

IZA DP No. 1083

**On the Specification of Mincerian Wage
Regressions with Heterogeneity, Non-Linearity,
Non-Separability, and Heteroskedasticity**

Christian Belzil

March 2004

On the Specification of Mincerian Wage Regressions with Heterogeneity, Non-Linearity, Non-Separability, and Heteroskedasticity

Christian Belzil

*GATE, University of Lyon 2,
CIRANO and IZA Bonn*

Discussion Paper No. 1083
March 2004

IZA

P.O. Box 7240
53072 Bonn
Germany

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

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available on the IZA website (www.iza.org) or directly from the author.

ABSTRACT

On the Specification of Mincerian Wage Regressions with Heterogeneity, Non-Linearity, Non-Separability, and Heteroskedasticity*

Using panel data taken from the NLSY, I perform the joint estimation of i) a reduced-form dynamic model of the transition from one grade level to the next with observed and unobserved heterogeneity, and ii) a flexible version of the celebrated Mincerian wage regression with skill heterogeneity, non-linearity in schooling, non-separability between the effects of schooling and experience and heteroskedasticity (after conditioning on unobserved skills). 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, schooling has a causal effect on wage growth. I find that estimates obtained in a standard framework (assuming separability) may underestimate the returns to schooling upon labor market entrance by as much as 15%. Finally, I find that the variance of the stochastic wage shock decreases with accumulated experience but is more or less independent of schooling.

JEL Classification: J2, J3

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

Christian Belzil
IZA Bonn
P.O. Box 7240
53072 Bonn
Germany
Email: belzil@iza.org

* I thank participants at the SOLE Transatlantic Meeting (June 2003) and the Social Sciences and Humanities Research Council of Canada for generous support.

1 Introduction

The Mincerian wage regression is one of the most widely used tools of empirical economics. The literature making use of empirical wage regressions is vast. Mincerian wage regressions have been applied to numerous areas of Labor Economics. They play a central part in the literature devoted to the returns to education (Card, 2001, and Belzil and Hansen, 2002a) as well as in the literature on the increase in wage inequality (Card and Lemieux, 2002, Gosling, Machin and Meghir, 2000, and Taber, 2001). They are also used to investigate statistical discrimination and gender differences in wages (Fortin and Lemieux, 1998), occupation choices (Keane and Wolpin, 1997) and are also present in the equilibrium search literature (Postel-Vinay and Robin, 2002, and Eckstein and van den Berg, 2003).

While the vast majority of the literature is quite homogeneous at the model specification level, the more recent literature is marked by important differences at the level of the functional form, the stochastic specification as well as the estimation method.

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 regression 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 (74).¹ Consequently, economists have started to pay particular attention to the introduction of heterogeneity in the slopes of the wage regression, to potential non-linearity (the convexity of the wage schooling relationship) and to the separability between education and experience. On top of that, the link between risk and education is becoming a major topic of ongoing research.

With regards to skill heterogeneity, the random coefficient representation of the wage regression function has gained in popularity, along with the literature on estimating average treatment effects.² At the same time, others

¹For more details, see Heckman, Lochner and Todd, forthcoming and Lemieux, 2003.

²The 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, 2000), Meghir and Palme (2002), Woolridge (1998, 2000), Angrist and Imbens (1994) and Card

have paid a particular attention to potential non-linearities explained by differences in local returns to the schooling across grade levels (Belzil and Hansen, 2002a).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 regression function might only reflect dynamic self-selection (merely a composition effect). That is, as we move toward higher levels of schooling, the local returns to schooling may turn out to be estimated from an increasingly large proportion of high ability workers. If so, allowing for cross-sectional heterogeneity in the slope parameter (s) of the wage regression might obviate the need for a flexible (non-linear) specification of the wage regression function and facilitate estimation. Equally, if the wage regression 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.

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 regression function using Instrumental Variable (IV) techniques and interpret the estimates within a linear random coefficient framework. The linearity assumption is therefore crucial.⁴ However, if the linear wage re-

(2000)). Belzil and Hansen (2002,b) present a structural analysis of the correlated random coefficient wage regression model.

³Belzil 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) 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. The degree of convexity of the wage regression (as well as its change over time) is also pointed in Mincer (1997), Lemieux (2003) and Deschênes (2001).

⁴In general, the use of IV techniques requires separability between the instruments and

gression is not supported by the data and the form of the wage regression 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 evidence.

On a similar note, 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 and Todd, forthcoming and Lemieux, 2003). This suggests that log wages regression may not be separable in education and experience and, in particular, that the return to experience may be affected by schooling. Various economic models may be able to explain this. These include models of endogenous post-schooling human capital investments as well as various lifecycle incentive models where wages are upward sloping.⁵

Specification issues are not only limited to non-linearity, separability and skill heterogeneity. In the applied literature based on cross-section techniques, the distinction between pure random shocks and unobserved persistent attributes is not possible. For this reason, heteroskedasticity is typically ignored, even though the relationship between risk and education is a rather fundamental economic question. The effect of schooling on earnings dispersion (or wage and employment rate dispersion) is a difficult question to tackle and few people have tried to quantify the link between education and labor market earnings risk. As of now, there is no definite answer to the question of the link between education (or human capital) and risk and, in particular, the link between the variance of future wage shocks and accumulated human capital.⁶

Finally, it should be noted that the literature is not only characterized by

the error term in the treatment equation.

