Separating risk from heterogeneity in education: a semiparametric approach[*]

Jacopo Mazza[†]
University of Essex

Hans van Ophem[‡]
University of Amsterdam

October 2, 2016

Abstract

Returns to education are variable both within and between educational group. If uncertain payoffs are a concern to individuals when selecting an education, wage variance is relevant. The variation is a combination of unobserved heterogeneity and pure uncertainty or risk. The first element is known to the individual, but unknown to the researcher, the second is unknown to both. As a result, the variance of wages observed in the data will overestimate the real magnitude of educational uncertainty and the impact that risk has on educational decisions. In this paper we apply a semiparametric estimation technique to tackle the selectivity issues. This method does not rely on distributional assumptions of the errors in the schooling choice and wage equations. Our results suggest that risk is decreasing in schooling. Private information accounts for a share varying between 0% and 13% of total wage variance observed depending on the educational level. Finally, we conclude that the estimation results are very sensitive to the functional relation imposed on the error structure.

Keywords: Wage inequality; Wage uncertainty; Unobserved heterogeneity; Variance differential; Selection bias, Return to education, Semiparametric estimation

JEL classification: C14; C24; D81; J31

[*]All data, computer programs and additional estimation results are available on request. We would like to thank Peter Backus, James Banks, Matias Cortes, Joop Hartog, an anonymous associate editor and two anonymous referees for valuable comments.

[†]Address: Department of Economics, University of Essex, Wivenhoe Park, Colchester, UK. Corresponding author: jacmazza@gmail.com

[‡]Address: Amsterdam School of Economics, Roetersstraat 11, 1018WB Amsterdam, The Netherlands

1 Introduction

Empirical evidence suggests that earnings inequality has increased in the US in the second half of the past century (Katz and Autor, 1999; Angrist et al., 2006). This observation involves both between and within educational group inequality (Acemoglu, 1999; Autor et al., 2006) and has attracted attention to the link between wage variance and schooling. If the variance of wages increases with schooling level and individuals are risk-averse, the increasing between educational group wage differences might reflect compensation for risk.

The identification of the causal effect of risk on education attainments is complicated by selection biases (Acemoglu, 2002). Observed wage inequality has to be calculated from truncated wages distributions. The truncation is an effect of private information: individuals possess information about their tastes and inclinations and will use this information to select their level of education assuring them the best risk/pay-off profile. Since this private information is not observed, researchers have to rely on the revealed schooling choices and observed wages and they should not make the mistake to confuse total observed variance with risk. In our terminology *risk* is the part of wage variability which is not foreseeable by the individual and the researcher even with the superior knowledge of private information. *Unobserved heterogeneity,* instead, is that part of wage variability that depends on factors known to the individual, but which are not observable by the researcher. Neglecting to disentangle risk and heterogeneity will cause an overestimation of risk and, in turn, an underestimation of risk premium offered in the labor market in the form of higher wages. The sum of risk and unobserved heterogeneity forms what we refer to as *wage dispersion.*

Understanding what is predictable and what is unknown to the individuals at the time of the schooling decision has obvious implications for the human capital literature. Disentangling what can or cannot be anticipated by individuals is also highly relevant for researchers interested in optimal taxation, saving decisions or consumption smoothing, to name a few[1].

In this paper we develop a two stage semiparametric estimation method able to estimate the importance of uncertainty in measured income variability and to disentangle it from predictable variability, allowing us to determine how much of income variability is due to luck and risk and how much is forecastable at the time of schooling decision.

Empirical investigations on this subject are relatively scarce and lead to substantially different conclusions. Cunha, Heckman and Navarro (2005) (CHN henceforth) and Cunha and Heckman (2007) (CH henceforth) investigate the effect of risk and unobserved heterogeneity on high school graduate's decision to proceed or not to college. They find that around 60% of income variability is predictable by the individuals *ex-ante.* Even more substantial is the share of income that Keane and Wolpin (1997) (KW henceforth) estimate to be foreseeable: about 90%. Chen (2008), on the other hand, finds substantially different results: unobserved heterogeneity is much smaller for both high school (7%) and college graduates (18%)[2]. In the replication of Chen's paper in Mazza et al. (2013) different figures are found: a very substantial unobserved heterogeneity, around 69% for high school graduates and a very modest figure of 10% for college graduates [3].

[1]For a more detailed discussion see, for example, Banks and Diamond (2010)

[2]Be aware that the table in the published paper of Chen is incorrect. At a later date she published a corrected version on her website. It is this version that we consider here.

[3]In Mazza et al. (2013) also UK and German data are considered. For these countries unobserved heterogeneity is

These different estimates are even more striking if we consider that all the mentioned papers exploit the same data-set: the National Longitudinal Survey of Youth (NLSY). There might be several reasons for these diverging conclusions. For one, even though the data source is homogeneous among all these works, the actual samples used differ. Both CHN and KW concentrate their analysis on short panels of white males, while Chen extends her sample in both dimensions by including African Americans and Hispanic minorities and by stretching the panel covering more time periods. An alternative reason can be the difference in the estimation methods employed. Although the starting point of all investigations mentioned is the Roy (1951) model, CHN and CH use a nonparametric estimation technique whereas Chen uses a fully parametric technique and restricts all parameters to be constant across educational categories. By exploiting a more restrictive model, Chen is able to estimate risk and unobserved heterogeneity parameters for the full distribution of educational attainments from high school dropouts to college graduates. On the other hand, both non parametric estimates of CHN and CH are limited to the binary choice between opting in or out of college.

As we explain below, Chen establishes a causal relation between education and wage dispersion, estimates a proper measure of risk for all educational categories and disentangles proper risk from unobserved heterogeneity. In her model, dispersion of wages is the outcome of interest and she applies the standard parametric selection model *a la* Heckman (1979) to the polychotomous case. We apply the same formalization, but we depart from it in an essential aspect: we do not impose normality on the distribution of disturbances. On top of that, we allow for different wage equations across educational levels. Our estimation method allows us to produce estimates for the full set of possible educational choices as in Chen and, at the same time, relax the normality assumption as in CHN and CH.

Parametric methods have undergone increasing criticism for imposing excessive restrictions on a model (Goldberger, 1983; Vella, 1998; Moretti, 2000). The main issue being that incorrect specification of the joint distribution of errors of wage and selection equation may lead to inconsistent estimates. In response researchers (Robinson, 1988; Cosslett, 1991; Ahn and Powell, 1993; Newey, 2009) have developed alternative approaches relying often on semiparametric estimation methods.

In this paper we contribute to this stream of literature by identifying the model parameters using a semiparametric estimation method. Our method consists of three steps: (i) the wage equation is estimated by exploiting the panel structure of the data. From this we retrieve the residuals necessary in step (iii). (ii) Conditional moments of the schooling choice distribution are estimated using the semiparametric estimation method proposed by Gallant and Nychka (1987) (GN henceforth) and its application to polychotomous choice as in Stewart (2004). (iii) The residuals of step (i) are related to the conditional moments estimated in step (ii) to identify the remaining parameters of our model. The estimates retrieved by this method are robust to misspecification of error distribution functions and allow us to minimize the distributional assumptions.

Semiparametric methods so far proposed in the literature tackle either self-selection or unobserved heterogeneity. Chen and Khan (2007) use kernel weighting schemes and symmetry conditions on the joint distribution of outcome and selection equation errors obtaining estimates for wage inequality among college graduates corrected for selection, but do not examine the impact of unobserved heterogeneity. Abadie (2002) proposes a method based on instrumental variables concerned with estimation

virtually not existent.

of causal effects on the entire distribution and not only mean effects, while Abadie et al. (2002) propose a generalization of the quantile treatment effect estimator for the case that selection into treatment is endogenous where the first step of the econometric model is estimated non parametrically. Abadie (2002) and Abadie et al. (2002) do not distinguishing between intrinsic heterogeneity and uncertain shocks.

The contributions of this paper are multiple. Methodologically, we show how risk and unobserved heterogeneity can be decomposed also in the polychotomous setting by applying a relatively straightforward and established semi-parametric method. This will allow us to provide evidence on the causal effect of education on income variability and estimate the level of uncertainty all across the schooling distribution. Furthermore we show how parameters estimated with a two step selection model are greatly affected by the assumption imposed on the error structure. By estimating three different specification for our model, one assuming joint normality, the second relaxing the joint normality assumption, but maintain a linear relationship between the error terms in the choice and outcome equation and a third flexible specification able to relax both assumptions, we are able to demonstrate how semiparametric methods relying on linearity in the error terms offer little advantages over the parametric counterparts and that more than the assumption of joint normality is the assumption of linearity that seems to play a crucial role for the estimated parameters in our setting.

With this work we hope to shed some light on the possible reason for discrepant results encountered in the literature by investigating whether the use of a fully parametric estimation technique is a determining factor.

Our major empirical finding is that only a minimal share of future income variability can be predicted by the individual when selecting into education. The only exception to this result is for high school graduates. By estimating the parametric counterpart *a la* Chen on our data and compare it directly to our semiparametric specification, we are able to show that the extent of selectivity detected is clearly affected by the distributional assumptions imposed. Our estimates place the amount of private information acted upon below the very substantial one encountered by non-parametric estimation methods (Cunha et al., 2005; Cunha and Heckman, 2007), but even below the very minimal one found in previous parametric estimates (Chen, 2008; Mazza et al., 2013).

2 Econometric specification

To identify the magnitude of risk in each education and its impact on individual choices and wages, two obstacles have to be bypassed: a) observed wage dispersion is not the correct quantification of true wage dispersion due to self-selection and b) even if we were able to correct for self-selection the corrected wage dispersion would still pool risk and unobserved heterogeneity together. Chen (2008) offers a solution to both these identification puzzles by dividing the problem into two parts. First the wage inequality corrected for self-selection is identified and then the unobserved heterogeneity is separated from risk. The model exploits the panel structure of NLSY to control for time invariant individual fixed effects, as for example taste for education, which are not observable and therefore could bias estimates if they are not accounted for.

In the remainder of this section we first consider the general setup of the model, then the consequences of this model from the individual view point and lastly, the consequences of the model from the view point of the researcher or observer.

2.1 General setup

The model presented in Chen is an extension of a classical Roy model (Roy, 1951) with four possible choices, in which the choice for "occupation" is substituted with a choice for educational level. In this model individuals (i = 1,..., N) have four possible schooling choices (s_i): no high school diploma ($s_i = 0$); high school diploma ($s_i = 1$); some college ($s_i = 2$); and four years of college or more ($s_i = 3$). Individuals are observed in $t = 1, ..., T$ periods; each time period is indexed by subscript (t). The total number of individuals in the sample will be indicated by N.

For each individual i we will observe one wage y_{it} for each time period t given and the time invariant educational level s_i. Which of the four possible wages is observed is determined by the relation:

$$y_{it} = y_{0it}I(s_i = 0) + y_{1it}I(s_i = 1) + y_{2it}I(s_i = 2) + y_{3it}I(s_i = 3),$$

where $I(.)$ is the indicator function equal to one if that particular schooling level is selected and zero otherwise.

