^w(.)$. In order to estimate the model, I choose a tractable form for $\varphi_i(S_i, Exper_{it})$, which is

$$\varphi_i(S_i, Exper_{it}) = \varphi_i^S(S_i) + \varphi_i^E(S_i) \cdot Exper_{it} \tag{3}$$

with

- $\varphi_i^S(.) = \varphi_{i1} \cdot S_i + \delta_2 \cdot S_{ic}$
  - where $\varphi_{i1} = \exp(X_i'\beta^s + \theta_i^s)$
  - where $S_{ic} = S_i - 12$ if $S_i > 12$ and $S_{ic} = 0$ if $S_i \leq 12$. I choose 12 years as a threshold in order to capture the highschool-college wage premium.
- $\varphi_i^E = \exp(X_i'\beta^E + \tau_1 \cdot S_i + \tau_2 \cdot S_{ic} + \theta_i^E)$
- $\varphi_i^w = X_i'\beta^w + \theta_i^w$

I assume that $(\theta_i^G, \theta_i^s, \theta_i^E, \theta_i^w)$ are jointly distributed with CDF $H(.)$. In order to approximate $H(.)$ as accurately as possible, I assume that there are 4 types of individuals. Each type is therefore endowed with a vector $(\theta_i^G, \theta_i^s, \theta_i^E, \theta_i^w)$ for $k = 1, 2...4$. The probability of belonging to type k, p_k, are estimated using logistic transforms.

$$p_k = \frac{\exp(q_k^0)}{\sum_{j=1}^4 \exp(q_j^0)}$$

where the $q_j^{0\prime}s$ are parameters to be estimated and with the restriction that $q_4^0 = 0$.

- $F_{S=s}(.)$ is approximated with a mean-mixture of 5 normal random variables; that is

$$F_{S=s}(.) = \sum_{m=1}^{M} P_m^*(s) \cdot \Phi(\mu_m(s), \sigma_m)$$

where $\Phi(\mu_m, \sigma_m)$ denotes the normal cdf, $P_{m(s)}^*$ are the mixing probabilities, and where $\sigma_m = 1$ for $m = 1, 2, ..M$. Because I allow for 4 different (type specific) intercepts, I impose the following identification conditions; $\mu_m(s) = 0$ for one m. This is true for all possible s.

- $f^w_{S=s}(.)$ is approximated with a mixture of 5 unrestricted normal densities; that is

$$f^w_{S=s}(\varepsilon^w_{it}) = \sum_{m=1}^{M} Q^*_m \cdot \phi(\mu^w_m, \sigma^w_m)$$

where $\phi(\mu_m, \sigma_m)$ denoted the standard normal density, and where $Q^*_{m(s)}$ are the mixing probabilities. For identification purposes, I impose an ordering condition and I also impose $\mu_m(s) = 0$ for one m.

Altogether, the definitions of $\varphi^S_i(.), \varphi^E_i$ and φ^w_i allow for skill heterogeneity, non-linearities in the return to schooling (with two levels) and for a causal effect of accumulated schooling on the return to experience.[10] The positivity of $\varphi_{i1}$and φ^E_i are imposed in order to eliminate the possibility of unrealistic values for predicted wages or for the returns to schooling and experience. Note that I focus on linear returns to experience because the model is fit on a sample of young workers and wages are observed over a period over which the concavity of earnings profile has most likely not set in yet. The allowance for a possible correlation between θ^G_i and labor market skill heterogeneity $(\theta^s_i, \theta^E_i, \theta^w_i)$ will capture any endogeneity in schooling which may persists even after conditioning on X.

An inspection of equation (3) reveals that, in this particular framework, the returns to schooling vary with experience (education causes wage growth). For a given number of years of experience, the marginal effect of a year of schooling is given by

$$\frac{\delta \log w_{it}}{\delta S_i} = \frac{\delta \varphi^S_i(.)}{\delta S_i} + \frac{\delta \varphi^E_i(.)}{\delta S_i} \cdot Exper_{it} \tag{4}$$

Focussing on the marginal effect of post-high school training, we get that

$$\frac{\delta \log w_{it}}{\delta S_i} = \varphi_{i1} + \delta_2 + [\varphi^E_i(S_i) \cdot (\tau_1 + \tau_2) \cdot Exper_{it}] \tag{5}$$

In the literature, it is customary to assume that $(\tau_1 = \tau_2 = 0)$, so that there is no distinction between returns to schooling measured at entrance in the labor market and the returns measured several years beyond school completion. In the present model, $\varphi^S_i(S_i) + \delta_2$ is a measure of the marginal effect of schooling on wages only at entrance in the labor market (when $Exper_{it} = 0$). The growing pattern of the returns to schooling will be illustrated in Section 6.

[10] Another type of non-separability (ignored in this paper) could arise if the returns to schooling decline with experience (or age) because of depreciation.