⁵See Rosenzweig and Wolpin (2000) for a discussion of endogeneous post-schooling investments and Belzil and Bognanno (2004) for a dynamic promotion models which implies that wage growth may depend on schooling.

⁶In practice, evaluating the risk associated to human capital investment requires to incorporate various dimensions such as the variance of the idiosyncratic shocks affecting wages, the failure probability (in school), expectations about future tuitions, technological change foresight and even uncertainty about one's ability (see Carneiro and Heckman, 2002, Chen, 2003 and Belzil and Hansen, 2002a).

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, 2002a) are much lower.⁷ These results are difficult to reconcile, as each estimation method has its own weaknesses.⁸ As indicated above, in the IV/OLS literature, nonlinearities are practically never allowed (at least formally) and the possibility of relying on “weak instruments” is rarely envisaged.⁹ At the same time, structural estimation requires parametric (distributional) assumptions which are not needed in the linear IV framework.¹⁰ The robustness of the results are therefore not always easy to evaluate.

This paper is driven by one major objective. It is to obtain estimates of all the components of the Mincerian wage regression within an econometric specification where schooling is endogenous, and where the wage regression is estimated as flexibly as in the structural estimation literature but which restricts the number of parametric/distributional assumptions to a

⁷The 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 Mincerian wage regression. 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 plays a central part.

⁸In Keane and Wolpin (1997), the returns to schooling are linear in schooling but are occupation specific. The population average return is between 5% and 8% per year (depending on the model). In Belzil and Hansen (2002,a), the returns vary with grade level and range between less than 1% per year (in high school) to more than 12% at college graduation. Both papers are based on a sample of young white males taken from the NLSY.

⁹See Staiger and Stock, 1997.

¹⁰The degree of under-identification (non-parametric) found in standard empirical dynamic programming models is well established in Rust (1994) and Magnac and Thesmar (2002).

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 (as in Cameron and Heckman, 1998) and a non-linear Mincerian wage regression 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 Mincerian wage regression 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), iii) the returns to experience depend on accumulated schooling and iv) the distribution of wage offers are conditionally (on skill heterogeneity) heteroskedastic. Secondly, I perform statistical tests of these various hypotheses (skill heterogeneity, non-linearity, separability and heteroskedasticity) in order to shed light on the optimal specification of the celebrated Mincerian wage regression 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 the returns to schooling and experience obtained under various scenarios.

The model is estimated using data from the National longitudinal survey of Youth (79-90). This is the same sample used by Belzil and Hansen (2002a) and Keane and Wolpin (1997). For this reason, comparing the returns to schooling and experience from structural models and reduced-form models is particularly enlightening. 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 regression (intercept term, returns to schooling and returns to experience) require 11 parameters

(4 support points and 7 observable regressors). Along with the variance of the idiosyncratic shock, the wage regression requires 38 parameters. As far as I know, this is the most general Mincer wage regression ever estimated.

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 regression, 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 regression 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, skill heterogeneity is also found to be quite important and I find that there is more cross-sectional variability in the returns to experience than in the returns to schooling. I find that those endowed with high returns to schooling will also be endowed with high returns to experience, although the correlation is quite small. 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 the individual specific returns to experience. This is consistent with the view that accumulated schooling may have a causal effect on wage growth. Aside from the major defect of underestimating the returns to schooling as individuals accumulate work experience, estimates obtained in a standard separable framework also underestimate the returns to schooling upon labor market entrance by a significant margin (as much as 15%).

Finally, wages are found to be homoskedastic in schooling but heteroskedastic in experience. This result is consistent with the possibility that senior workers are less exposed to business cycle fluctuations or other stochastic shocks affecting the labor market.

The results reported here are in line with those found in the structural literature. The estimates of the returns to schooling, much lower than those 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 regressions and not by the stronger parametric assumptions required in structural estimation.

The paper is structured as follows. The empirical model is exposed in Section 2. Section 3 is devoted to the results of the statistical tests. The

structural parameter estimates are discussed in Section 4 and the relative importance of skill heterogeneity and non-linearities is studied in Section 5. In Section 6, I investigate the importance of allowing for non-separability. In Section 7, I discuss the importance of heteroskedasticity (in comparison with skill heterogeneity). The conclusion is in Section 8.

2 A Reduced-form Dynamic Model of Schooling and Wages

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

2.1 Schooling attainments

The econometric model used to deal with the endogeneity of schooling attainment is a hazard function model of grade transition. I denote the conditional probability of stopping school with grade level S (the hazard rate) by $F(\gamma_{i,S})$ where $F(\cdot)$ is a cumulative distribution function and where γ_{iS}^* an individual/grade specific index. The index, γ_{iS}^* , is expressed as

$$\gamma_{iS}^* = \gamma_{0S} + X_i' \gamma_1 + \theta_i^G \quad (1)$$

where X_i is a vector of exogenous regressors, γ_1 is a vector of unknown parameters and γ_{0S} is a vector of grade specific intercept terms to be estimated. The continuation probability is therefore 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.¹¹ Note that θ_i^G is assumed to be orthogonal

¹¹To do so, I regressed AFQT scores on schooling and kept the residual.

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.