The potential wage (y^*_{sit}) is a latent variable and represents the wage that we would observe in each category if the subject would have chosen that particular educational level. In other words, the potential wage is the hypothetical wage that the individual would earn if one of the other three counterfactual educational levels had been chosen. We assume that it is determined by a linear model:

$$y^*_{sit} = \alpha_s + x_{it}\beta_s + \sigma_s e_{si} + \psi_{st}\epsilon_{it} \text{ only observed if } s_i = s, \quad (1)$$

α_s is a schooling specific constant, β_s is a schooling specific vector of coefficients for the vector of observable covariates x_{it}, the individual fixed effect, capturing unobserved earning potential at schooling level s, is represented by the time invariant term $\sigma_s e_{si}$, while and the error term $\psi_{st}\epsilon_{it}$ denotes transitory shocks uncorrelated with personal characteristics and across time.[4] The transitory shocks incorporate institutional or macroeconomic shocks uncorrelated with the individual fixed effects. The e_{si} and ϵ_{it} are random variables with zero mean and variance 1. Inequality in potential wages within schooling levels equals $\sigma_s^2 + \psi_{st}^2$: the sum of a permanent component created by variation in individual specific effect and a transitory component.

Individuals first select into an education according to their personal tastes and inclinations, in the second step their wage is revealed and they earn a wage dependent on their schooling choice. Specifically, we observe the wage y_{it}. The assignment to one of the four schooling categories is governed by the rule:

$$s_i = s \text{ if } a_{si} \leq \sigma_\nu \nu_i < a_{s+1,i} \text{ for } s = 0, 1, 2, 3. \quad (2)$$

[4]In the empirical part of Chen it is assumed that β_s is constant across schooling levels, apart from the constant. We allow for the more general case also in our estimations.

In this expression ν_i is the unobserved schooling factor known to the individual. It includes taste for education, motivation and all other factors influencing the educational choice of the individual. ν_i is unobservable to the researcher. a_{si} is the minimal or maximum level of ν_i for individuals that choose schooling level s and it is determined by the relation:

$$a_{si} = \kappa_s - z_i\theta. \tag{3}$$

The vector z_i contains time invariant observable characteristics, θ is the vector of coefficients for z_i and κ_s (with s =0, 1, 2, 3) are constants with $\kappa_0 = -\infty$ and $\kappa_4 = \infty$, respectively. We assume that $E[\epsilon_{it}|\nu_i, x_{it}, z_i] = 0$ and $Var[\psi_{st}\epsilon_{it}|\nu_i, x_{it}, z_i] = \psi_{st}^2$. ν_i is allowed to be correlated with the fixed effect e_{si}. Furthermore, it is assumed that $E[\nu_i|z_i] = 0$ and $Var[\nu_i|z_i] = 1$.

To disentangle the share of wage variance due to risk from that caused by unobserved heterogeneity, we rely on some additional assumptions regarding the disturbances in the outcome and selection equation. In the literature on risk and heterogeneity the error process is often modeled in a linear and additive fashion (Carneiro et al., 2003; Cunha et al., 2005; Cunha and Heckman, 2007; Chen, 2008). Although, it has been noted (Vella, 1998) that semiparametric models relying on a linear structure in the error terms offer little advantage over their parametric counterpart. Therefore, we do not impose linearity and we allow the relation between e_{si} and ν_i to be nonlinear, although the part of e_{si} not correlated to ν_i (ξ_{si}) is additive. Note that the assumption of a linear relation (i.e. $\sigma_s e_{si} = \rho_s\sigma_s\nu_i + \sigma_{\xi s}\xi_{si}$), is a special case of our specification. Such a linear relation conforms closely with the assumption of a normal distributed error term in the wage equation.

In the empirical part of this paper we compare two different semiparametric specifications of our model relying and not relying on linearity and we show how results are considerably affected by this assumption. In order to be able to separately identify risk and unobserved heterogeneity, we specify the relation between the individual fixed effect in the wage regression and the unobserved schooling factor as:

$$\sigma_s e_{si} = \rho_s\sigma_s g(\nu_i) + \sigma_{\xi s}\xi_{si}, \tag{4}$$

with $g(\nu_i)$ an unknown function with variance 1, $E[\xi_{si}|\nu_i, x_{it}, z_i] = 0$, $Var[\sigma_{\xi s}\xi_{si}|\nu_i, x_{it}, z_i] = \sigma_{\xi s}^2$ and $Cov[\sigma_s e_{si}, \sigma_g g(\nu_i)] = Cov[\rho_s\sigma_s g(\nu_i) + \sigma_{\xi s}\xi_{si}, \sigma_g g(\nu_i)] = \rho_s\sigma_s\sigma_g$. Similar as before, the variance of $g(\nu_i)$ is put to 1, so that σ_g^2 is the relevant variance. In essence, this equation captures the idea that unobserved schooling component and unobserved earning potential are correlated, where the correlation between the two is indicated by ρ_s, but since this correlation, whose functional form we leave unspecified, is not perfect, a random term ($\sigma_{\xi s}\xi_s$) is added to the relation.

At this point it is useful to consider the relation between the variance of the fixed effects and the parameters introduced:

$$Var[\sigma_s e_{si}] = \sigma_s^2 = Var[\rho_s\sigma_s g(\nu_i) + \sigma_{\xi s}\xi_{si}] = \rho_s^2\sigma_s^2 + \sigma_{\xi s}^2. \tag{5}$$

From this it follows that:

$$\sigma_{\xi s}^2 = \sigma_s^2(1 - \rho_s^2). \tag{6}$$

From this relation we can conclude that potential wage inequality can only be bigger or equal to the observed one.

2.2 Individual point of view

From the individual viewpoint the expected value of future wages is given by:

$$E[y_{sit}|\nu_i, x_{it}, z_i] = \alpha_s + x_{it}\beta_s + \rho_s\sigma_s g(\nu_i). \tag{7}$$

We assume that the individual knows $g(.)$ and ν_i. This assumption explains the difference between the individual's and the researcher's view point: the researcher only knows that $\sigma_\nu\nu_i$ is in an interval (cf. equation 2), and therefore the conditioning has to be on an interval in our empirical analysis. This decomposition of expected wages introduces an important feature of the model. When selection is positive (i.e.: $\rho_s > 0$) the labor market rewards workers with a high taste for education whilst the opposite occurs when selection is negative (i.e.: $\rho_s < 0$).

Since individuals possess a more accurate assessment of their own abilities than researchers, private information ($\sigma_\nu\nu_i$) has to be accounted for in order to build a true measure of risk. Risk about wage per schooling level is the variance of permanent and transitory component from the individual viewpoint that needs to be separated from unobserved heterogeneity. Following Chen, we indicate this risk with τ^2_{st}. Using equation (4) and our assumptions about the moments of the disturbances given above, we obtain an expression for risk as the variance of the error term in (1) given observed and unobserved heterogeneity:

$$\tau^2_{st} = Var[\sigma_s e_{si} + \psi_{st}\epsilon_{it}|\nu_i, x_{it},, z_i] = \sigma^2_{\xi s} + \psi^2_{st} = \sigma^2_s(1-\rho^2_s) + \psi^2_{st}. \tag{8}$$

If we look at equation (5) we note that, if we condition on ν_i the individual randomness only comes from ξ_{si}. As the extent to which wage dispersion is predictable from the personal standpoint is expressed by the correlation coefficient ρ_s, equation (8) makes the formal link between wage dispersion and private information explicit. In fact, if the correlation between unobserved schooling factor ($g(\nu_i)$) and permanent component of wage dispersion (e_{si}) is perfect (i.e.: $\rho_s = \pm 1$) the individual can predict exactly the permanent part of the wage variability and risk is only caused by transitory shocks (ψ^2_{st}). Alternatively, if $\rho_s = 0$ then the individual does not possess any additional information compared to the researcher and risk is directly observed in the data.

Rearranging equation (8) as $\tau^2_{st} + \rho^2_s\sigma^2_s = \sigma^2_s + \psi^2_{st}$ helps visualizing how potential wage dispersion ($\sigma^2_s + \psi^2_{st}$) is the sum of two elements: the variance of unobserved heterogeneity ($\rho^2_s\sigma^2_s$) and risk (τ^2_{st}). Note also that if correlation between schooling and unobserved tastes for education exists (i.e.: $\rho_s \neq 0$) potential wage dispersion overstates the real degree of risk ($\tau^2_{st} < \sigma^2_s + \psi^2_{st}$).

2.3 Researcher's point of view

We next consider the model from the researcher's point of view. Contrary to the individual situation ν_i is unknown, but the research does observe the chosen level of schooling and this implies, according to our model (cf. 2): $s_i = s$ if $a_{si} \leq \sigma_\nu\nu_i \leq a_{s+1,i}$. This is the usual sample selection model with ordered censoring rules. From the assumptions expressed in (4) we obtain that:

$$E[\sigma_s e_{si} + \psi_{st}\epsilon_{it}|s_i = s] = \rho_s \sigma_s E[g(\nu_i)|s_i = s] \quad (9)$$

and

$$Var[\sigma_s e_{si} + \psi_{st}\epsilon_{it}|s_i = s] = \rho_s^2 \sigma_s^2 Var[g(\nu_i)|s_i = s] + \sigma_{\xi s}^2 + \psi_{st}^2. \quad (10)$$

Ignoring the transitory shocks, substituting (6) into (10) and rearranging, we obtain the variance of observed wages corrected for truncation:

$$Var[\sigma_s e_{si}|s_i = s] = \rho_s^2 \sigma_s^2 Var[g(\nu_i)|s_i = s] + \sigma_s^2(1 - \rho_s^2) = \sigma_s^2 \left(1 - \rho_s^2(1 - Var[g(\nu_i)|s_i = s]\right), \quad (11)$$

where $0 \leq 1 - Var(g(\nu_i)|s_i = s) \leq 1$ and as a result:

$$0 \leq 1 - \rho_s^2(1 - Var[g(\nu_i)|s_i = s]) \leq 1. \quad (12)$$

In our model the error term is composed by two elements, the permanent component ($\sigma_s e_{si}$), for which we have explicated the variance in (10), and the transitory shocks $\psi_{st}\epsilon_{it}$. The expression for the variance of the complete error term is:

$$Var[\sigma_s e_{si} + \psi_{st}\epsilon_{it}|s_i = s] = Var[\sigma_s e_{si}|s_i = s] + \psi_{st}^2 = \sigma_s^2 \left(1 - \rho_s^2(1 - Var[g(\nu_i)|s_i = s]\right) + \psi_{st}^2 \quad (13)$$

and this variance is smaller or equal than $\sigma_s^2 + \psi_{st}^2$ because of (12). We now turn to the estimation of the model and the identification of the parameters.

3 Semiparametric estimation and identification

A fertile line of research (Cosslett, 1983; Robinson, 1988; Powell, 1989; Cosslett, 1991; Ahn and Powell, 1993; Dahl, 2002; Newey, 2009) has produced semiparametric methods to correct for self-selection with more limited reliance on distributional assumptions. Generally all these methods imply a two-step approach, with a specified selection and structural equation and generic selection correction function and error term density. The assumption that those methods usually imply is:

$$E[\sigma_\nu \nu_i|s_i = s; x_i, z_i] = E[\sigma_\nu \nu_i|a_{si} \leq \sigma_\nu \nu_i < a_{s+1,i}; x_i, z_i] = h_s(z_i\theta)$$

,with $h_s(.)$ some unknown function. The semiparametric approach differs from the parametric one in two important dimensions: a) no distributional assumption on ν_i is specified and b) no assumptions on the joint distribution of the error terms in the selection and outcome equation are made when estimating $E[\sigma_s e_{si} + \psi_{st}\epsilon_{it}|s_i = s]$.