3.3 The Likelihood Function

The likelihood function is the joint probability of observing a level of schooling attainment, S_i, and a particular wage history $(w_{i1}...w_{iT})$. Given type k, and dropping the i subscript, the likelihood for one observation, $L(k)$, has two components (the probability of having continued in school until S years of schooling is achieved and the density of observed wages until 1990). More specifically, L_k is equal to

$$L(k) = \Pi_{j=1}^{S-1}(1-\{\sum_{m=1}^{M} P_m^*(j)\cdot\Phi(\frac{\gamma_{kj}^*-\mu_m(j)}{\sigma_m})\})\cdot\sum_{m=1}^{M} P_m^*(S)\cdot\Phi(\frac{\gamma_{kS}^*-\mu_m(S)}{\sigma_m})\cdot$$
$$\Pi_{t=1}^{T}[\sum_{m=1}^{M} Q_m^*\cdot\phi(\frac{w_t-\varphi_k^w-\varphi_k(S,Exper_t)-\mu_m^w}{\sigma_m^w})] \quad (6)$$

The total log likelihood function to be maximized is

$$\log L = \log\sum_{k=1}^{K} p_k\cdot L(k) \quad (7)$$

where each p_k represents the population proportion of type k.[11]

3.4 Identification

In order to understand how the model is identified, it is informative to consider the literature on estimating hazard functions, as well as from results on estimating discrete choices with mixtures of normals. To estimate the model, I choose to approximate the error shock by a mixture of normal distributions. As shown in Geweke and Keane (1995), a mixture of normals is able to approximate a wide range of possible distribution in the context of a binary discrete choice model (provided some scaling conditions). It significantly outperform standard probit or logit models.

At the same time, and consistent with panel data models, the repeated observations on labor market wages allow me to identify the person specific intercept and slopes. Again, using mixture of normals (with no restrictions) allows me to approximate the error shock as flexibly as possible. Because both the distribution of wage shocks and error shocks generating the sequence of discrete choices are estimated flexibly, and because I interpret the model as a flexible approximation to a dynamic discrete choice model, I estimate the joint likelihood without any exclusion restriction (a feature of most structural models).

Note that the grade transition model set in (1) is a special case of the reduced-form discrete choice model analyzed in Heckman and Navarro (2006). They prove

[11]Estimation is performed using Fortran-based statistical routines, although it could also easily be performed with popular canned programs.

non-parametric identification of several classes of dynamic discrete choice models (including discrete hazard functions) to which they append outcome equations. Unlike Heckman and Navarro (2006), I do not impose a curvature condition on the latent utility equation. Instead, I rely solely on the flexibility of the normal mixture specification to obtain parametric identification. Intuitively, identification may be more easily understood by noting that the joint likelihood uses not only the first moment of the wage equation, but also higher moments.

To summarize, even though the allowance for normal mixtures allows for a high degree of flexibility, it is important to see that the identification of the model is still parametric. IV models, on the other hand, are identified primarily from orthogonality conditions which follow directly from ad hoc assumptions regarding the effect of some policy shocks on the error term of the outcome equation. However, as IV estimates are obtained in a context where only the first moment of the outcome equation is considered, their identification is semi-parametric.[12]

[12] As discussed in Keane (2006) and Belzil (2007), the distinction between the IV approach and the structural approach may be coined in terms of a trade off between behavioral and statistical assumptions. Without knowing the true data generating process, it is not possible to determine which approach to estimation is more flexible.

4 Searching for the Best Specification

The first step in estimation is the choice of the number of mixture components (m) and the number of groups (K). In order to choose the number of mixtures (M=5), I experienced with two and three and noticed very little changes at the level of the estimates of returns to schooling and experience. In order to implement the model, I have initially investigated a version with observed heterogeneity and gradually included unobserved types. Various experiments have indicated that it is not necessary to go beyond 4 types.[13] It should be noted that the 7 regressors representing family background are highly correlated. As a consequence, I treated the set of regressors as a single block which can proxy skill heterogeneity and chose not to remove the variables that may turn out to be insignificant in one of the components of the wage equation.

As a first step, I estimated the most general model specification and re–estimated several restricted versions that allowed me to perform likelihood ratio tests. There are 3 natural hypotheses of interest. The first one is that the effect of schooling on log wages is linear ($\delta_2 = 0$). The second hypothesis is that, conditional on unobserved heterogeneity, returns to experience are unaffected by accumulated schooling ($\tau_1 = \tau_2 = 0$). The third hypothesis is that skill heterogeneity is accounted for in the wage intercept and that a random coefficient specification is not required. This boils down to imposing $\beta^s = \beta^E = 0$, $\theta_1^s = \theta_2^s = ...\theta_k^s$, and $\theta_1^E = \theta_2^E = ...\theta_k^E$ (the "classical representation" of the Mincer wage equation). This test hinges on the assumption that there is a fixed (known) number of types.[14]

A summary of the likelihood ratio tests is found in Table 1 below. As is clear from the test statistics reported in Table 1, all three hypotheses are strongly rejected at the 1% level and, as a consequence, the optimal specification requires non-linearities in schooling, dependence between the returns to experience and accumulated schooling, and skill heterogeneity in the slopes. The evidence is overwhelming and does not require further discussion. This specification is now the baseline model which can be used to investigate several issues, which are addressed below.