2.2 The Mincerian Wage Regression with Heterogeneity, Non-linearity, Non-separability and Heteroskedasticity.

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

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

where S_i denotes schooling and $Exper_{it}$ is accumulated years of experience at date t . I use actual experience as opposed to potential experience (see Appendix 1). I assume that $\varepsilon_{it}^w \sim i.i.d N(0, \sigma_{it}^2)$. In order to estimate the model, I choose a tractable form for $\varphi_i(S_i, Exper_{it})$, that is

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

- $\varphi_i^S(\cdot) = \varphi_{i1} \cdot S_i + \delta_2 \cdot S_{ic}$
 - 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 $S_i \leq 12$
- $\varphi_i^E = \exp(X_i' \beta^E + \tau_1 \cdot S_i + \tau_2 \cdot S_{ic} + \theta_i^E)$
- $\varphi_i^w = X_i' \beta^w + \theta_i^w$
- $(\theta_i^G, \theta_i^s, \theta_i^E, \theta_i^w)$ are jointly distributed with CDF $H(\cdot)$
- $\sigma_{it} = \exp(\sigma_0 + \sigma_1 \cdot S_i + \sigma_2 \cdot S_i^2 + \sigma_3 \cdot Exper_{it} + \sigma_4 \cdot Exper_{it}^2)$

Altogether, the definitions of $\varphi_i^S(\cdot)$, φ_i^E and φ_i^w 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.¹² The parameters ranging from σ_1 to σ_4 enable me to take into account that the variance of wage offers varies with accumulated schooling and experience. The positivity of φ_{i1} and φ_i^E 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_i^G and labor market skill heterogeneity ($\theta_i^s, \theta_i^E, \theta_i^w$) will capture any endogeneity in schooling which may persist 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

$$\frac{\delta \log w_{it}}{\delta S_i} = \frac{\delta \varphi_i^S(\cdot)}{\delta S_i} + \frac{\delta \varphi_i^E(\cdot)}{\delta S_i} \cdot \text{Exper}_{it} \quad (4)$$

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

$$\frac{\delta \log w_{it}}{\delta S_i} = \varphi_i^S(\cdot) + \delta_2 + [\varphi_i^E(S_i) \cdot (\tau_1 + \tau_2) \cdot \text{Exper}_{it}] \quad (5)$$

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

2.3 Unobserved Heterogeneity

In order to approximate $H(\cdot)$ as accurately as possible, I assume that there are 4 types of individuals. Each type is therefore endowed with a vector

¹²Another type of non-separability (ignored in this paper) could arise if the returns to schooling decline with experience (or age) because of depreciation.

$(\theta_i^G, \theta_i^s, \theta_i^E, \theta_i^w)$ for $k = 1, 2, \dots, 4$. The probability of belonging to type k , p_k , are estimated using logistic transforms.

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

and with the restriction that $q_4 = 0$.

2.4 The Likelihood Function

The likelihood function is the joint probability of observing a level of schooling attainment, S_i , and a particular wage history $(w_{i1} \dots w_{i1990})$. Given type k , it has three components; the probability of having continued in school until S years of schooling is achieved (L_{1k}), the joint probability of stopping school with S years of schooling and entering the labor market at observed wage w_{i1} (denoted L_{2k}) and the density of observed wages until 1990 (denoted L_{3k}). The entry wage (w_{i1}) and the decision to stop school must be treated as dependent, even after conditioning on unobserved heterogeneity, because the schooling decision is most likely based, among other things, on starting wages.¹³

Given a distributional assumption for the hazard (and therefore the school continuation probabilities), L_{1k} is simply $\prod_{j=1}^{S-1} (1 - F_j(\cdot))$ while L_{2k} would typically require the evaluation of the conditional probability of stopping school (conditional on the observed entry wage) times the marginal distribution of the entry wage. Finally L_{3k} is just the product of wages densities $(w_{i2}, \dots, w_{i1990})$. For a given type k , the likelihood is therefore $L_k = L_{1k} \cdot L_{2k} \cdot L_{3k}$ and the log likelihood function to be maximized is

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

where each p_k represents the population proportion of type k .

2.5 Identification

The identification strategy borrows substantially from the literature on estimating hazard functions. The treatment of the grade transition as a hazard

¹³This type of interdependence is more clearly defined within a structural model.

function allows me to identify unobserved heterogeneity (conditional on exogenous parents' background variables) affecting grade transition (θ^G). While non-parametric identification results exist only for proportional hazard functions (which is not the case here), the estimation of non-proportional hazards with person specific unobserved factors is also quite common (see Baker and Melino, 2000). It is therefore reasonable to assume that the grade transition hazard (set in discrete time) will allow me to capture the unobserved component affecting schooling decisions. 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.

3 Searching for the Best Specification

The estimation of the model requires a choice for the cumulative distribution function used in the grade transition, $F(\cdot)$. The standard normal is a natural choice, as it will easily allow for the correlation between the school leaving decision and the wage observed in the first year of full time work. If one is willing to assume independence (for the entry year), then the logit hazard function is also attractive.¹⁴ The results presented below are for the normal case. Evaluating the type specific likelihood components (L_{1k} , L_{2k} and L_{3k}) is rendered easy with the normality assumption

As a first step, I estimated the most general model possible and re-estimated several restricted versions that allowed me to perform likelihood ratio tests. In order to implement the model, I have initially fitted a version with observed heterogeneity and gradually included unobserved types. Various experiments have indicated that it is not necessary to go beyond 4 types.¹⁵ It should be noted that the 7 regressors representing family background are relatively 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 regression.