Our starting points is the following factorization of the error term:

$$\sigma_s e_{si} + \psi_{st}\epsilon_{it} = \rho_s \sigma_s g(\nu_i) + \sigma_{\xi s}\xi_{si} + \psi_{st}\epsilon_{it},$$

where both ϵ_{it} and ξ_{si} are independent of ν_i. β_s from (1) can be consistently estimated by the fixed effect model. It eliminates $\sigma_s e_{si}$ by estimating:

$$(y_{sit} - \bar{y}_{si}) = (x_{it} - \bar{x}_i)\beta_s + (\psi_{st}\epsilon_{it} - \overline{\psi}_s \bar{\epsilon}_i), \tag{14}$$

where $\bar{y}_{si}$, $\bar{x}_i$, $\overline{\psi}_s$and $\bar{\epsilon}_i$ denote the average over time of the corresponding variables. Chen shows that the transitory component ψ_{st} of the wage inequality is identified through the variance of the error term in equation (14). Estimates of α_s (s=0,1,2,3) are given by:

$$\hat{\alpha}_s = \frac{1}{N_s T}\sum_{t=1}^{T}\sum_{i=1}^{N_S}(y_{sit} - x_{it}\hat{\beta}_s). \tag{15}$$

As a result we can estimate $(\sigma_s e_{si} + \psi_{st}\epsilon_{it} | s_i = s)$ in the following way:

$$\sigma_s e_{si} \widehat{+ \psi}_{st}\epsilon_{it} = y_{sit} - \hat{\alpha}_s - x_{it}\hat{\beta}_s \tag{16}$$

for every $i = 1, ..N; t = 1, .., T$ and for given s.

We assume that $\sigma_g g(\nu_i)$ can be approximated by a polynomial of order Q:

$$\sigma_g g(\nu_i) = \sum_{j=0}^{Q}\tilde{\alpha_j}\,(\sigma_\nu \nu)^j. \tag{17}$$

By increasing Q the function $g(\nu_i)$ can be approximated to any precision required. Note that by setting $Q = 1$ this assumption reduces to a linear relation i.e. $\sigma_s e_{si} = \rho_s \sigma_s g(\nu_i) + \sigma_{\xi s}\xi_{si}$. Such a linear relation closely conforms with the assumption of a normal distributed error term in the wage equation. From assumption (17) we get:

$$E[\sigma_g g(\nu_i)|s_i = s] = \sum_{j=0}^{Q}\tilde{\alpha_j} E\left[(\sigma_\nu \nu_i)^j | s_i = s\right]. \tag{18}$$

As we will show below these conditional expectations $E\left[(\sigma_\nu \nu_i)^j | s_i = s\right]$ can be consistently estimated for every individual and for every level of schooling. In the usual econometric fashion, we now specify:

$$\begin{aligned}(\sigma_s e_{si} + \psi_{st}\epsilon_{it}|s_i &= s) = E[\sigma_s e_{si} + \psi_{st}\epsilon_{it}|s_i = s] + \varsigma_{sit} &(19)\\ &= \frac{\rho_s \sigma_s}{\sigma_g}\sum_{j=0}^{Q}\tilde{\alpha_j} E\left[(\sigma_\nu \nu_i)^j | s_i = s\right] + \varsigma_{sit}. &(20)\end{aligned}$$

So we can fit the $Q + 1$ conditional moments (including the 0-moment which corresponds to $Pr[s_i = s]$, $s = 0, 1, 2, 3$) to the residuals given s as described in (16). As shown in the appendix, from this relation we can identify all the parameters of the error structure except for the sign of the correlation ρ_s.

It is worth nothing that for the identification of our parameters of interests, the sign of the correlation coefficient ρ_s is not needed as both the parameters capturing risk and unobserved heterogeneity are only dependent on the square of it.

To estimate the conditional expectations in (20) we estimate the schooling choice by adopting the semiparametric estimation strategy proposed by Gallant and Nychka (1987) . This estimator does not require assumptions about the distribution of the error term ν_i in the selection equation to estimate θ. The underlying idea of this methodology is to approximate the true density by the product of an order K series of polynomials and a normal density. In this way, many different features of the unknown density - the density itself, its mean, variance and higher moments - can be consistently estimated. The approximation is specified as:

$$f_K(\nu) = \frac{1}{\eta} \sum_{k=0}^{2K} \iota_k^* \nu^k \phi(\nu), \tag{21}$$

where η is defined such that the density integrates to 1.

GN show that estimates of θ are consistent provided that the order of polynomials K increases with sample size. The non-parametric feature is that the number of terms can increases to infinity with the number of observations, but at a slower rate. We focus on the practical application of the method and not on its asymptotic properties if the number of polynomial expansions tends to infinity. Thus, as in van Soest et al. (2002), we work under the assumption that the length of the series approximation is given. As a result standard properties of (parametric) maximum likelihood apply. The choice of K is a standard model selection problem that we tackle by applying two different model selection criteria: the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) and chose the preferred one according to these measures (van Soest et al., 2002; Stewart, 2004). In principle any moment generating density other than the normal could be used; the normal density is a convenient choice since this form nests the ordered probit model which becomes a special case with $K = 1$ and $K = 2$ (Stewart, 2004)[5]. Using the estimation method of GN will yield consistent estimates of θ, the interval limits κ_2 and κ_3 and the variance of the error term ν_i (σ_ν^2). Since the functional form of the density of the error term is estimated we can numerically approximate the truncated first and second moments and the variance of the error term from the following expressions:

$$E[\sigma_\nu \nu_i | s_i = s] = \frac{\int_{a_{si}}^{a_{s+1,i}} \sigma_\nu \nu_i f_K(\sigma_\nu \nu_i) d\nu}{Pr[a_{si} \leq \sigma_\nu \nu_i \leq a_{s+1,i}]}, \tag{22}$$

$$E[(\sigma_\nu \nu_i)^2 | s_i = s] = \frac{\int_{a_{si}}^{a_{s+1,i}} (\sigma_\nu \nu_i)^2 f_K(\sigma_\nu \nu_i) d\nu}{Pr[a_{si} \leq \sigma_\nu \nu_i \leq a_{s+1,i}]}, \tag{23}$$

or

[5] Ordered response models are structurally unidentified because we only observe an unordered discrete dependent variable. Usually, the following identification restrictions are imposed: (i) one of the interval limits to be estimated is set to 0 or alternatively no constant is used in the specification and (ii) the variance of the error term is set to unity. We follow (Stewart, 2004), and use different identification restrictions to ease the estimation: no constant is used and the first interval limit is fixed to its ordered probit estimate. As a result we do estimate the variance of the error term (σ_ν^2 in the present setting).

$$Var[\sigma_\nu \nu_i | s_i = s] = E[\sigma_\nu^2 \nu_i^2 | s_i = s] - E[\sigma_\nu \nu_i | s_i = s]^2. \quad (24)$$

Higher order conditional moments (order $j > 2$) can also be estimated using[6]:

$$E[(\sigma_\nu \nu_i)^j | s_i = s] = \frac{\int_{a_{si}}^{a_{s+1,i}} (\sigma_\nu \nu_i)^j f_K(\sigma_\nu \nu_i) d\nu}{Pr[a_{si} \leq \sigma_\nu \nu_i \leq a_{s+1,i}]}. \quad (25)$$

We have now discussed how to identified all parameters of the model without assuming (joint) normality of the error terms in the wage and choice equations. The only two assumptions that we need to establish identification are:

1. the error term of the wage equation is given by: $\sigma_s e_s + \psi_{st}\epsilon_{it} = \rho_s \sigma_s g(\nu_i) + \sigma_{\xi s}\xi_{si} + \psi_{st}\epsilon_{it}$; where $g(.)$ is left unspecified;

2. $E[\epsilon_{it}|\nu_i, x_{it}, z_i] = 0$, $Var[\psi_{st}\epsilon_{it}|\nu_i, x_{it}, z_i] = \psi_{st}^2$, $E[\xi_{si}|\nu_i, x_{it}, z_i] = 0$ and $Var[\sigma_{\xi s}\xi_{si}|\nu_i, x_{it}, z_i] = \sigma_{\xi s}^2$.

Bootstrapping will be used to estimate the standard errors of all estimated parameters.

4 Data

We use the National Longitudinal Survey of Youth 1979 (NLSY79) to estimate the parameters of interest. The survey is a widely exploited data set of 12,686 young American citizens who were 14 to 22 years old in 1979. The participants in the survey were interviewed annually from 1979 until 1994 and biennially from then on. NLSY79 provides information on schooling, labor market experiences, training expenses, family income, health condition, household composition, region of residence and environmental characteristics.

4.1 NLSY

The NLSY data comprise four samples: a random sample, an economically disadvantaged sample, a sample drawn from the military and a random sample of black and Hispanic populations. We limit our sample in several ways. First we base our estimations on the random sample only. Second, we restrict our analysis to men between the survey years 1991 and 2010 (calendar years 1990 to 2009) . We decide to consider only men as their labour supply decisions is not complicated by fertility and labour market participation issues. As the labour market participation of women is more elastic, if we were to include them in the analysis we wold have to first model their participation decisions, instead we decide to follow standard practice in the returns to education literature (Willis and Rosen, 1979; Angrist and Krueger, 1991; Card, 1993, 2001; Cameron and Taber, 2004) and in the literature more directly related to this work (Cunha et al., 2005; Cunha and Heckman, 2007; Chen, 2008) and restrict our analysis to men only. The wave restriction will allow us to focus on individuals already out of school and into the labor market. Third, we exclude 1,360 respondents (about 11% of the

[6] The numerical integration of these conditional moments is relatively straightforward in a software package like R. The computational burden is relatively modest (more than 3000 approximations in a few seconds) but results become increasingly imprecise as j increases.

entire sample) who do not provide information about parental education[7], highest grade completed, exact work experience history, and ability index as defined below. Fourth, we drop 133 individuals not providing any information on hourly rate of pay, our outcome variable. Fifth, since the information prior to 1978 is limited, we circumscribe our analysis to males aged between 13 and 18 as of that date, in order to rely on precise information about region of residence at 17. Census region of residence is essential for the construction of our instrument since we need it in order to construct measures of local labor market conditions. After having selected the sample according to these guidelines, we have a balanced panel sample consisting of 3,040 individuals observed in 12 subsequent waves.

Our dependent variables are two: schooling for the choice equation and earnings for the outcome equation. Schooling is measured as highest schooling level completed in 1990. From this information we construct four dummies for the highest educational achievement: no high school, high school, college drop outs and college graduates or beyond. Earnings are defined as the logarithm of hourly earnings in 1992 dollars. In our wage regression we exclude observations occurring before 1990 when individuals were between 25 and 30 years of age. In this way we ensure that all wages are recorded once the final schooling stage has been reached[8].

The control variables added both to the schooling and wage equations and presented in Table 1, are the highest education completed for both parents, the Armed Forces Qualification Test score (AFQT), the family income, the number of siblings and the ethnic origin. All these variables are meant to control for intrinsic ability and family background of the individual. To control for characteristics of the geographical area of origin we include a set of dummies for urban area and for the region of residence at 14 (Northeast, Mid-West, South or West).