[13]Indeed, I tried with 6 types but it turned out that the parameter estimates and the correlation estimates were practically not affected by the decision to go to 4 types. This is most likely explained by the relatively large number of observed regressors already included.

[14]Considering the number of types as fixed is relatively standard in the empirical literature where the estimation method consists of a relatively complicated mixed likelihood function. Aside from the case of a single spell duration model, non-parametric estimation of K is rarely achieved in the empirical literature (see Heckman and Singer, 1984).

5 The Parameter Estimates

In what follows, I discuss the parameter estimates and, in particular those pertaining to the wage returns to schooling and experience. The entire set of parameter estimates for the most general model specification is found in a sequence of tables ranging from Table 2A to Table 2F. Because there is a very large number of parameters required to approximate the distribution of the random shocks, and because these same parameters are relatively difficult to interpret, I report the resulting means, variances, and skewness coefficients for both the grade transition (Table 2G) and the wage distribution (Table 2H). The first and second moments of returns to schooling and experience are found in Table 3A. A Variance decomposition is documented in Table 4.

5.1 The Effects of Parents Background Variables

Parents background variables affect grade transition as well as wages. As mentioned earlier, these variables are highly collinear. The objective is to treat them as a single block of variables that are used as proxies for individual endowments. Indeed, their relative explanatory power with respect to grade transition and to wages will be discussed below (in the variance decomposition section).

5.1.1 Parents Background Variables and Schooling Attainments

The estimates of the effects of parents background variables on the hazard rate are found in the first column of Table 2A. After taking into account the interaction term between mother's and father's schooling, the estimates indicate that the school continuation probability increases with parents' schooling.[15] The parameter estimates also imply that schooling attainments will increase with household income, AFQT scores and decrease with the number of siblings. Those raised with both biological parents also tend to leave school later. These results are consistent with what has been reported in Belzil and Hansen (2002), Eckstein and Wolpin (1999) and Cameron and Heckman (1998 and 2000). Similar results are also present in numerous other studies. They do not require more discussion.

5.1.2 Parents Background Variables and Wages

Overall, the level of significance of the family background variables in the wage equation (reported in Table 2B) is somewhat lower than what was found in the grade transition equation. Notwithstanding this, there is evidence that most variables associated with higher schooling attainments (lower hazards) are also

[15] This may be seen after noting that, given the range of father's and mother's schooling, the negative effect of the interaction term (found at each possible grade level) will dominate when evaluating the marginal effects for both mother's schooling and father's schooling.

associated with higher returns to schooling, higher returns to experience and higher wage intercepts. This conclusion is reached after the examination of the effects of parents background variables and after taking into account the interaction terms. In particular, the returns to schooling and experience increase with father's schooling and mother's schooling. Both the intercept term and the returns to experience also increase with family income and AFQT scores. However, the returns to schooling appear to be decreasing with both parents income and AFQT scores, although the effects are relatively small.

5.2 Non-Linearity

The results found in Table 2C show strong evidence in favor of the convexity of the wage-schooling relationship. Without loss of generality, returns to education are measured upon entrance in the labor market (when $Exper_{it} = 0$). The parameter estimate for δ_2 (equal to 0.0406), along with the estimates for β^s (Table 2B), imply an average return to schooling equal to 0.0403 per year of schooling prior to high school graduation and 0.0804 in college (Table 3A). This is consistent with evidence presented in Belzil and Hansen (2002) and seems to indicate that the non-linear (convex) shape of the wage schooling relationship is acute and, furthermore, not a reflection of omitted skill heterogeneity. The standard deviation measures cross-sectional dispersion across types and across regressors). For a given type, and given some regressors, the degree of non-linearity is highly significant. This issue will be addressed in Section 6.

Aside from convexity, it should also be noted that, when compared to the estimates of the returns to schooling reported in the IV literature, these estimates are small.[16] However, they are comparable with the relatively lower estimates obtained in the structural literature.[17]

5.3 Non-Separability: The Effect of Schooling on Wage Growth

After conditioning on skill heterogeneity, and taking into account the endogeneity of schooling, the estimate for τ_1 and τ_2 (found in Table 2C) indicate that there is a positive correlation between accumulated schooling and individual specific returns to experience. However, this positive correlation is mostly explained by schooling acquired beyond high school graduation. This is illustrated by the relatively small value of the estimate for τ_1 (0.0033) and the much larger value for τ_2 (0.0573). This is consistent with the view that accumulated schooling

[16]The OLS estimate fluctuates depending on which year of the panel is chosen. However, for those periods considered, it averages around 10% per year. Indeed, the pooled OLS estimate is equal to 9.9%.