There are 4 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, the 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 $\beta^s = \beta^E = 0$, $\theta_1^s = \dots\theta_k^s$, and $\theta_1^E = \dots\theta_k^E$ (the “classical representation” of the Mincer wage regression). This test hinges on a fixed (known) number of types.¹⁶ Finally, the last null hypothesis is that the wage regression function

¹⁴Indeed, the estimation of the model with logistic hazard function has delivered practically identical results for the log wage regression parameters and comparable results (aside from an obvious scaling problem) for the grade transition equation.

¹⁵Indeed, 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.

¹⁶Considering the number of types as fixed is relatively standard in the empirical literature where the estimation method consists of a relatively complicated mixed likelihood

is homoskedastic in accumulated human capital (schooling and experience). This is accomplished by imposing $\sigma_1 = \dots\sigma_4 = 0$.

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 four 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, skill heterogeneity in the slopes as well as heteroskedasticity. 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 to be addressed below.

4 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 best model specification is found in a sequence of tables ranging from Table 2A to Table 2F. Without loss of generality, the returns to education are measured upon entrance in the labor market (when $Exper_{it} = 0$). The first and second moments of the slope parameters as well as the key correlations are found in table 3A, 3B and 3C. A Variance decomposition is documented in Table 4.

4.1 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. This may be seen after noting that, given the range of father's and mother's schooling, the negative effect of the interaction term (-0.0181) will dominate when evaluating the marginal effects for both mother's schooling and father's schooling. The parameter estimates also imply that schooling attainments will increase with household income, AFQT scores and de-

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).

crease 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 (2002a), 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.

4.2 Parents Background Variables and Wages

Overall, It is clear that the level of significance of the family background variables in the wage regression 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 associated with higher returns to schooling, higher returns to experience and higher wage intercepts (Table 2A). 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.

4.3 Non-linearity

The results found in Table 2B show strong evidence in favor of the convexity of the wage-schooling relationship. The parameter estimate for δ_2 (equal to 0.0435), along with the estimates for β^s (Table 2A), imply an average return to schooling equal to 0.0426 per year of schooling prior to high school graduation and 0.0863 in college (Table 3A). This is consistent with evidence presented in Belzil and Hansen (2002a) 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. This issue will be addressed in Section 7.¹⁷

Aside from the convexity, it should also be noted that, when compared to the OLS estimate of the returns to schooling applied over the entire sample

¹⁷The increasing degree of convexity of the wage regression is also pointed in Mincer (1997), Lemieux (2003) and Deschênes (2001).

(9.9%) and to those reported in the IV literature, these estimates are small. They are in accordance with the relatively lower estimates obtained in the structural literature. As do structural models, the maximum likelihood estimates presented therein imply a relatively strong ability bias.¹⁸ Finally, the returns to experience, averaging 6% per year in the population, exceed the return to schooling in high-school but are below the returns to college.

4.4 Unobserved Heterogeneity

An inspection of the unobserved heterogeneity support points, along with the type probabilities (Table 2C and Table 2D), reveals that those types who will experience higher schooling attainments (lower hazard rates) will also experience higher returns to schooling and experience. More precisely, and after conditioning on observed attributes, type 3 individuals (representing 54% of the population) will experience the highest level of schooling (the lowest hazard) and will be endowed with the second highest returns to schooling and experience. The highest return to schooling and experience is achieved by type 2 individuals while type 4 individuals will obtain the lowest returns.

4.5 The Causal Effect of Schooling on Wage Growth

After conditioning on skill heterogeneity, and taking into account the endogeneity of schooling, there is a positive correlation between accumulated schooling and the individual specific returns to experience (Table 2B). 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 may have a causal effect on wage growth.¹⁹

¹⁸In the OLS/IV literature, the difference between IV and OLS estimates is often imputed to potential measurement error in schooling. The reader will however notice that, in a very general model specification such as the one proposed here, this argument breaks down.

¹⁹In the recent (but growing) empirical literature devoted to personnel economics, it is reported that schooling has a relatively small effect on promotion outcomes (see Baker, Gibbs and Holmstrom, 1993, Lazear, 1999, and Belzil and Bognanno, 2004).

4.6 Heteroskedasticity

The wage regression function appears to be heteroskedastic in experience (Table 2F). The estimates for experience (-0.0298) and experience² (0.0012) indicate that the wage variance will decrease with experience over the relevant range. However, the insignificance of the parameters associated to education and its square seems to indicate that there seems to be very little correlation between risk (at the wage level) and education.

4.7 Cross-Sectional Dispersion in the Returns to Schooling and Experience

The second moments (the standard deviations) of all the determinants of the wage regressions, also found in table 3A, indicate that there is more cross-sectional variability in the returns to experience than in the returns to schooling. Interestingly, and as indicated by the correlations found in Table 3B those endowed with high returns to schooling will also be endowed with high returns to experience, although the correlation (0.05) is quite small. The unobserved components of the returns to schooling and experience (Table 3C) are however more strongly correlated (the correlation is 0.76).