The AFQT is a series of four tests in mathematics, science, vocabulary and automotive knowledge. The test was administered in 1980 to all subjects regardless their age and schooling level. For this reasons it can include age and schooling effects in the ability index that the test is meant to construct. To correct for these undesirable effects we follow a frequently used procedure first introduced by Kane and Rouse (1995) and then adopted by many others (Neal and Johnson, 1996, Cameron and Heckman, 2001, Hansen et al., 2004; Carneiro and Lee, 2009). First we regress the original test score on age dummies and quarter of birth, then we replace the original test score with the residuals obtained from this regression.

For family income we use family income at age 17. If no measure for family income at 17 is recorded, we replace it by family income at the closest age to 17 available.

To allow for a meaningful comparison of our results with previous parametric estimates our sample selection rules follow quite closely those adopted by Chen (2008). The only noticeable difference is the exclusion of the older cohorts of individuals from our sample due to lack of precise geographical location information prior to 1978 which is a necessary information for constructing one of our two instruments. The exclusion of the 1,004 individuals exceeding the imposed age threshold accounts almost completely for the discrepancy in the sizes between ours and Chen's samples. Sample selection

[7]Of these 1,360 individuals, 992 are dropped because of missing information for one (or both) parent's highest educational achievement. For 120 respondents this information is missing because the parent is unknown. The remaining 872 missing values are due to the respondent not knowing this information. For single parent families the value for the missing parent education is recorded as long as the individual knows it, otherwise the observation is dropped from our final sample.

[8]Starting our panel at an earlier stage would also create an unbalanced panel in which college graduates would be underrepresented since it would contain more observations for high school dropouts and graduates.

could pose a threat to the comparability of our results if older cohorts would systematically differ from younger ones in some unobservable manner correlating with productivity traits. In order to attenuate concerns of selectivity we estimate a simple linear probability model for a series of ability proxies and demographic characteristics on the probability of belonging to the excluded cohorts. Quite reassuringly, all variables proxying for ability have an insignificant relationship with the probability of being dropped from our sample. The only two demographics estimated to have a statistically non zero effect are mother education which is positively correlated, but whose effect at 0.6% for each additional year is fairly small and being of Hispanic ethnicity which decreases the probability of being excluded by 4%. Overall, we believe that these results do not point towards systematic differences able to account for the different estimates encountered in the two papers.

A direct comparison between the sample used in our analysis with that used by CHN, which is the other closely related work to the present one is more problematic. In fact, CHN restrict their sample to working white males who at least graduated from high school. A direct consequence of this choice is that all their estimates are based on a smaller NLSY sub sample than ours and Chen's. Intuitively, this stricter sample selection should eliminate a lot of the observed heterogeneity in the data. Furthermore, we might expect that this more homogeneous and probably privileged[9] group of individuals might have access to better information and therefore be in the condition of forecasting their future performance in the labor market more precisely. If this conjecture is true, we might expect that the higher level of predictability of future earnings on the students' part encountered in CHN might be, at least partially, driven by the particular sample selected.

Means and standard deviations of dependent and independent variables are given in Table 1. We see that distribution among the four educational category is quite equal, but a substantial share stopped after high school, African-Americans are overrepresented and the large majority of respondents was raised in a urban environment.

The covariates included in the first differenced wage equation , are experience in the labor market and its square and local unemployment. Work experience is here defined as the cumulative number of working weeks divided by 49: the amount of working weeks in a calendar year. In this way we transform work weeks in work years[10].

4.2 Choice of instruments

The two instruments that we exploit for identification of the conditional moments are the unemployment rate in the region of residence at age 17 and an indicator for whether the individual was born in

[9]Just to form an idea of the socio-economic intra ethnic differences existing in the NLSY sample, in our sample Blacks' and Hispanics' AFQT average score is 53% and 67% that of Whites respectively, while the family income gap is approximately 53% and 29% in favor of Whites when compared to Blacks and Hispanic respectively.

[10]A possible concern of including work experience in a wage regression is endogeneity and in order to overcome the potential bias, in the returns to schooling literature, it is common practice to replace actual work experience with potential experience defined as age minus schooling years minus six. Unfortunately, it is not clear whether this procedure is to be preferred. In fact, if schooling is endogenous and potential experience depends directly on it, potential experience would be endogenous as well. Furthermore, some authors (Oaxaca and Regan, 2009; Blau and Kahn, 2013) argue that replacing potential for actual work experience could introduce additional bias in the estimated coefficients. For these reasons and since the paper more closely related to ours (Chen) controls for actual experience in the wage regression, we have decided to follow in her steps. Nonetheless, as a robustness check, we have performed our estimation replacing actual with potential work experience. Our main results are not sensitive to this different specification. Results available on request.

Table 1: Summary statistics

Time Invariant Variables			
(a) Schooling Variables		Black	.23
Years of schooling	12.79		(.42)
	(2.42)	Hispanic	.16
Categorical education:			(.37)
No high school	.17	*(c) Geographic Controls at age 14*	
	(.38)	Urban	.78
High school	.44		(.41)
	(.50)	Residence in Northeast	.19
Some college	.21		(.39)
	(.40)	Residence in South	.35
Four year college or beyond	.20		(.47)
	(.40)	Residence in West	.17
(b) Ability and Family Background			(.38)
Armed Forces Qualifying Test score (adjusted)	50.22	Residence in Mid-West	.28
	(28.02)		(.45)
Highest grade mother	10.96	*(d) Instruments for Schooling*	
	(3.16)	Unemployment rate (%)	7.34
Highest grade father	10.96		(1.99)
	(3.87)	Month of Birth	6.71
Number of siblings	3.65		(3.30)
	(2.52)	Mean local unemployment over working life	6.50
Family income (1999 dollars)	41,752.07		(.47)
	(32,361.58)		

	Time Variant Variables				
Calendar year	1990	1994	2000	2004	2010
Actual work experience	7.97	10.75	14.83	17.25	20.40
	(3.09)	(3.97)	(5.62)	(6.78)	(8.57)
Log hourly earnings	2.14	2.39	2.74	2.95	3.09
	(.67)	(.66)	(.82)	(.81)	(.82)

Note: Standard deviations in parentheses. Unemployment rates calculated from CPS data.

the first or in the second half of the year[11]. Having an overidentified model allows us to test for the validity of the combination of the instruments selected via a Sargan-Hansen test. We will present the results for this test in section 5.1.

Information about unemployment rates are taken from the Current Population Survey (CPS) which is conducted by the American Bureau of Census for the Bureau of Labor Statistics on a sample of 50,000 American families each month for the last 50 years. The resulting index is differentiated by census macro areas - north-east, mid-west, south and west - corresponding to those available in the NLSY allowing us to match individuals with their region specific index. Unemployment rate is a good indicator for the general conditions in the economy that any new entrant in the labor market would have to face. It affects human capital accumulation via changes in the opportunity costs of investing into further education. Even though a large body of empirical evidence on the effect of unemployment rate on schooling decisions point towards a positive correlation for the two variables (Betts and McFarland, 1995; Clark, 2011), theoretically the effect is ambiguous. In fact, while we observe poor labor market conditions decreasing the opportunity costs of schooling encouraging schooling, they also lead to a decrease in the resources of credit constrained families causing the opposite effect. Unemployment rates have regularly been used as an instrument for schooling decisions (Cameron and Heckman, 2001; Carneiro et al., 2003, 2011). The unemployment rate of interest is usually the county or state level. Given the data restriction policy of the NLSY data administrators, we are forced to aggregate at the census region level[12]. Therefore, this instrument is constructed similarly to the established procedure with the only difference being the geographical unit of reference. This coarser aggregation level could spark some concerns for the relevance of our instrument. Fortunately first stage estimates show an effect strong enough to rule out concerns about weak instruments.

In Figure 1 we show the unemployment rate in each of the four census regions that our sample is divided into for the relevant years. From this figure we can appreciate the between and within cohort variation that we exploit for our identification. Evidently the first is more prominent than the latter.

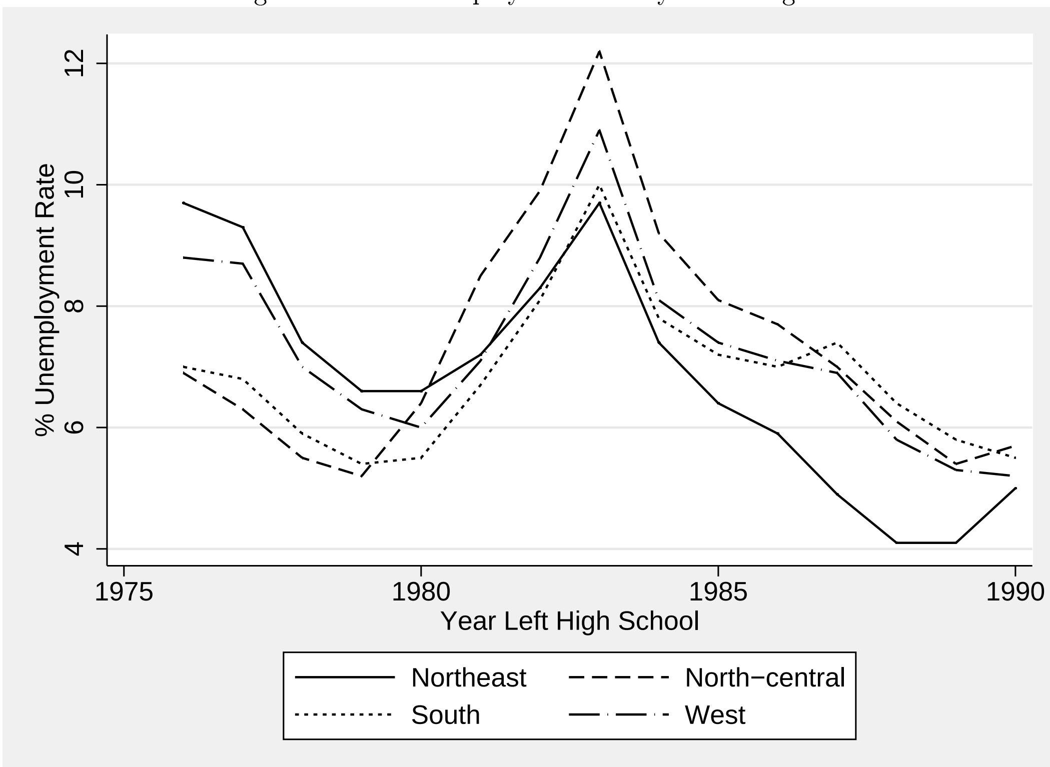
Previous research exploiting local economic conditions at the time of choice as an instrument emphasizes the importance of controlling for the current local conditions (Cameron and Heckman, 1998, 2001; Carneiro et al. 2003; Cameron and Taber 2004; Carneiro and Lee 2009). In fact, it is reasonable to doubt that local unemployment at 17 might correlate with unobservables in wage equations for the subsequent periods. If that is the case, the instrument would not be exogenous and therefore invalid. To mitigate this concern we follow Cameron and Taber (2004) and include a time-varying measure for unemployment rate in the current region of residence directly in the wage regression. Effectively we are using the innovations of this measure as instruments. It is worth noting that, on average, there are 8 years elapsing between the time when we observe our selection terms and when we first observe wages. The assumption is then that, conditional on current unemployment rate, past economic conditions have no effect on present wages.

The second instrument that we employ is an indicator variable for being born early in the calendar year. A growing literature on the determinants of schooling achievement has found that being among

[11] We are grateful to one anonymous referee for suggesting this instrument.