[17]As pointed out in Belzil (2006), IV estimates are typically obtained in a framework where accumulated experience is ignored and implicitly included in the error term.

may have a causal effect on wage growth. Obviously, this result may raise several economic interpretations. In a competitive market setting, this may arise if higher education reduces the costs of learning new skills (say on-the-job training) and may therefore stimulate post-schooling human capital accumulation. In a non-competitive framework with search frictions, higher wage growth could arise if, for instance, offer arrival rates depend on higher education status. Finally, if higher education is used as a signal in order to promote workers, wage growth would also be higher for more educated workers. Without further structure, it is impossible to say more.[18]

5.4 Unobserved Heterogeneity

The classical ability bias hypothesis is usually discussed in the context of an additive heterogeneity term affecting the wage equation. It arises when the wage intercept is positively correlated with schooling attainment.[19] In this model, the issue is complicated by the high dimensionality of the heterogeneity vector. To grasp the importance of heterogeneity, I also report the correlation between each component in Table 3B.

The negative correlations between θ^G and θ^S (-0.42) and between θ^G and θ^E (–0.28) reveal that those types who will tend to experience higher schooling attainments (lower hazard rates) will also experience higher returns to schooling and experience. This is a form of ability bias that is explained by self-selection based on individual specific slopes. Indeed, the positive correlation between the wage intercept (θ^W) and the grade transition equation (θ^G), equal to 0.92, appear to be the consequence of using a more flexible heterogeneity specification.

5.5 Heteroskedasticity and Skewness

Because I assume a separate distribution for the random shocks, for each grade level, the results are therefore obtained under arbitrary form of heteroskedasticity, skewness or kurtosis. Because of the relative complexity of the formulas that link each parameter (the $P_m^{*\prime}s$, the $\mu_m s$ and the $\sigma_m' s$) to the resulting moments (variance, skewness and kurtosis), it is difficult to perform a formal test of equality of these moments across schooling levels.

In view of the recent literature aimed at distinguishing ex-ante risk from unobserved heterogeneity, it may be particularly interesting to investigate how the variance of log wages behave across grade levels. The sequence of grade level specific variances is found in Table 2H (column 2). The variance of the stochastic component of log wages fluctuates around 0.30. When considering

[18] Indeed, as far as I know, the relative importance of human capital, search frictions and incentive provisions in explaining life cycle earnings is not known. It is much beyond the scope of this paper.

[19] For more details, see Belzil and Hansen (2002).

the average variances below (and including) grade 12, which is around 0.41 and above (including) grade 13 (around 0.32), we see that higher education may reduce exposure to income fluctuations. Obviously, as the model is not fully structural (and not based on a rational expectation assumption), it is impossible to say whether this difference in heteroskedasticity is used as an ex-ante input.

Finally, the degree of skewness resulting from the distributions appears quite mild. Negative skewness appears as frequent as positive skewness, but furthermore, the degree of skewness rarely exceeds 0 by a large amount.

5.6 Variance Decompositions

Some variance decompositions are found in Table 4. These may be used to infer the relative importance of parents background variables, unobserved skills and schooling (for the returns to experience) in explaining skill heterogeneity. The main findings seem to indicate that, while modeling wage equations in a context where the coefficients are allowed to be correlated with observed characteristics is important, skill heterogeneity is captured mostly through unobserved skills. More precisely,

- Only 9% of the cross sectional variations in returns to schooling is explained by parents background variables while 91 % is explained by unobserved skills
- 24% of the cross-sectional variations in the returns to experience are explained by parents background variables while 60% are explained by unobserved skills. Interestingly, accumulated schooling explains 18% of the returns to experience.
- 38% of the cross sectional variations in the wage intercept are explained by parents background variables while 62% are explained by unobserved skills.

6 Assessing the Relative Importance of Heterogeneity and Non-linearity.

At this stage, it is natural to investigate the consequence of ignoring either skill heterogeneity or non-linearity on the accuracy of the estimates of the returns to schooling. After all, most estimates published in the literature (based on IV methods) are based on cross-section data and on model specifications where the wage equations are assumed to be linear in schooling. While determining the degree of convexity might appear as a pure statistical issue at first glance, it is not really so. As discussed in Belzil 2006, the college/high-school wage premium may be easily explained in presence of high psychic costs. Determining the importance of non-linearities may therefore be a key step in evaluating the importance of psychic costs in college attendance decisions.[20]

The reliability of various model specifications may be investigated by comparing estimates of the returns to schooling upon labor market entrance obtained when various dimensions of the most general model specification are removed. In Table 5, I perform such comparisons. I also report estimates of the returns to schooling and experience in the case where skill heterogeneity and non-linearity are omitted (in column 4) and compare them to the estimates already reported. For the specifications with skill heterogeneity, I report the mean return to schooling as well as the standard deviation.

The results indicate that, to a certain extent, the degree of convexity of the wage schooling relationship is affected by the omission of skill heterogeneity. The difference between the returns in high school and in college, of the order of 4.9 percentage points in a flexible model (column 3), which allows for both skill heterogeneity and non-linearities, is now increased to 6.1 percentage point when skill heterogeneity is not controlled for (column 2). This result illustrates the degree of importance of dynamic self-selection.