4.8 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 regressions 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 10% of the cross sectional variations in returns to schooling is explained by parents background variables while 90 % is explained by unobserved skills
- 19% of the cross-sectional variations in the returns to experience are explained by parents background variables while 65% are explained by

unobserved skills. Interestingly, accumulated schooling explains 16% of the returns to experience.

- 41% of the cross sectional variations in the wage intercept are explained by parents background variables while 59% are explained by unobserved skills.

5 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 regressions are assumed to be linear in schooling.

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.

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 4.5 percentage points in a flexible model which allows for both skill heterogeneity and non-linearities (column 3), is now increased to 7 percentage point when skill heterogeneity is not controlled for (column 2). This results is not surprising and illustrates eloquently the 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 6.5% per year (column 1), seriously over estimates the return to high school training and under estimates the return to post high school training.

Ignoring both nonlinearity and skill heterogeneity, raises the return to schooling to 0.0892 (column 4) and creates the statistical illusion that log wages are increased by more than 9% per year, what ever is the level of schooling. This estimate, 8.9%, is still below what would be obtained by OLS (around 9.9% per year) but it is the largest obtained so far. Somehow, this indicates that differences in the intercept term alone, are not capable of capturing a level of dynamic self-selection compatible with the ability bias implied by the most general model specification(s).

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 0.0259 (when non-linearity is accounted for) to 0.0362 (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.

6 The Importance of Non-Separability

Among all particular dimensions that I have examined, the issue of separability of log wages in education and the 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 may not only be seen from a human capital perspective but also from a pure incentive angle. 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. Despite the seemingly small estimate for τ_2 , it is clear that the return to post high-school training rises 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 esti-

mation level. To illustrate this, I re-estimated a conventional form of the regression (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. I find that imposing separability reduces the return to schooling (beyond high school graduation) by around one percentage point. In this restricted version, the population average return to college training is 0.0795. As such, these estimates indicate that evaluating the returns to schooling using a traditional (separable) functional form, induces estimated returns to schooling upon entrance in the market which are 14% to 15% lower than they should. This is a severe under-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.

7 Heteroskedasticity or Heterogeneity?

Empirical wage regressions are virtually always assumed to be homoskedastic. This is surprising as the relationship between risk and education is a rather fundamental economic question. The effect of schooling on earnings dispersion (or wage and employment rate dispersion) is a difficult question to tackle. While few individuals have recently investigated the determinants of school financing in a context where individuals are willing to smooth consumption (Keane and Wolpin, 2001, Sauer, 2002 and Cameron and Taber, forthcoming), few people have tried to quantify the link between education and labor market earnings risk (Belzil and Hansen, forthcoming, Carneiro, Hansen and Heckman, 2002, and Chen, 2003, are recent examples). In stylized “implicit contract” frameworks, in which risk averse individuals are willing to trade wage rigidity against stable employment patterns, it is reasonable to assume that the need for risk sharing will be smaller for low educated workers who benefit from a relatively high level of social insurance. However, at the same time, wage dispersion may also vary with factors such as union status, occupation type and the like. As a consequence, the link between education and wage/earnings dispersion is not trivial.²⁰

²⁰For a survey of the contract literature, see Rosen (1985).

Evidence already presented in Table 2A suggests that Mincerian wage regressions estimated from the NLSY are homoskedastic in schooling but heteroskedastic in experience. This result is consistent with the possibility that senior workers are less exposed to business cycle fluctuations or other stochastic shocks affecting the labor market. However, given the reduced-form nature of the model, it is not possible to give a “structural” interpretation to the level (or absence) of heteroskedasticity. It nevertheless appears that modeling the channels by which accumulated experience affects earnings or wage dispersion would be most appropriate.

In classical random coefficient framework, in which regressors are exogenous, the regression model is often represented as a linear regression with an heteroskedastic error term. The degree of heteroskedasticity reflects the degree of cross sectional variation in the coefficients. It is therefore relevant to ask if the level of heteroskedasticity found in wage regression is strongly affected by the absence of control for skill heterogeneity. To provide a transparent example, I re-estimated the model with a linear relationship for the variance function (excluding the quadratic term for schooling and experience). As a first step, I estimate the unrestricted version of the linear heteroskedasticity model (with skill heterogeneity, non-linearity and non-separability). The estimates are found in column 1 of Table 8. Subsequently, I re-estimated it while imposing population homogeneity in the returns to schooling and experience (column 2 of Table 8). The results illustrate that the degree of heteroskedasticity is inflated when skill heterogeneity is ignored. For instance, the variance of wages seem to increase with schooling (see column 2) while the homoskedasticity hypothesis would be practically rejected in presence of skill heterogeneity (column 1). Again, the evidence suggests that heteroskedasticity is created by labor market experience, not schooling.

8 Conclusion

In this paper, I present econometric estimates of the celebrated Mincer wage regression obtained with a degree of flexibility which, as far as I know, has not been achieved before. The econometric model is novel. It is based on 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 wage regression with the following attributes; skill heterogeneity, non linearity in schooling, non-separability between schooling and experience and heteroskedasticity. The model is estimated from a panel of young males taken from the National longitudinal survey of Youth (79-90). The empirical likelihood function maximizes the joint probability of the schooling attainment and the wage histories observed between 1979 and 1990.