[12] We do not have access to the geocode data since the use is limited to resident researchers at American institutions.

Figure 1: U.S. Unemployment rate by macro region 1976-1990



the youngest in your cohort has negative effects on the likelihood of acquiring higher levels of education (Crawford et al., 2011, 2014). This is because younger children tend to perform worse in their first years of education as they are less mature. Falling behind in the first years of education might have long lasting effect on educational achievements if regularly being among the lowest achievers among your peers affects motivation and determination, or if it has an impact on ones self-esteem. Crawford et al. (2014) find that for a sample of UK students, the August born individuals, who are the youngest possible individuals in any given cohort in the UK setting, score sensibly lower than the oldest individuals within their cohorts in standardized tests. This effect is still detectable at age 16. In the US, the cutoff month for student enrollment is January so that all individuals born in the same calendar year are starting education in September of the year they are turning 6. Therefore, we create an indicator variable taking value one for individuals born in the first half of the year. If older individuals perform better at the beginning of their educational path due to their relative higher maturity and if this effect persists throughout the educational career, we expect this dummy variable to be positively correlated with educational achievements.

Our two instruments are orthogonal one from the other and therefore should operate via distinct channels on the decision to acquire additional schooling, should affect different people and obviously have different range of variation.

5 Empirical results

As described in Section 3, we first estimate the two stages of the selection model estimated on NLSY79 data and then we identify the key parameters of our model: the permanent component of wage dispersion (σ_s^2); the transitory component of wage dispersion (ψ_{st}^2); the variance of unobserved heterogeneity ($\gamma_s^2 \sigma_\nu^2$) and of risk (τ_{st}^2).

5.1 Selection of the preferred model and first stage

In the first stage we estimate the multivariate choice equation via the GN method discussed in Section 3. In this way we obtain estimates for the density function of the unobserved heterogeneity component and we can estimate the conditional moments required. From a purely theoretical standpoint, it is essential for the degree of polynomial K to increase with sample size. Only then the estimates are consistent. In practical applications, however, this is not possible. To select the best approximation we apply two standard methods for selection: AIC and BIC. The two methods differ on how steeply they penalize model complexity. AIC tends to penalize complexity less than BIC, thus if parsimony is important BIC should be the preferred criteria. In Table 2 we present the two criteria. On top of that we present LR-tests on the coefficients of the polynomial and on the ordered probit.

Table 2: Model comparison

K	log likelihood	LR-test of OP	p-value	LR-test of K=3	p-value	LR-test of K-1	p-value	AIC	BIC
OP	-2,856.53							5,763.06	5,913.60
3	-2,844.32	24.423	.000					5,742.63	5,905.22
4	-2,844.33	24.410	.000	.01	1.00	.01	1.00	5,744.65	5,913.26
5	-2,844.25	24.559	.000	.14	.93	.15	.70	5,746.50	5,921.13
6	-2,843.57	25.917	.000	1.49	.68	1.36	.24	5,747.14	5,927.79

We start from the 3rd degree polynomial since this is the first model generalizing the ordered probit (OP) to the semiparametric case. The results emerging from the three tests are unambiguous. All tests select the third degree polynomial expansion as the favored one, therefore we use this specification to estimate the first stage of our selection model[13]. The LR-test of the SNP models against the OP formally tests departure from normality. The results shown in table 2 indicate that the OP model is rejected in all cases demonstrating how we do observe significant differences between the fitted density and the standard normal.

The results of the ordered probit model and the GN procedure at 3rd and 4th degree polynomial are presented in Table 3[14]. It has to be noted that estimates of θ cannot be compared directly across models because the variances differ (Stewart, 2004). What we can compare are ratios of different

[13] In table 3 we show first stage estimates adopting also a 4th degree polynomial. From that table it is evident that first stage results are unaffected by the choice of the degree of expansion. We have experimented also with five and six degree polynomial expansions, but the results remain largely unchanged, the only noteworthy effect of increasing the order of the polynomial expansion is that of increasing the level of significance for the birth in the first half instrument.

[14] Estimation of the 3rd degree polynomial in Stata 14 only takes few seconds and converges quite rapidly after only 7 iterations. Estimation of the complete model is a matter of 1-2 minutes.

coefficients from different models. For our purposes, the most relevant result shown in this section concerns our instruments and their impact on schooling choices irrespective of the selected model. In our preferred GN(3) specification, the *t-statistic* of the unemployment rate at 17 is a reassuring 5.62. Birth in the first half of the year is statistically significant, but only at 10% level; the associated *t-statistic* for this variable is 1.89. In any case, the combinations of our two instruments produce strong first stages. Conventional test for weak identification produce *F-statistic* of 54.28 ruling out concerns of weak instruments for this specification, allowing us to conclude that the combination of our two instruments correlate with schooling decision. Since our model is overidentified, we can check whether our instruments are truly exogenous by performing a Sargan-Hansen test for overidentifying restrictions. The null hypothesis that the overidentifying restriction is valid cannot be rejected since the corresponding *p-value* is 0.56, it thus appears that our combination of instruments is also exogenous[15]. Our two instruments show the expected effects on schooling length by being both positive. The first stage estimate for the unemployment instrument supports the opportunity cost story according to which high unemployment rates encourage individuals to postpone the timing of their entry into the labor market. The birth instrument is also conforming with the theoretical prediction of relatively older students being more likely to access higher levels of education probably due to long lasting effects of their relatively older age at the time of school access.

Since the variable measuring schooling is constant over time, we follow Cameron and Taber (2004) and construct a time-invariance local labor market variable by averaging out for each individual the values of local unemployment over the years in which the individual is included in the wage regressions and include this new measure in the first stage.

The other covariates all show the expected signs. Parental education, ability and family income are positively correlated with educational achievements while number of siblings negatively so, although not significantly. African-American and Hispanic students appear to reach higher levels of education. Although this is somewhat unexpected, the same result was found by Cameron and Taber (2004) and Chen (2008) using the same data. It might be explained by the notion that our control variables capture the initial disadvantage of these groups quite well.

5.2 Wage equation

In Table 4 we report estimates of equation (1). All these estimates are obtained via the first difference model as in equation (14) running four separate regression by schooling level allowing, in this way, for the β's to vary accordingly.

The set of controls includes all the time varying variables: actual experience and its square and the actual unemployment rate for the year and census area where the individual is currently employed when wages are measured. The signs for the estimated covariates are the expected ones. Wages

[15] Specifying the month of birth instrument as a linear trend spanning from 1 to 12 or as four separate birth quartiles does not change our main results. By varying the functional form in our first stage, the probability of acquiring further education is negatively influenced by how late within each year one individual is born, leaving the other covariates unaffected. Furthermore, the combination of the two exclusion restriction would still be strong - F-test 54.26 for the linear case and 27.59 for the quartile specification - and the Sargan-Hansen test for the validity of our instrument cannot reject the null hypothesis of exogenous instruments - p-value .564 in the first case and .930 in the second. We decide to adopt the dichotomous definition since in this specification, the variable is statistically different from zero at lower level of significance. All results available on request.

Table 3: First stage estimates for different values of K. Dependent variable: educational category.

	OP		GN(3)		GN(4)	
Unemployment rate at 17	.113***	(.015)	.096***	(.017)	.106***	(.019)
Born first half	.063	(.042)	.071*	(.038)	.076*	(.041)
Mean local unemp. over working life	-1.169***	(.091)	-1.104***	(.076)	-1.178***	(.083)
Highest grade mother	.026**	(.010)	.023***	(.009)	.026***	(.010)
Highest grade father	.049***	(.008)	.040***	(.007)	.046***	(.008)
Number of siblings	-.014	(.010)	-.010	(.008)	-.012	(.009)
Family income bottom quartile	-.069	(.147)	-.099	(.121)	-.099	(.138)
Family income second quartile	-.018	(.147)	-.043	(.122)	-.039	(.138)
Family income third quartile	.060	(.146)	-.011	(.122)	.026	(.140)
Family income top quartile	.210	(.148)	.124	(.124)	.142	(.143)
AFQT score (adjusted)	.026***	(.001)	.022***	(.002)	.025***	(.002)
Black	.523***	(.061)	.421***	(.061)	.479***	(.072)
Hispanic	.319***	(.073)	.231***	(.062)	.275***	(.073)
Cut-off point (κ_1)	-4.997***	(.619)	-4.997	(fixed)	-4.997	(fixed)
Cut-off point (κ_2)	-3.088***	(.613)	-3.402***	(.124)	-3.244***	(.146)
Cut-off point (κ_3)	-2.136***	(.608)	-2.571***	(.189)	-2.336***	(.219)
Polynomial:						
1			-.163	(.157)	.041	(.237)
2			-.108***	(.019)	-.111**	(.046)
3			.056***	(.012)	.014	(.019)
4					.013**	(.006)
Estimated Moments for ν_i:						
Mean			-.079		.167	
Variance			.784		.895	
Skewness			.194		.300	
Kurtosis			4.753		3.138	
N	3,040		3,040		3,040	

Note: Geographic and cohort controls added. Geographic controls include the urban dummy and regional dummies for residence in the four standard census region at 17. Cohort controls are a set of four indicator variables for age, which ranges between 13 to 18 in 1978. Reference categories: family income missing, whites. */**/*** indicate confidence levels of 10/5/1 percent respectively. Robust standard errors in parentheses.

Table 4: Wage equation. Dependent variable: log hourly wages.

	Less than high school	High school	Some college	4 yrs. college and beyond
Experience	.065***	.069***	.078***	.109***
	(.010)	(.004)	(.006)	(.006)
Experience^2	-.000*	-.000***	-.001***	-.001***
	(.000)	(.000)	(.000)	(.000)
Unemployment rate	-.010	-.014***	.001	-.011**
	(.008)	(.004)	(.005)	(.005)
Constant	1.402***	1.620***	1.581***	1.715***
	(.103)	(.048)	(.068)	(.065)
R^2	.175	.193	.216	.233
N	508	1,321	592	619

*/**/*** indicate confidence levels of 10/5/1 percent respectively.

increase with experience at a decreasing rate and this effect is stronger for college graduates, while high unemployment rates in the area are associated with lower wages.

5.3 Main results

We consider three measures of dispersion: i) the observed wage dispersion given the choice of schooling ($Var[y_{sit}|s_i = s, x_{it}, z_i]$); ii) the potential wage dispersion purged of selection and truncation biases ($\sigma_s^2 + \psi_{st}^2$); and iii) risk in potential wages, after removing truncation and selection biases and incorporating unobserved heterogeneity factors (τ_{st}^2). The results are summarized in Table 5 and Figure 2. We first discuss the sensitivity of our results to the different econometric strategies adopted, we then turn our attention to the consequences of self-selection and truncation and we conclude by examining the causal link between schooling level and risk.

5.3.1 Model comparison

Table 5 reports three different approaches to estimate the various inequality measures distinguished by educational categories. The first column, labeled "parametric" reports the results of a Heckman type selection model in which joint normality for the error terms in the selection and outcome equations is imposed. The second two columns under the "semiparametric" label relax the joint normality assumption, but differ in a crucial aspect: while the estimates shown in the "linear" columns (LSP henceforth) impose a linear relationship between the two error terms, linearity is no longer imposed for identifying the parameters displayed in the non-linear column (NLSP henceforth). The linear case (Q = 1) can be tested against the more general nonlinear specification (in our case Q = 3). Not only are the second and third order polynomial coefficients simultaneously significant at a 1% level (test statistic: $\chi^2(2) = 33.03$) also both coefficients are individually significant at a 1% level (t-ratio's: 6.80 and 5.94). As a result we favor the NLSP model.