However, as indicated by the rejection of the linear model, a fair degree of convexity persists. At the same time, the consequences of ignoring non-linearity are also quite spectacular. The estimate for the population average return to schooling in a linear model, which is around 6% per year (column 1), seriously over estimates the return to high school training (averaging 0.0376) and under estimates the return to post high school training (0.0864). It is important to remember that the relatively high degree of dispersion across individuals (types or observable regressors) may give the impression that the difference between the return to high-school and post high school training are insignificant, but that they are significant for each individual.

Ignoring both nonlinearity and skill heterogeneity (including in the intercept term), raises the return to schooling to 9% (column 4) and creates the statistical illusion that log wages are increased by more than 9% per year, regardless of the level of schooling.

Finally, another consequence of ignoring non-linearity is the exaggeration of the importance of skill heterogeneity. This is illustrated by the increase in the standard deviation of the returns to schooling from 0.0266 (when non-linearity is accounted for) to 0.0359 (when it is ignored).

In short, the results indicate that both non-linearity and heterogeneity are important and, perhaps more importantly, that ignoring either of those aspects may have serious consequences.

[20] This is issue is at the center of most structural models of schooling decisions. See Belzil (2006), for a survey.

7 The Importance of Non-Separability

Among all particular dimensions that I have examined, the issue of separability of log wages in education and returns to experience may be the most interesting from an economic standpoint. While non-linearity and heterogeneity may be seen as "statistical" issues, the absence of separability suggests the relevance of modeling wage growth. As stated earlier, wage growth is not only a key feature of human capital models but is also central to incentive models as well as search theory. Indeed, in the literature on labor market incentives and personnel economics, wage growth is also related to firm payment mechanisms (promotions, tournaments and various delayed payment schemes).

At this stage, two issues naturally arise. First, if schooling affects wage growth (given unobserved skills), the returns to schooling must be redefined so to incorporate the fact that schooling facilitates access to high wage growth. The returns to schooling, defined for the early years of labor market experience, are found in Table 6. As the returns depend on heterogeneity and on schooling level itself, the value is average over the values of X_i and realized schooling (from grade 13 onward). Despite the seemingly small estimate for τ_2, which was found to be equal to 0.0437, the return to post high-school training appears to rise relatively significantly. It goes from 0.08 (at entrance in the labor market) to 0.11 after 8 years of experience.

A second issue relates to the effect of assuming separability at the estimation level. To illustrate this, I re-estimated a conventional form of the equation (setting τ_1 and τ_2 to 0), and re-evaluated the returns to schooling upon entrance in the labor market and experience. The results are in Table 7. Contrary to intuition, I find that imposing separability does not affect much the return to post high-school training. In this restricted version, the population average return to college training is 0.0876. However, the return to high school training is largely inflated by imposing separability. Its average, now equal to 0.0498, is almost 50% higher than in the non-separable model (0.0498/0.0376). This is a severe over-estimation which, as far as I know, is practically never discussed in the literature. Although non-linearity and skill heterogeneity have been investigated before and will likely be investigated by researchers in the future, the cause of non-separability deserves some more attention. Modeling the channels by which schooling affects wage growth (training opportunities, promotions,..etc.) appears to be most appropriate.

8 Conclusion

In this paper, I present econometric estimates of the celebrated Mincer wage equation obtained with a degree of flexibility which, as far as I know, is virtually never achieved. The data reject all simplifying assumptions common in the empirical literature. I find that the degree of convexity of the wage equation, as measured by the difference in the local returns to schooling before and after high school graduation, is dependent on the allowance for skill heterogeneity. However, the log wage equation remains highly convex, even after conditioning on unobserved and observed skills. The convexity is acute and is therefore not solely a reflection of omitted skill heterogeneity. Not surprisingly, skill heterogeneity is also found to be quite important. After conditioning on skill heterogeneity, there is a positive correlation between accumulated schooling and the individual specific returns to experience. Standard models based on the separability assumption have two major defects. First, they ignore the positive benefit of education on future wage growth. Secondly, they appear to over-estimate the returns to high-school education by a significant margin (as much as 15%). Finally, I find some evidence that the variance of the idiosyncratic wage shock is reduced by obtaining higher education.

Overall, the results presented therein are much more in line with those reported in the structural literature than in the OLS/IV literature. For instance, the population average return to college education upon entrance in the labor market, around 8% per year, is much inferior to IV estimates often exceeding 15%. In the applied labor economics literature, the instrumental (IV) approach is often thought to be the most flexible because it requires less parametric assumptions than the structural approach.[21] To the extent that the estimation method suggested in the paper is a relatively flexible model of schooling decisions that does not require to model per-period utilities and subjective beliefs for each component of the decision process, there is no obvious reason to believe that the low returns reported in the structural literature are an artifact of the structural approach.

The discrepancy between structural estimates and reduced-form estimates has been noticed relatively recently. As of now, it is certainly not well understood. It remains an open question to see if the differences between structural estimates and OLS/IV estimates are attributable to differences in functional form of the Mincer equation or to the identifying restrictions. Only further investigation will enable applied econometricians to fully understand these marked differences.