The data reject all simplifying assumptions common in the empirical literature. I find that the degree of convexity of the wage regression, 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 regression 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 and I find that there is more cross-sectional variability in the returns to experience than in the returns to schooling. Interestingly, those endowed with high returns to schooling will also be endowed with high returns to experience, although the correlation is quite small.

After conditioning on skill heterogeneity, there is a positive correlation between accumulated schooling and the individual specific returns to experience. This is consistent with the view that accumulated schooling may have a causal effect on wage growth. 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 under-estimate the returns to schooling upon labor market entrance by a significant margin (as much as 15%).

Finally, wages are found to be homoskedastic in schooling but heteroskedastic in experience. This result is consistent with the possibility that senior workers are less exposed to business cycle fluctuations or other stochastic shocks affecting the labor market.

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%. The huge discrepancy between structural estimates and OLS/IV estimates does not seem to be attributable to the strong parametric

assumptions required in structural models. After all, the estimation strategy proposed in this paper is based on a wage regression more generally specified than in the literature and on a relatively flexible way to treat the endogeneity of schooling decisions. The discrepancy between structural estimates and reduced-form estimates has been noticed relatively recently. As of now, it is certainly not well understood. The results reported here seem to suggest that it may well be explained by differences in the econometric specification of the wage regressions but only further investigation will enable applied econometricians to fully understand these marked differences.

References

- [1] Baker, Michael and Angelo Melino (2000) "Duration Dependence and Non-Parametric Heterogeneity: A Monte Carlo Study" *Journal of Econometrics*, vol 96, 2:357-393.
- [2] Baker, George, Gibbs, Michael and Holmstrom, Bengt. "Hierarchies and Compensation: A Case Study," *European Economic Review*, 37 (1993): 366-378.
- [3] Belzil, Christian and Michael Bognanno (2004) "The Promotion Dynamics of American Executives" Working Paper, IZA and CIRANO.
- [4] Belzil, Christian and Hansen, Jörgen (2002a) "Unobserved Ability and the Return to Schooling", *Econometrica*, vol 70, 575-591.
- [5] Belzil, Christian and Hansen, Jörgen (2003) "Structural Estimates of the Intergenerational Education Correlation", *Journal of Applied Econometrics*, vol18: 679-696.
- [6] Belzil, Christian and Hansen, Jörgen (2002b) "A Structural Analysis of the Correlated Random Coefficient Wage Regression Model" Working Paper, IZA.
- [7] Belzil, Christian and Hansen, Jörgen (forthcoming) "Earnings Dispersion" forthcoming in *Research in Labor Economics*, volume 23.
- [8] 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.
- [9] Cameron, Stephen and Chris Taber (forthcoming) "Estimation of Educational Borrowing Constraints" *Journal of Political Economy*
- [10] Card, David (2001) "The Causal Effect of Education on Earnings" *Handbook of Labor Economics*, edited by David Card and Orley Ashenfelter, North-Holland Publishers.

- [11] Card, David and Thomas Lemieux (2001) “Can Falling Supply Explain the rising Return to college for Younger men? A Cohort-Based Analysis” *Quarterly Journal of Economics* 116, May 2001, 705-46
- [12] 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.
- [13] Chen, Stacey (2003) “ Does College Attendance Increase wage Volatility? Working Paper, SUNY Albany.
- [14] Deschênes, Olivier “unobserved Ability, Comparative Advantage and the Rising Return to Education in the United States: A Cohort Based Approach”
- [15] Eckstein, Zvi and Kenneth Wolpin (1999) “Youth Employment and Academic Performance in High School”, *Econometrica* 67 (6,)
- [16] Gosling, Amanda, Stephen Machin and Costas Meghir (2000) “The Changing Distribution of Male Wages in the UK” *Review of Economic Studies* 67, October, 635-66.
- [17] Heckman, James, Lochner, Lance and Petra Todd (2001) “Fifty Years of Mincer Earnings Regressions” forthcoming in *The Journal of Political Economy*
- [18] 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.
- [19] Heckman, James and E. Vytlačil (2000) “Local Instrumental variables and Latent Variable Models for Identifying and Bounding Treatment Effects”, Working paper, Department of Economics, University of Chicago.
- [20] Heckman, James and E. Vytlačil (1998) “Instrumental Variables Methods for the Correlated Random Coefficient Model”, *Journal of Human Resources*, Volume 33, (4), 974-987.
- [21] Hsiao, Cheng (1988) *The Analysis of Panel data*, Cambridge Press

- [22] Imbens, Guido and J. Angrist (1994) "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, 62, 4, 467-76.
- [23] Keane, Michael P. and Wolpin, Kenneth (1997) "The Career Decisions of Young Men" *Journal of Political Economy*, 105 (3), 473-522.
- [24] Lazear, Edward (1999), "Personnel Economics: Past Lessons and Future Directions" *Journal of Labor Economics* vol 17, no 2, pp 199-236
- [25] Lemieux, Thomas "The Mincer Equation Thirty years after Schooling, Experience and Earnings", Working Paper, Center for Labor Economics, UC Berkeley.
- [26] Fortin, Nicole and Lemieux, Thomas (1998) "Rank Regressions, Wage Distributions and the Gender Wage Gap" *Journal of Human Resources*, 33, summer, 610-643.
- [27] Magnac, T. and D. Thesmar (2001): "Identifying Dynamic Discrete Decision Processes," *Econometrica*
- [28] Meghir, Costas and Marten Palme (2001) "The Effect of a Social Experiment in Education" Working Paper, UCL.
- [29] Mincer, Jacob (1997) "Changes in Wage Inequality, *Research in Labor Economics* 16, 1-18.
- [30] Mincer, Jacob (1974) "Education and Earnings, Columbia University
- [31] Rosen, Sherwin (1985) "Implicit Contracts: A Survey" *Journal of Economic Literature*
- [32] Rosenzweig Mark and K. Wolpin (2000) "Natural Natural Experiments in Economics" *Journal of Economic Literature*, December, 827-74.
- [33] Taber, Christopher (2001) "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability", *Review of Economic Studies*.
- [34] Sauer, Robert (2001) "Education Financing and Lifetime Earnings" *Review of Economic Studies*.