The first striking feature of the results reported in Table 5 is the similarity of the parameters shown in the first two columns. Whether normality is imposed or not makes very little difference for the estimation of our parameters if the assumption of linearity is maintained. As noted by Vella (1998) a semiparametric framework relying on a linear relationship in the error terms presents no particular advantage over the Heckman model and that the coefficients estimated in this way generally produce similar results to the parametric counterpart. Our estimates further corroborate this observation.

Equally remarkable is the difference that allowing for a non-linear relationship implies for our results. The parameters in the column labeled as non-linear differ substantially from the other two model specifications. Only for the college graduates category are the results somewhat robust to the different assumptions imposed by the three models.

5.3.2 Importance of self-selection and truncation corrections

Observed wage inequality. Panel I reports estimates for observed wage inequality distinguished by its two components, the transitory and the permanent one. Remember that in our model the transitory component is an exogenous shock that individuals can not act upon and is estimated via a fixed effect model. Therefore, by construction it is not influenced by self-selection nor distributional assumptions and its estimates are necessarily equal between the three estimation frameworks. From the table we can appreciate how the contribution of the transitory component to the observed inequality is very marginal and falling with educational level. Our result indicate that education offers some protection against idiosyncratic shocks. Transitory variability for college graduates is 75% of the transitory variability experienced by high school graduates.

For each of our three specifications observed wage dispersion is mainly determined by its permanent component. This result is at odds with Chen for whom observed inequality was almost evenly split between permanent and transitory components.

Potential wage inequality. If individuals act upon private information about their personal tastes and inclinations to select the preferred risk/pay-off profile, estimates presented in panel I are biased measures of the real level of wage dispersion that each schooling level entails.In panel II part C we show our estimates of the permanent component corrected for selection and truncation biases for our two semiparametric specifications and we compare it directly to our parametric estimates. By confronting the parameters in panel I.A. with those in II.C. we can immediately note the level of bias detected by our estimation procedures. Given our theoretical model, correcting for endogenous education will always result in an underestimation of the potential wage variance by the observed measures of wage variance [16] and this is what we encounter in our results. It has to be noted that even for the categories for which we detect some bias, the underestimation of potential by observed wage dispersion is far from being substantial. Observed wage dispersion accounts for 91% of high school graduates' potential wage dispersion, and for around 98% for college graduates. These results are higher than any other previous parametric and non parametric estimations. For Chen the difference varies between 2% and 30% depending on educational level. As in the case of the observed wage inequality, the impact of the transitory component on potential inequality is practically absent. In our preferred specification, both observed and potential inequality decrease - non-monotonically - by schooling level and high school dropouts exhibit the largest variance.

The permanent component presented in panel C is corrected for self-selection and truncation, but it does not account for unobserved schooling factor ν_i which is included in estimates presented in panel III section E. It is interesting to compare the two estimates of panel C and panel E since from this comparison we see the importance of unobserved heterogeneity for wage dispersion. Controlling for unobserved heterogeneity has little effect for each of our three models. In the case of the NLSP model, it diminishes the estimates of the permanent component for the two middle categories, but has no effect on the estimates for high school dropouts and college graduates, while for the other two models it has an effect on the estimates for the two highest educational categories.

These results are driven by the size of the correlation between wages and schooling factor. For the NLSP model, the parameter $|\rho_s|$ in panel D describes a marginal, but statistically significant correlation between the error terms for the lowest and highest educational categories and a larger correlation for high school graduates and college dropouts. If we compare the correlation coefficients estimated semiparametrically with the parametric and the LSP ones, it is clear that the latter two tend to be larger compared with the NLSP model, with the only exception of high school graduates.

Given our model, estimates of unobserved heterogeneity in panel E have to be directly affected by

[16] See equation (12).

Table 5: Parameters of interest estimation based on full sample

	No high school			High school			Some college			4 yr. college		
	Par.	Semipar.		Par.	Semipar.		Par.	Semipar.		Par.	Semipar.	
I. Observed wage inequality		Linear	Non-lin.		Linear	Non-lin.		Linear	Non-lin.		Linear	Non-lin.
A. Permanent component	.318***	.327***	.731***	.301***	.304***	.514***	.339***	.337***	.537***	.402***	.385***	.461***
$\left(\sigma_s^2\left(1-\rho_s^2(1-Var[g(\nu_i)\|s_i=s]\right)\right)$	(.017)	(.017)	(.196)	(.013)	(.013)	(.107)	(.019)	(.019)	(.099)	(.022)	(.022)	(.060)
B.Transitory component (ψ_{st}^2)	.028***	.028***	.028***	.022***	.022***	.022***	.021***	.021***	.021***	.020***	.020***	.020***
	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)	(.001)
Observed inequality (A+B)	**.346*****	**.355*****	**.759*****	**.323*****	**.326*****	**.536*****	**.360*****	**.358*****	**.558*****	**.422*****	**.405*****	**.481*****
	(.019)	**(.019)**	**(.196)**	**(.013)**	**(.013)**	**(.107)**	**(.020)**	**(.020)**	**(.099)**	**(.023)**	**(.023)**	**(.060)**
II. Potential wage inequality												
C. Permanent component (σ_s^2)	.320***	.333***	.731***	.301***	.304***	.565***	.366***	.359***	.543***	.451***	.402***	.461***
- Adjusted for selection and truncation	(0.19)	(0.19)	(.197)	(.013)	(.013)	(.107)	(.023)	(.023)	(.099)	(.028)	(.028)	(.060)
Potential wage inequality (C+B)	**.348*****	**.361*****	**.759*****	**.323*****	**.326*****	**.587*****	**.387*****	**.380*****	**.564*****	**.471*****	**.422*****	**.481*****
	(.020)	**(.020)**	**(.196)**	**(.014)**	**(.014)**	**(.107)**	**(.021)**	**(.021)**	**(.099)**	**(.024)**	**(.024)**	**(.060)**
III. Wage uncertainty												
D. Correlation coefficient (ρ_s)	-.090	.170	.010***	-.013	.042	.355***	-.280**	.259**	.140***	-.430***	.293***	.034***
	(.107)	(.107)	(.001)	(.076)	(.076)	(.033)	(.058)	(.058)	(.012)	(.086)	(.086)	(.002)
E. Permanent component ($\sigma_s^2(1-\rho_s^2)$)	.317***	.324***	.731***	.301***	.303***	.494***	.337***	.335***	.533***	.367***	.367***	.461***
- Accounted for Unobs. Schooling Factor	(.019)	(.019)	(.196)	(.013)	(.013)	(.107)	(.018)	(.018)	(.099)	(.021)	(.021)	(.060)
Degree of wage uncertainty (τ_s^2)	**.345*****	**.350*****	**.759*****	**.323*****	**.325*****	**.516*****	**.359*****	**.349*****	**.553*****	**.387*****	**.377*****	**.480*****
	(.020)	**(.020)**	**(.196)**	**(.011)**	**(.011)**	**(.107)**	**(.020)**	**(.020)**	**(.099)**	**(.022)**	**(.022)**	**(.060)**
Unobserved heterogeneity ($\gamma_s^2\sigma_\nu^2$)	**.003**	**.010**	**.000**	**.000**	**.000**	**.071*****	**.029*****	**.024**	**.010*****	**.083*****	**.035**	**.000**
	(.012)	**(.012)**	**(.000)**	**(.005)**	**(.005)**	**(.000)**	**(.008)**	**(.008)**	**(.000)**	**(.020)**	**(.020)**	**(.000)**

Note: Estimates of transitory volatility are obtained by regressing ψ_{st}^2 on age dummies and categorical educational variables. */**/*** indicate confidence levels of 10/5/1 percent respectively. Bootstrapped standard errors based on 500 replications in parentheses. For the two semiparametric specifications ρ_s is expressed in absolute values.

the estimated correlation coefficients. Unobserved heterogeneity in the NLSP model is absent for high school dropouts and college graduates, but present, albeit small, for high school graduates and college dropouts. Conversely, in the parametric and LSP case, this parameter explains a bigger share of wage dispersion for college drop outs and college graduates.

Predictability of wage variance. The decomposition of the two determinants of dispersion - uncertainty and heterogeneity - shows that the contribution of private information to the identification of a causal relation between risk and wages varies between the four categories. If we concentrate on our favored NLSP procedure we see that at one extreme, risk explains 100% of wage variability for the highest and lowest educational categories, at the other, private information accounts for around 12% of potential wage variability for high school graduates, while for college dropouts private information accounts for only 2%, which is exactly the same estimate obtained by Chen.

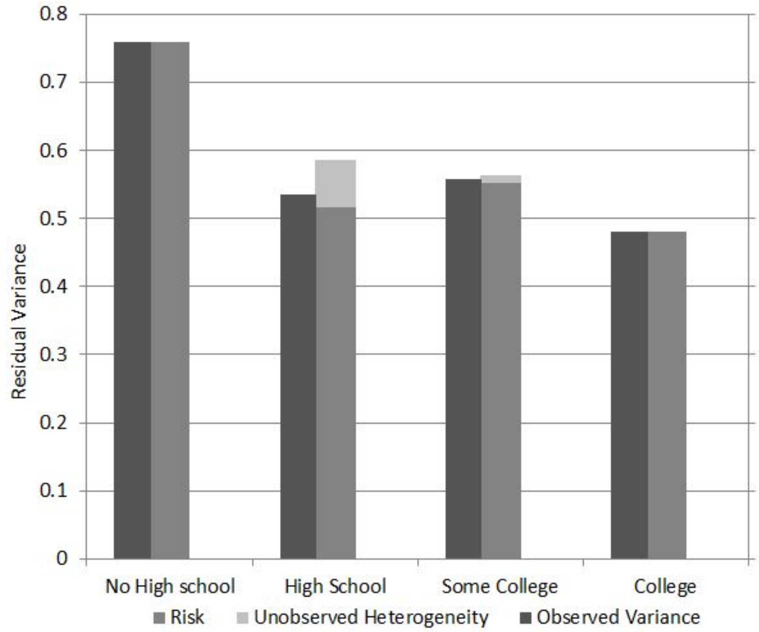
It is interesting to contrast our favored semiparametric with our parametric and linear semiparametric estimates for the contribution of unobserved heterogeneity for the identification of the potential wage dispersion. The estimates diverge considerably. In general the contribution of unobserved heterogeneity, even if small also for the parametric and LSP case, is clearly larger than the one detected for the NLSP case. Additionally, the ordering of the four categories with respect to the unpredictability that each involve diverges completely. In the first two specifications high school graduates are those exposed to higher unpredictability while college graduates sit at the other extreme.

The estimated level of unobserved heterogeneity detected in our NLSP estimation falls short of that found in Chen and *a fortiori* also than those found by CHN, but it conforms with the results encountered in Carneiro et al. (2003) who also found substantial uncertainty when forecasting future payoffs at the time when schooling decisions are made. Only for one out of four educational categories self-selection considerably biases estimates of the causal impact of schooling on risk and with the exception of high school dropouts, individuals seem unable to beat the econometrician in predicting wage variability.