[21] Surrounding issues are discussed in Keane (2007) and Belzil (2007).

References

[1] Belzil, Christian (2007), "The Return to Schooling in Structural Dynamic Models: A Survey, *European Economic Review (in press)*

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

[3] Belzil, Christian and Hansen, Jörgen (2003) "Structural Estimates of the Intergenerational Education Correlation", *Journal of Applied Econometrics, vol18: 679-696.*

[4] Belzil, Christian and Hansen, Jörgen (2007) "A Structural Analysis of the Correlated Random Coefficient Wage Regression Model, *Journal of Econometrics (in press).*

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

[7] Carneiro, Pedro, Hansen, Karsten and James Heckman (2003) "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and measurement of the Effects of Uncertainty on College Choice", *International Economic Review, 44 (2), 361-422.*

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

[9] Geweke, John and Michael, P. Keane (1995) "Mixture of Normals Probit Models" Federal Reserve Bank of Minneapolis, Staff Report # 237.

[10] Heckman, James, Lochner, Lance and Petra Todd (2005) "Earnings Functions, Rates of Return, and Treatment Effects: The Mincer Equation and Beyond, Working Paper 11544, NBER.

[11] Heckman, James and Salvador Navarro (2006) "Dynamic Discrete Choice and Dynamic Treatment Effects" forthcoming in *Journal of Econometrics.*

[12] Heckman, James and Singer, Burton (1984) “A Method for Minimizing the impact of distributional Assumptions in Econometric models for Duration Data, *Econometrica, 52 (2):* pp 271-320.

[13] Heckman, James and E. Vytlacil (2005) “Structural Equations, Treatment Effects and Econometric Policy Evaluations. *Econometrica 73(3), 669-738.*

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

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

[16] Keane, Michael (2006) “Structural vs. Atheoretic Approaches to Econometrics, forthcoming in *Journal of Econometrics*

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

[18] Lazear, Edward (1999), “Personnel Economics: Past Lessons and Future Directions” Journal of Labor Economics” vol 17, no 2, pp 199-236 .

[19] Magnac, T. and D. Thesmar (2001): “Identifying Dynamic Discrete Decision Processes,” *Econometrica*

[20] Mincer, Jacob (1997) “Changes in Wage Inequality, *Research in Labor Economics 16, 1-18.*

[21] Mincer, Jacob (1974) “Education and Earnings, Columbia University

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

[23] Taber, Christopher (2001) “The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability”, *Review of Economic Studies.*

[24] Staiger, Douglas and James H. Stock (1997),“Instrumental Variables Regression with Weak Instruments” *Econometrica, 65*: 557-586.

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

Table 1
Testing for skill heterogeneity, non-linearities and heteroskedasticity

Null Hypothesis	Lik. Ratio Statistics	number of restrictions	critical value at 1% level
Linear returns to schooling	15.7	1	6.6
Effect of schooling on return to experience	11.0	2	9.2
Homogenous returns schooling/experience	260.3	14	29.1

Table 2A
The Effects of Parents Background Variables on grade transition
(asymptotic t-ratios)

	Grade level					
	$\gamma_{1,6}$	$\gamma_{1,7}$	$\gamma_{1,8}$	$\gamma_{1,9}$	$\gamma_{1,10}$	$\gamma_{1,11}$
Family background variables						
father's educ	0.0824	0.0964	0.1003	0.1225	0.1267	0.1375
	(4.25)	(6.04)	(6.62)	(6.00)	(5.63)	(5.03)
mother's educ	0.0925	0.1036	0.1523	0.1628	0.1552	0.1646
	(9.24)	(9.79)	(8.38)	(6.38)	(6.94)	(7.06)
father's ed.*mother's ed	-0.0026	-0.0256	-0.0045	-0.0103	-0.0203	-0.0181
	(10.04)	(10.01)	(9.38)	(7.48)	(8.86)	(8.01)
fam. Income	-0.0053	-0.0091	-0.0034	-0.0048	-0.0064	-0.0071
	(3.47)	(5.49)	(6.74)	(5.95)	(5.44)	(5.93)
AFQT scores	-0.2534	-0.2634	-0.2758	-0.4002	-0.3001	-0.3336
	(10.46)	(19.33)	(5.89)	(5.28)	(5.02)	(4.95)
siblings	0.0976	0.0798	0.1005	0.1226	0.0927	0.1033
	(2.78)	(5.21)	(6.39)	(8.04)	(7.03)	(6.44)
nuclear family	-0.0423	-0.1000	-0.0987	-0.1103	-0.1056	-0.1143
	(2.00)	(1.94)	(2.00)	(1.38)	(1.79)	(1.63)

Table 2A (continued)-
The Effects of Parents Background Variables on grade transition
(asymptotic t-ratios)