- [35] Staiger, Douglas and James H. Stock (1997), “Instrumental Variables Regression with Weak Instruments” *Econometrica*, 65: 557-586.
- [36] Willis, R. and S. Rosen (1979) “Education and Self-Selection”, *Journal of Political Economy*, 87, S-7-S36.
- [37] Wooldridge, Jeffrey M.(2000) “Instrumental Variable Estimation of the Average Treatment Effect in the Correlated Random Coefficient Model”, Working Paper
- [38] 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	Likelihood Ratio statistics	# of restrictions	critical value at 1% level
Linear returns to schooling	17.1	1	6.6
Effect of schooling on return to experience	10.9	2	9.2
Homogenous returns schooling/experience	271.9	14	29.1
Homoskedasticity	19.6	4	13.3

**Table 2A- The Effects of Parents Background Variables
(asymptotic t-ratios)**

parameters	Grade	Wage Regression		
	transition	intercept	return	return
	γ_1	term	to schooling	to experience
		β^w	β^S	β^E
Family background variables				
father's educ	0.1375 (7.04)	0.0010 (0.13)	0.0822 (4.31)	-0.0415 (2.76)
mother's educ	0.1646 (10.79)	-0.0074 (-1.04)	0.0481 (2.59)	-0.0387 (3.11)
father's ed.*mother's ed	-0.0181 (12.01)	-0.0003 (0.56)	-0.0048 (3.45)	0.0033 (3.21)
fam. Income	-0.0071 (4.69)	0.0014 (3.25)	-0.0027 (2.45)	0.0043 (4.90)
AFQT scores	-0.3336 (21.35)	0.0258 (6.72)	-0.0425 (4.03)	0.0466 (4.97)
siblings	0.1033 (6.21)	0.0008 (0.18)	0.0282 (2.02)	-0.0410 (3.03)
nuclear family	-0.1143 (1.54)	0.0067 (3.06)	0.0124 (0.20)	-0.0421 (1.75)

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

Non-linearity	-	-	
	(δ_2)		0.0435 (8.07)
Non-Separability			
	(τ_1)		0.0033 (1.20)
	(τ_2)		0.0573 (4.27)

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

Parameter	Grade Transition	Wage regression		
	θ	θ^w	θ^S	θ^E
type 1	0.0000 (normalized)	1.6598 (18.37)	-6.9796 (44.22)	-1.4075 (10.910)
type 2	-1.5302 (8.29)	1.6020 (20.61)	-3.3086 (15.70)	-2.5175 (19.57)
type 3	-2.8651 (19.52)	1.3792 (18.23)	-3.6724 (16.85)	-2.5537 (19.62)
type 4	-1.5762 (8.87)	1.4791 (19.79)	-13.3947 (3.79)	-3.1716 (20.48)

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

	q_k^0	P_k	
type 1	-2.1757 (10.04)	0.0225 (13.23)	-
type 2	0.1481 (1.11)	0.2365 (19.67)	-
type 3	0.9684 (8.39)	0.5371 (20.06)	-
type 4	0.00	0.2036	-
	-	-	

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

grade level	
grade 6	-3.3499 (8.47)
grade 7	-1.6928 (4.01)
grade 8	-0.9949 (2.45)
grade 9	-0.5181 (1.27)
grade 10	-1.4109 (4.40)
grade 11	2.3609 (6.37)
grade 12	1.8589 (4.85)
grade 13	2.5220 (6.75)
grade 14	1.2321 (5.02)
grade 15	4.8649 (14.52)
grade 16	3.5557 (10.47)
grade 17 or more	4.7731 (13.09)

Table 2F
Heteroskedasticity in the wage regression
(with asymptotic t-ratios)
Wage variance

constant (σ_0)	-1.1705 (22.50)
schooling (σ_1)	0.0074 (0.53)
schooling ² (σ_2)	-0.00002 (0.25)
experience (σ_3)	-0.0258 (3.76)
experience ² (σ_4)	0.0012 (1.65)
cov(ε_{it} , <i>hazard</i>)	0.1545 (1.95)

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

	Mean	St. Dev	Minimum	Maximum
Returns to Schooling				
<i>until grade 12</i>	0.0428	0.0259	0.00001	0.1491
<i>grade13-more</i>	0.0864	0.0259	-	-
Returns to experience	0.0606	0.0288	0.0157	0.2813
wage intercept	1.5293	0.1256	1.2762	1.9186

Note: The returns to schooling are measured at entrance in the labor market.