Possible mechanisms able to explain the negligible advantage that students display over the econometrician when predicting future wages is a misuse of their *private information*. In fact, in light of a growing literature in both economics (Jerrim, 2011) and psychology (Kruger and Dunning, 1999) on students' overconfidence[17] we cannot exclude *a priori* the possibility that possessing superior information does not immediately translate into risk minimizing choices. Kruger and Dunnig find that less able individuals tend to overestimate their abilities and are generally less able to produce accurate prediction of their relative positioning in the wage distribution. Jerrim (2011), for a sample of UK college students, finds evidence of a substantial overestimation of their starting salaries. These two distinct mechanism operating at the two extremes of the educational distribution could, at least partially, explain the absence of any detectable self-selection that we encounter for those two categories. On the other hand, we could think that for the type of professions that a high school degree gives access to, such as clerical occupations or routine manual ones, the variance of wages is less pronounced, making an accurate estimate of one's future economic outcomes more viable.

[17]We are grateful to one anonymous referee for pointing us towards this stream of literature and suggesting this possible mechanism.

Figure 2: Wage inequality measures by schooling levels



5.3.3 The causal impact of education on risk

Figure 2 summarizes our main findings from our preferred specification: the NLSP model. To understand this figure, concentrate on high school graduates. The observed variance equals 0.536 for this group (A+B in Table 5), unobserved heterogeneity is estimated at 0.071 and potential wage equality at 0.587 (C+B in Table 5) for this group. The difference is risk: 0.587-0.071 =0.516. Figure 2 shows that observed wage inequality noticeably understates potential wage inequality only for the high school graduates category. The highest and lowest educational category do not seem to be affected by selection issues. High school drop outs are exposed to a considerable amount of unforeseeable risk and in general education decreases wage variance and risk. Interestingly, even though potential wage inequality is higher for high school graduates than for college dropouts, risk is lower for the former compared to the latter and the higher potential wage variance is explained by increasing unobserved heterogeneity. Our results point towards the existence of a sheepskin effect in the causal relation between education and risk since high school and college graduates are those experiencing the lowest level of risk. As in Chen and CHN, we do not find any evidence for the existence of a risk return tradeoff for college education. This contrasts with previous literature on this topic which neglected to account for selection and truncation biases (Christiansen et al., 2007; Diaz-Serrano et al., 2008). On the contrary, our results show that investing in college education is not risky *per se* as it actually decreases the level of uncertainty that individuals are exposed to. What causes an increase in risk is dropping out of college. College entry causes a 7% increase in risk, while obtaining a college degree decreases risk by 15%.

6 Conclusion

In this paper we apply a two step semiparametric estimation technique imposing minimal assumptions on the structure of the error terms to distinguish various components of wage variance. We extend the original parametric technique from the dichotomous to the polychotomous case and to selection models providing consistent estimates of within education potential wage variation, accounting for selection and truncation biases and the degree of private information owned by the individuals that might be used to select their favorite level of education. We are thus able to evaluate the magnitude of wage risk that every education level entails.

Our analysis offers several insights. First, it clearly suggests that procedures that equate variability with uncertainty might overstate risk and, hence, understate the pricing of it. In fact, at least for high school graduates, observed wage dispersion does underestimates potential wage dispersion. This gives an indication that individuals try to act upon their private information when selecting their educational level, at least in the most predictable cases. Second, we find that investing in education has a reducing effect on wage variability, but at a decreasing rate with the largest marginal effect encountered when obtaining an high school diploma; additionally, education offers some protection against unpredictable shocks as our estimated transitory components suggest. Third, in most cases, individuals cannot predict a larger share of their lifetime income variability compared to the econometrician. Our assessment of the relevance of unobserved heterogeneity places the degree of private information possessed and acted upon by individuals below those previously encountered by Cunha et al. (2005) and Cunha and Heckman (2007) for the semiparametric case and in Chen (2008) for the parametric one. Fourth, by directly comparing parametric and semiparametric estimates on the same sample we have shown how relaxing the distributional assumptions does matter greatly. This result indicate that different distributional assumptions might play a role in the contrasting results encountered so far in the literature. Lastly, we have shown that the type of nonparametric estimation method used and the assumption imposed when correcting for self-selection and separately identifying unobserved heterogeneity from risk really matters. Semiparametric methods relying on a linear relationship between the disturbances in the selection and outcome equation offer little advantage over the parametric case.

If our results are correct, we have to conclude that individuals cannot do much better than econometricians in predicting future wages. Nonetheless, even though private information is probably present, what our results show is that uncertainty is empirically dominating. The vast majority of wage variability is unpredictable to the individuals and econometricians alike and it does not decrease with schooling. Uncertainty might as well be more complex than mere wage variability across educational categories. An obvious source of uncertainty is risk of unemployment. If more education reduces the likelihood of unemployment spells, risk can be reduced by the prospective of a continuous work career. A complete study of educational risk has to account for both sources of uncertainty: wage variations and risk of unemployment. We leave this task to future research.

References

Abadie, A.: 2002, 'Bootstrap Test for Distributional Treatment Effects in Instrumental Variable Models'. *Journal of the American Statistical Association* **97**(457), 284–292.

Abadie, A., J. Angrist, and G. Imbens: 2002, 'Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings'. *Econometrica* **70**(1), 91–117.

Acemoglu, D.: 1999, 'Changes in Unemployment and Wage Inequality:An Alternative Theory and Some Evidence'. *American Economic Review* **89**(5), 1259–1278.

Acemoglu, D.: 2002, 'Technical Change, Inequality, and the Labor Market'. *Journal of Economic Literature* **40**(1), 7–72.

Ahn, H. and J. Powell: 1993, 'Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism'. *Journal of Econometrics* **58**(1/2), 3–29.

Angrist, J., V. Chernozhukov, and I. Fernández-Val: 2006, 'Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure'. *Econometrica* **74**(2), 539–563.

Angrist, J. and A. Krueger: 1991, 'Does Compulsory School Attendance Affect Schooling and Earnings?'. *Quarterly Journal of Economics* **106**(4), 979–1014.

Autor, D. H., L. F. Katz, and M. S. Kearney: 2006, 'The Polarization of the U.S. Labor Market'. *American Economic Review* **96**(2), 189–194.

Banks, J. and P. A. Diamond: 2010, 'The Base for Direct Taxation'. In: J. Mirrlees, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles, and J. Poterba (eds.): *Dimensions of Tax Design: The Mirrlees Review*. Oxford University Press.

Betts, J. R. and L. L. McFarland: 1995, 'Safe Port in a Storm: The Impact of Labor Market Conditions on Community College Enrollments'. *The Journal of Human Resources* **30**(4), 741–765.

Blau, F. D. and L. Kahn: 2013, 'The Feasibility and Importance of Adding Measures of Actual Experience to Cross-Sectional Data Collection'. *Journal of Labor Economics* **31**(2), S17–S58.

Cameron, S. V. and J. J. Heckman: 1998, 'Life Cycle Schooling and Dynamic Selection Bias: Models and Evidence for Five Cohorts of American Males'. *Journal of Political Economy* **106**(2), pp. 262–333.

Cameron, S. V. and J. J. Heckman: 2001, 'The Dynamics of Educational Attainment for Black, Hispanic, and White Males'. *Journal of Political Economy* **109**(3), 455–499.

Cameron, S. V. and C. Taber: 2004, 'Estimation of Educational Borrowing Constraints Using Returns to Schooling'. *Journal of Political Economy* **112**(1), 132–182.

Card, D.: 1993, 'Using Geographic Variation in College Proximity to Estimate the Return to Schooling'. NBER Working Papers 4483, National Bureau of Economic Research, Inc.

Card, D.: 2001, 'Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems'. *Econometrica* **69**(5), 1127–1160.

Carneiro, P., K. Hansen, and J. J. Heckman: 2003, 'Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choices'. *International Economic Review* **44**(2), 361–422.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil: 2011, 'Estimating Marginal Returns to Education'. *American Economic Review* **101**(6), 2754–81.

Carneiro, P. and S. Lee: 2009, 'Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality'. *Journal of Econometrics* **149**(2), 191 – 208.

Chen, S.: 2008, 'Estimating the Variance of Wages in the Presence of Selection and Unobserved Heterogeneity'. *The Review of Economics and Statistics* **90**(2), 275–289.

Chen, S. and S. Khan: 2007, 'Estimating the Casual Effect of Education on Wage Inequality Using IV Methods and Sample Selection Models'. Working paper, University at Albany-SUNY.

Christiansen, C., J. S. Joensen, and H. S. Nielsen: 2007, 'The risk-return trade-off in human capital investment'. *Labour Economics* **14**(6), 971 – 986.

Clark, D.: 2011, 'Do Recessions Keep Students in School? The Impact of Youth Unemployment on Enrolment in Post-compulsory Education in England'. *Economica* **78**(311), 523–545.

Cosslett, S.: 1983, 'Distribution Free Maximum Likelihood Estimator of the Binary Choice Model'. *Econometrica* **51**(3), 765–782.

Cosslett, S.: 1991, 'Nonparametric and Semiparametric Estimation Methods in Econometrics and Statistics'. In: W. Barnett, J. Powell, and G. Tauchen (eds.): *Semiparametric Estimation of a Regression Model with Sample Selectivity.* Cambridge, UK: Cambridge University Press, pp. 175–197.

Crawford, C., L. Dearden, and E. Greaves: 2011, 'When You Are Born Matters: Evidence for England'. IFS Report R80, Institute for Fiscal Studies.

Crawford, C., L. Dearden, and E. Greaves: 2014, 'The drivers of month-of-birth differences in children's cognitive and non-cognitive skills'. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **177**(4), 829–860.

Cunha, F. and J. Heckman: 2007, 'Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling'. *Labour Economics* **14**(6), 870–893.

Cunha, F., J. Heckman, and S. Navarro: 2005, 'Separating Uncertainty from Heterogeneity in Life Cycle Earnings'. *Oxford Economic Papers* **57**(2), 191–261.

Dahl, G. B.: 2002, 'Mobility and the Return to Education: Testing a Roy Model with Multiple Markets'. *Econometrica* **70**(6), 2367–2420.

Diaz-Serrano, L., J. Hartog, and H. S. Nielsen: 2008, 'Compensating Wage Differentials for Schooling Risk in Denmark'. *Scandinavian Journal of Economics* **110**(4), 711–731.

Gallant, R. A. and D. W. Nychka: 1987, 'Semi-Nonparametric Maximum Likelihood Estimation'. *Econometrica* **55**(2), 363–390.

Goldberger, A.: 1983, 'Abnormal Selection Bias'. In: S. Karlin, T. Amemiya, and L. Goodman (eds.): *Studies in Econometrics, Time Series, and Multivariate Statistics.* New York: Academic Press.

Hansen, K. T., J. J. Heckman, and K. J. Mullen: 2004, 'The effect of schooling and ability on achievement test scores'. *Journal of Econometrics* **121**(12), 39–98.

Heckman, J.: 1979, 'Sample Selection Bias as a Specification Error'. *Econometrica* **1**(74), 153–162.

Jerrim, J.: 2011, 'Do UK Higher Education Students Overestimate Their Starting Salary?*'. *Fiscal Studies* **32**(4), 483–509.

Kane, T. J. and C. E. Rouse: 1995, 'Labor-Market Returns to Two and Four Years College'. *American Economic Review* **85**(3), 600–614.