	Grade level					
parameters	$\gamma_{1,12}$	$\gamma_{1,13}$	$\gamma_{1,14}$	γ_{15}	γ_{16}	$\gamma_{17-more}$
Family background variables						
father's educ	0.1325	0.1402	0.1522	0.1463	0.1823	0.1620
	(7.04)	(6.65)	(5.96)	(5.68)	(6.74)	(6.94)
mother's educ	0.1616	0.1487	0.1287	0.1302	0.1723	0.1729
	(10.79)	(8.48)	(9.38)	(9.86)	(10.04)	(9.46)
father's ed.*mother's ed	-0.0131	-0.0200	-0.0108	-0.0120	-0.0204	-0.0145
	(12.01)	(12.05)	(13,28)	(14.03)	(14.58)	(12.58)
fam. Income	-0.0062	-0.0053	-0.0048	-0.0050	-0.0040	-0.0063
	(4.69)	(4.12)	(5.29)	(5.94)	(6.38)	(4.86)
AFQT scores	-0.3085	-0.2854	-0.2389	-0.2056	-0.2927	-0.3198
	(21.35)	(16.83)	(20.44)	(17.49)	(18.38)	(19.93)
siblings	0.1003	0.1009	0.0996	0.1058	0.0899	0.1115
	(6.21)	(4.94)	(5.24)	(5.29)	(5.39)	(6.03)
nuclear family	-0.1223	-0.1337	-0.1087	-0.1124	-0.1046	-0.1196
	(1.54)	(1.59)	(2.38)	(2.04)	(2.05)	(1.86)

Table 2B-
The Effects of Parents Background Variables in the Wage equation
(asymptotic t-ratios)

	Wage equation		
parameters	intercept term β^w	return to schooling β^S	return to experience β^E
Family background variables			
father's educ	0.0009	0.0818	-0.0405
	(0.29)	(4.31)	(2.76)
mother's educ	-0.0072	0.0428	-0.0357
	(-1.04)	(2.59)	(3.11)
father's ed.*mother's ed	-0.0005	-0.0055	0.0028
	(0.86)	(3.45)	(3.21)
fam. Income	0.0016	-0.0020	0.0040
	(3.20)	(2.45)	(4.30)
AFQT scores	0.0232	-0.0403	0.0462
	(5.29)	(4.03)	(4.29)
siblings	0.0004	0.0290	-0.0405
	(0.29)	(2.02)	(3.29)
nuclear family	0.0060	0.0138	-0.0441
	(2.95)	(0.20)	(1.65)

Table 2C
Non-Linearity and Non-Separability
(with asymptotic t-ratios)

Non-linearity	- -	
	(δ_2)	0.0406
		(6.73)
Non-Separability		
	(τ_1)	0.0035
		(1.62)
	(τ_2)	0.0437
		(5.01)

Table 2D
Unobserved Heterogeneity Support points
(with asymptotic t-ratios)

	Grade Transition	**Wage equation**		
Parameter	θ^G	θ^w	θ^S	θ^E
type 1	-0.0241	1.6598	-5.2934	-1.5555
	(0.14)	(18.37)	(44.22)	(11.04)
type 2	-1.3726	1.6020	-2.8635	-2.5175
	(9.29)	(20.61)	(15.02)	(19.03)
type 3	-2.3987	1.3792	-3.0329	-2.2997
	(12.87)	(18.23)	(18.03)	(10.63)
type 4	-1.5003	1.4791	-3.0329	-3.1716
	(8.12)	(19.79)	(4.29)	(19.25)

Table 2E
Type Probabilities
(with asymptotic t-ratios)

	q_k^0	p_k
type 1	-2.0326	0.03
	(12.25)	
type 2	0.1027	0.23
	(1.91)	
type 3	0.9034	0.52
	(6.79)	
type 4	0.0000	0.21
		-

Table 2F
Grade Specific Intercepts
(with asymptotic t-ratios)

grade level	Parameter	Estimate/ (t-ratio)
grade 6	$\gamma_{0,6}$	-3.0346 (9.03)
grade 7	$\gamma_{0,7}$	-1.3856 (3.78)
grade 8	$\gamma_{0,8}$	-0.9049 (2.37)
grade 9	$\gamma_{0,9}$	-0.4389 (1.38)
grade 10	$\gamma_{0,10}$	-1.5210 (3.20)
grade 11	$\gamma_{0,11}$	2.4129 (4.29)
grade 12	$\gamma_{0,12}$	1.5329 (4.87)
grade 13	$\gamma_{0,13}$	2.2638 (6.29)
grade 14	$\gamma_{0,14}$	1.3856 (4.39)
grade 15	$\gamma_{0,15}$	4.2948 (10.94)
grade 16	$\gamma_{0,16}$	3.2004 (11.29)
grade 17 or more	$\gamma_{0,17}$	4.2838 (12.92)

Table 2G
Distributions of the error shocks:
Mean Mixture of normals for grade transitions