Table 3B
Skill heterogeneity: Correlation table

	Returns to Schooling (φ_{i1})	Returns to experience (φ_i^E)	Wage intercept (φ_i^W)
Returns to Schooling (φ_{i1})	1.0000	-	-
Returns to experience (φ_i^E)	0.0503	1.0000	-
wage intercept (φ_i^W)	-0.0626	0.4335	1.0000

Note: The returns to schooling are measured at entrance in the labor market.

Table 3C
Unobserved heterogeneity: Some Correlations

	Returns to Schooling	Returns to experience	wage intercept
Returns to Schooling	1.0000	-	-
Returns to experience	0.7631	1.0000	-
wage intercept	-0.1177	0.1279	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	10%	90%	-
Returns to experience	19%	65%	16%
wage intercept	41%	59%	-

Table 5
Skill heterogeneity vs non-linearity:
Comparison of the returns to schooling and experience
(standard deviation in parentheses)

Specification	(1)	(2)	(3)	(4)
Skill Heterogeneity	yes	no	yes	no
Non-linearity	no	yes	yes	no
Returns to Schooling				
 High school	0.0654 (0.0367)	0.0020 (0.00)	0.0428 (0.0259)	0.0892 (0.00)
 Post high school	-	0.0725 (0.00)	0.0864 (0.0259)	-
Returns to experience	0.0611 (0.0301)	0.0596 (0.00)	0.0606 (0.0288)	0.0592 (0.00)

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

years of experience	returns to schooling
0	0.0864
2	0.0925
4	0.0986
6	0.1047
8	0.1092

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

	with non-separability	with separability
Returns to Schooling		
<i>until grade 12</i>	0.0428 (0.0259)	0.0409 (0.0213)
<i>grade 13 or more</i>	0.0864 (0.0259)	0.0795 (0.0213) - -
experience	0.0606 (0.0288)	0.0625 (0.0295)

Table 8
The Effect of ignoring skill heterogeneity on Heteroskedasticity
(with asymptotic t-ratios)
Wage variance

constant (σ_0)	-1.2432 (19.43)	-1.1245 (17.97)
schooling (σ_1)	0.0034 (0.93)	0.0044 (1.88)
experience (σ_3)	-0.0154 (3.76)	-0.0152 (2.01)
$\text{cov}(\varepsilon_{it}, \text{hazard})$	0.1145 (1.95)	0.4321 2.87)

Appendix 1-The Data

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

The original sample contained 3,790 white males. However, we lacked information on family background variables (such as family income as of 1978 and parents' education). We lost about 17% of the sample due to missing information regarding family income and about 6% due to missing information regarding parents' education. The age limit and missing information regarding actual work experience further reduced the sample to 1,710.

Descriptive statistics for the sample used in the estimation can be found in Table 1. The education length variable is the reported highest grade completed as of May 1 of the survey year and individuals are also asked if they are currently enrolled in school or not.²¹ 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).²² 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

²¹This feature of the NLSY implies that there is a relatively low level of measurement error in the education variable.

²²Overall, interruptions tend to be quite short. Almost half of the individuals (45 %) who experienced an interruption, returned to school within one year while 73% returned within 3 years.

relatively small.

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

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

Table A1 - Descriptive Statistics

	Mean	St dev.	# of individuals
Family Income/1000	36,904	27.61	1710
father's educ	11.69	3.47	1710
mother's educ	11.67	2.46	1710
# of siblings	3.18	2.13	1710
prop. raised in urban areas	0.73	-	1710
prop. raised in south	0.27	-	1710
prop in nuclear family	0.79	-	1710
AFQT/10	49.50	28.47	1710
Schooling completed (1990)	12.81	2.58	1710
# of interruptions	0.06	0.51	1710
duration of interruptions (year)	0.43	1.39	1710
wage 1979 (hour)	7.36	2.43	217
wage 1980 (hour)	7.17	2.74	422
wage 1981 (hour)	7.18	2.75	598
wage 1982 (hour)	7.43	3.17	819
wage 1983 (hour)	7.35	3.21	947
wage 1984 (hour)	7.66	3.60	1071
wage 1985 (hour)	8.08	3.54	1060
wage 1986 (hour)	8.75	3.87	1097
wage 1987 (hour)	9.64	4.44	1147
wage 1988 (hour)	10.32	4.89	1215
wage 1989 (hour)	10.47	4.97	1232
wage 1990 (hour)	10.99	5.23	1230
Experience 1990 (years)	8.05	11.55	1230

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

Appendix 2- An Alternative Identification Strategy

In the most general model estimated above, identification of non-linearities is based on an unknown spline function parameter estimated at the same time as individual specific terms explaining the baseline returns to high school training. It may be argued that identification is achieved through a particular parametric form. One alternative possibility is to use observed regressors in order to identify non-linearities. This is achieved by assuming that

$$\varphi_{i2} = \exp(X_i'\delta_2)$$

While this specification has the unappealing feature that imposing skill homogeneity in the slopes automatically results in imposing linearity, it is still an appealing specification for investigating the robustness of the results. I have re-estimated the model with this particular specification and found practically identical results. For instance, the returns to high school education is 0.0401 while the return to post high school training averages 0.0810. The average returns to experience was found to be 0.0616.²³

²³These results may be obtained upon request from the author.