Katz, L. F. and D. H. Autor: 1999, 'Changes in the Wage Structure and Earnings Inequality'. In: O. Ashenfelter and D. Card (eds.): *Handbook of Labor Economics*. Amsterdam, North-Holland: Elsevier Science, pp. 1463–1555.

Keane, M. P. and K. I. Wolpin: 1997, 'The Career Decisions of Young Men'. *Journal of Political Economy* **105**(3), pp. 473–522.

Kruger, J. and D. Dunning: 1999, 'Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments.'. *Journal of Personality and Social Psychology* **77**(6), 1121–1134.

Mazza, J., H. van Ophem, and J. Hartog: 2013, 'Unobserved heterogeneity and risk in wage variance: Does more schooling reduce earnings risk?'. *Labour Economics* **24**(C), 323–338.

Moretti, E.: 2000, 'Do Wages Compensate for Risk of Unemployment? Parametric and Semiparametric Evidence from Seasonal Jobs'. *Journal of Risk and Uncertainty* **20**(1), 45–66.

Neal, D. A. and W. R. Johnson: 1996, 'The Role of Premarket Factors in Black-White Wage Differences'. *Journal of Political Economy* **104**(5), 869–895.

Newey, W. K.: 2009, 'Two-Step Series Estimation of Sample Selection Models'. *Econometrics Journal* **12**(1).

Oaxaca, R. L. and T. Regan: 2009, 'Work Experience as a Source of Specification Error in Earnings Models: Implications for Gender Wage Decompositions'. *Journal of Population Economics* **22**(2), 463–499.

Powell, J.: 1989, 'Semiparametric estimation of Censored selection models'. Working paper, University of Wisconsin.

Robinson, P. M.: 1988, 'Root-N-Consistent Semiparametric Regression'. *Econometrica* **56**(4), 931–954.

Roy, A. D.: 1951, 'Some Thoughts on the Distribution of Earnings'. *Oxford Economic Papers* **3**(2), 135–146.

Stewart, M. B.: 2004, 'Semi-nonparametric estimation of extended ordered probit models'. *The Stata Journal* **4**(1), 27–39.

van Soest, A., M. Das, and X. Gong: 2002, 'A Structural Labour Supply Model with Flexible Preferences'. *Journal of Econometrics* **107**, 345–374.

Vella, F.: 1998, 'Estimating Models with Sample Selection Bias: A Survey'. *The Journal of Human Resources* **33**(1), 127–169.

Willis, R. J. and S. Rosen: 1979, 'Education and Self-Selection'. *Journal of Political Economy* **87**(5), S7–36.

A Identification of the parameters.

The starting point is the model specified in eqs. (1)-(4). For future reference it is relevant to note that we can calculate the average of the conditional errors across T or specify it for a specific $t = t_0$ in line with eq (10):

$$var\left[\frac{1}{T}\sum_{t=1}^{T}(\sigma_s e_{si} + \psi_{st}\epsilon_{it})|s_i = s\right] = \rho_s^2\sigma_s^2 var[g(\nu_i)|s_i = s] + \sigma_{\xi s}^2 + \frac{1}{T}\overline{\psi_s^2} \tag{26}$$

$$var[\sigma_s e_{si} + \psi_{st_0}\epsilon_{it_0}|s_i = s] = \rho_s^2\sigma_s^2 var[g(\nu_i)|s_i = s] + \sigma_{\xi s}^2 + \psi_{st_0}^2. \tag{27}$$

From our estimation procedure explained in section 3, we get consistent estimates of α_s β_s and ψ_{st} for $s = 0, 1, 2, 3$ and $t = 1, ..., T$ and from this we estimate $(\sigma_s e_{si} + \psi_{st}\epsilon_{it}|s_i = s)$ for every $i = 1, ..N;\ t = 1, .., T$ and for given s.

We assume that $\sigma_g g(\nu_i)$ can be approximated by a polynomial of order Q as in (17). If we calculate the condition expectation of this we get (18). The conditional expectations $E\left[(\sigma_\nu\nu_i)^j\,|s_i = s\right]$ can be consistently estimated by (25) for every individual and for every level of schooling.

In the usual econometric fashion we can write:

$$\begin{aligned}(\sigma_s e_{si} + \psi_{st}\epsilon_{it}|s_i \quad &= \quad s) = E[\sigma_s e_{si} + \psi_{st}\epsilon_{it}|s_i = s] + \varsigma_{sit} \\ &= \frac{\rho_s\sigma_s}{\sigma_g}\sum_{j=0}^{Q}\tilde{\alpha_j}E\left[(\sigma_\nu\nu_i)^j\,|s_i = s\right] + \varsigma_{sit}.\end{aligned} \tag{28}$$

So we can fit the $Q + 1$ conditional moments (including the 0-moment which equals $Pr[s_i = s]$, $s = 0, 1, 2, 3$) to the residuals $(\sigma_s e_{si} + \widehat{\psi_{st}\epsilon_{it}}|s_i = s)$. Since there are restrictions on the parameters (the scale factors $\frac{\rho_s\sigma_s}{\sigma_g}$ are different but the $\tilde{\alpha_j}'s$ $(j = 0, 1, .., q)$ are the same) we need to apply NLS to:

$$\sigma_s e_{si} + \psi_{st}\epsilon_{it} = \left(\sum_{s=0}^{3}\theta_s \cdot I(s_i = s)\right) \cdot \left(\sum_{j=0}^{Q}\frac{\rho_1\sigma_1}{\sigma_g}\tilde{\alpha_j}E\left[(\sigma_\nu\nu_i)^j\,|s_i = s\right]\right) + \varsigma_{sit}, \tag{29}$$

where $I(s = s) = 1$ if $s_i = s$ and 0 otherwise. Furthermore: $\theta_s = \frac{\rho_s\sigma_s}{\rho_1\sigma_1}$ Applying NLS, we obtain the consistent estimates: $\widehat{\frac{\rho_1\sigma_1}{\sigma_g}}\tilde{\alpha}_j, \widehat{\theta}_0, \widehat{\theta}_2$ and $\widehat{\theta}_3$, $j = 0, 1, .., q$. [18] We opt for s = 1 as the reference since the

[18] In the reported estimation we chose $Q = 3$. For higher order approximations the calculation of the conditional

estimated effect is the largest one (i.e. $\widehat{\theta}_0 < 1; \widehat{\theta}_2 < 1$ and $\widehat{\theta}_3 < 1$). In our experience this choice speeds up the NLS-estimation.

Note that

$$E\left[\frac{\rho_s\sigma_s}{\sigma_g}\sigma_g g(\nu_i)\right] = \sum_{s=0}^{3} E\left[\frac{\rho_s\sigma_s}{\sigma_g}\sigma_g g(\nu_i)|s_i = s\right] Pr[s_i = s] = M_{1s} \quad (30)$$

and (31)

$$E\left[\left(\frac{\rho_s\sigma_s}{\sigma_g}\sigma_g g(\nu_i)\right)^2\right] = \sum_{s=0}^{3} E\left[\left(\frac{\rho_s\sigma_s}{\sigma_g}\sigma_g g(\nu_i)\right)^2 |s_i = s\right] Pr[s_i = s] = M_{2s}, \quad (32)$$

where the conditional moments can be estimated using (25). From this we can estimate $var(\frac{\rho_s\sigma_s}{\sigma_g}\sigma_g g(\nu_i)) = \rho_s^2\sigma_s^2$ consistently by:

$$\frac{N_s}{N_s - 1}\frac{1}{N_s}\left(\sum_{i=1}^{N_S}\widehat{M_{2s}}\right) - \frac{N_s}{N_s - 1}\left(\frac{1}{N_s}\sum_{i=1}^{N_S}\widehat{M_{1s}}\right)^2. \quad (33)$$

We can estimate $var\left[\frac{\rho_s\sigma_s}{\sigma_g}\sigma_g g(\nu_i)|s_i = s\right]$ using both the first and second conditional moments that were already part of (30) and (32). Furthermore we can consistently estimate the variance of $(\sigma_s e_{si} + \psi_{st}\epsilon_{it}|s_i = s)$ by the sample variance of $\widehat{\sigma_s e_{si} + \psi_{st}\epsilon_{it}}$ for the sub sample $s_i = s$ and averaged across T or for a specific $t = t_0$. As a result by using (26) or (27) we can retrieve consistent estimates of $\sigma_{\xi s}^2$ for $s = 0, 1, 2, 3$, and therefore, by using

$$\rho_s = Cov[e_{si}, g(\nu_i)] = \frac{\rho_s\sigma_s}{\sqrt{\rho_s^2\sigma_s^2 + \sigma_{\xi s}^2}},$$

we can estimate $|\rho_s|$ and consequently σ_s^2 since we already have consistent estimates of $\rho_s^2\sigma_s^2$ for $s = 0, 1, 2, 3$. The sign of ρ_s is not identified. If we would know the sign of one of the ρ_s for $s = 0, 1, 2$ or 3, the signs of the other correlations would follow from the estimates of θ_0, θ_2 and θ_3. The reason that the sign of ρ_s is not identified is that ρ_s and $g(\nu_i)$ are always paired in our model and that there are no restrictions on the signs of the polynomial approximation of $g(\nu_i)$. Note that this is not important for the main parameters we want to identify, cf. Table 5.

jth-order moments became imprecise if $j > 6$.

B Parametric wage equation.

Table 6: Wage equation

	Less than high school	High school	Some college	4 yrs. college and beyond
Current local unemployment	-.045	.140**	.195**	.191**
	(.069)	(.044)	(.065)	(.074)
Experience	.125***	.080**	.042	.051
	(.032)	(.025)	(.031)	(045)
Experience^2	-.002*	-.001	-.000	-.001
	(.001)	(.001)	(.001)	(.001)
Black	-.056	-.034	-.178**	-.027
	(.095)	(.059)	(.073)	(.078)
Hispanic	.125	-.024	-.038	-.049
	(.089)	(.059)	(.069)	(.085)
AFQT score (adjusted)	.006*	.004**	-.003	.002
	(.003)	(.001)	(.002)	(.002)
Highest grade mother	.001	-.007	.007	.005
	(.012)	(.007)	(.009)	(.010)
Highest grade father	.002	-.004	-.001	-.002
	(.010)	(.007)	(.008)	(.009)
Number of siblings	-.002	-.009	.013	-.003
	(.010)	(.006)	(.008)	(.011)
Family income bottom quartile	-.362**	.036	-.137	.145
	(.165)	(.126)	(.189)	(.201)
Family income second quartile	-.342**	-.000	.014	.177
	(.158)	(.124)	(.188)	(.198)
Family income third quartile	-.176	.019	-.039	.179
	(.158)	(.122)	(.191)	(.194)
Family income top quartile	-.137	.057	.005	.296
	(.179)	(.122)	(.183)	(.192)
Correction terms :	.058	-.023	-.168**	-.220**
	(.116)	(.066)	(.071)	(.107)
Constant	1.506***	.616**	.952**	1.252**
	(.423)	(.258)	(.377)	(.450)
R^2	.205	.161	.118	.196
N	508	1,321	592	619

Note: Socio-economic controls including dummies for ethnicity, parents education and family income at 14 added. Geographic controls including the urban dummy and three regional dummies for residence at 14 added. */**/*** indicate confidence levels of 10/5/1 percent respectively. Bootstrapped standard errors based on 500 replications in parentheses.