	Mean	**Variance**	**Skewness**
grade level			
grade 6	0.2745	1.2298	0.6234
grade 7	0.1927	1.1836	-0.2236
grade 8	-0.3756	1.3332	-0.0028
grade 9	0.5319	1.3849	0.2935
grade 10	0.4429	1.4726	0.6349
grade 11	0.3620	2.1823.	-0.4727
grade 12	0.0387	1.0378	0.4223
grade 13	0.2987	1.6239	0.2925
grade 14	0.4190	1.5529	-0.4448
grade 15	0.7823	0.9238	0.5102
grade 16	0.2835	1.4965	0.4440
grade 17 or more	0.3956	1.2855	-0.3996

The different moments are calculated using each grade specific normal mixtures

Table 2H
Distributions of the error shocks:
Mixture of normals for the post-schooling wage distribution

	mean	variance	skewness
grade level			
grade 6	0.0325	0.2836	0.5398
grade 7	-0.0829	0.6725	0.9825
grade 8	-1.2356	0.2735	-0.2845
grade 9	0.2602	0.3002	-0.2856
grade 10	0.8428	0.3587	-0.6325
grade 11	-0.6438	0.5234	0.8639
grade 12	0.8845	0.4298	0.2745
grade 13	-1.6398	0.2839	0.2845
grade 14	1.0231	0.2019	-0.2536
grade 15	-0.8529	0.3976	-0.6724
grade 16	0.3856	0.2856	0.1288
grade 17 or more	0.2734	0.4523	0.2734

The different moments are calculated using each grade specific normal mixtures

Table 3A
Skill heterogeneity: first and second moments of the returns to schooling and experience

	Mean	**St. Dev**	**Minimum**	**Maximum**
Returns to Schooling				
until grade 12	**0.0376**	**0.0266**	**0.00001**	**0.1327**
grade13-more	**0.0864**	**0.0266**	**-**	**-**
Returns to experience	**0.0605**	**0.0201**	**0.0105**	**0.2939**
wage intercept	**1.4427**	**0.1056**	**1.2539**	**1.9645**

Note: The returns to schooling are measured at entrance in the labor market. The standard deviation measures cross-sectional dispersion across types and across regressors.

Table 3B
Correlation between Heterogeneity Components

	θ^G	θ^w	θ^S	θ^E
θ^G	1.0000	0.9160	-0.4221	-0.2785
θ^w		1.0000	-0.1307	-0.1710
θ^S			1.0000	0.8549
θ^E				1.0000

Table 4
Variance Decomposition of Skill heterogeneity:
Family Background, Unobserved Skills and Schooling

	% Variance explained		
	Parents' background variables	unobserved skills	accumulated schooling
Returns to Schooling	9%	91%	-
Returns to experience	24%	60%	18%
wage intercept	40%	60%	-

Note: The total shares for the returns to experience do not add to 100 because schooling is endogenous and therefore not orthogonal to background variables and skills.

Table 5A
Skill heterogeneity vs non-linearity:
Comparison of the returns to schooling and experience
(the st. deviations are in parentheses)

Specification					
		(1)	**(2)**	**(3)**	**(4)**
Skill Heterogeneity		**yes**	**no**	**yes**	**no**
Non-linearity		**no**	**yes**	**yes**	**no**
Non-Separability		**yes**	**yes**	**yes**	**yes**
Returns to Schooling					
in High school	**parameter**		**0.0101**		**0.0920**
	mean	**0.0604**		**0.0376**	
	st-dev	**0.0359**		**0.0266**	
			-		**-**
Post high school	**parameter**		**0.0725**		**0.0920**
	mean	**0.0604**		**0.0864**	**-**
	st.dev	**0.0359**	**-**	**0.0266**	

Note: The standard deviations provide a measure of cross-sectional dispersion.

Table 6
The returns to post-high-school education
in the early phase of labor market experience

years of experience	**returns to schooling**	**standard errors**
0	0.0823	(0.005)
2	0.0905	(0.006)
4	0.0997	(0.006)
6	0.1066	(0.007)
8	0.1104	(0.007)

Note: The value is average over the values of X_i and over realized schooling (from grade 13 onward). The standard errors are calculated accordingly (using the delta method).

Table 7
Returns to schooling with/without Separability
(standard deviations in parentheses)

	with separability	**with non-separability**	
Returns to Schooling			
until grade 12	**0.0498**	**0.0376**	
	(0.0234)	**(0.0211)**	
grade 13 or more	**0.0876**	**0.0864**	- -
	(0.0259)	**(0.0213)**	

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. The NLSY documents monthly activities for each months lying between two interviews. 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.[22] 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).[23] Given the age of the individuals in our sample, we assume that those who have already started to work full-time by 1990 (94% of our sample), will never return to school beyond 1990. Finally, one notes that the number of interruptions is relatively small.

Unlike many reduced-form studies which use proxies for post-schooling labor market experience (see Rosenzweig and Wolpin), we use actual labor market experience. Actual experience accumulated is computed using the fraction of the

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

[23]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.

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 wage variable is the hourly wage variable in the principle job declared by the individual. We do not distinguish between those working part-time and those working full-time.

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.