Separating risk in education from heterogeneity: a semiparametric approach

Mazza, J.; van Ophem, H.

[Link to publication](#)

UvA ECONOMETRICS

Separating risk in education from heterogeneity: a semiparametric approach

Jacopo Mazza and Hans van Ophem

UvA UNIVERSITEIT VAN AMSTERDAM

Separating risk in education from heterogeneity: a semiparametric approach.

Jacopo Mazza *

and

Hans van Ophem

Amsterdam School of Economics

Universiteit van Amsterdam

Roetersstraat 11 1018WB Amsterdam, The Netherlands

December 13, 2010

Abstract

Returns to education are variable within the same educational group. If uncertain payoffs are a concern to individuals when selecting education, wage variance is relevant. It is the resultant of unobserved heterogeneity and pure uncertainty. The first element is known to the individual, but unknown to the researcher, the second is unknown to both. If individuals exploit private information to select their level of education the variance 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 will apply a semi parametric estimator to tackle the selectivity issues. This method does not rely on the joint normality of errors in the selection and wage equations and is robust to misspecification of the errors distribution. Results suggest that pure risk tends to increase with education. Private information accounts for a bigger share of total wage variance observed in the data than previously thought.

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

JEL classification: C14; C24; D81; J31

---

*Corresponding author: j.mazza@uva.nl. All data and computer programs are available on request.

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 wage variance increases with schooling level and individuals are risk-averse, the increasing 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 is calculated from truncated wages distributions. The truncation is an effect of private information: individuals possess information about their tastes and inclinations and might use this information to select their level of education assuring them the best risk/pay-off profile. Since this 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 wage uncertainty or *risk* is the part of wage variability which is not foreseeable by the individual even with the superior knowledge that he possesses about himself. *Unobserved heterogeneity,* instead, is that part of wage variability that depends on factors known to the individual, but not observable by the researcher. Forgetting to disentangle risk and heterogeneity will cause an overestimation of real risk and, in turn, an underestimation of risk premium offered in the labor market in the form of higher wages.

Our two main goals are to establish a causal relation between education and wage inequality and to estimate a proper measure of *risk* that various educational categories entail disentangling it from *unobserved heterogeneity.* The literature on this issue is scarce. Chen (2008) tackles the issues of selectivity and unobserved heterogeneity taking dispersion of wages as the outcome of interest implementing a standard parametric selection model based on normal distributions with instrumental variables as originally proposed by Heckman (1979) and extending it to the polychotomous case. We apply the same formalization, but we depart from it in an essential aspect: we do not impose restriction of normality on the distribution of disturbances in the primary and secondary equation.

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 joint distribution of errors of wage and selection equation may lead to inconsistent estimates. This criticism spurred a growing literature (Robinson, 1988; Cosslett, 1991; Ahn and Powell, 1993; Newey, 2009) proposing alternative and usually semi-parametric estimation methods.

In this paper we apply a two stage semi parametric estimation method. Similarly to the two step Heckman method we first estimate a selection equation from which we build four

selection correction terms, one for each category of schools, and include them as additional regressors in the outcome equation to re-establish zero conditional mean of the error term. We do so by exploiting the semi parametric[1] estimator developed by Gallant and Nychka (1987) in the first stage and a procedure proposed by Cosslett (1983; 1991) in the second. Additionally, we propose an alternative method to correct for self-selection via the direct inclusion of estimates of the expected value of unobserved heterogeneity obtained by Gallant and Nychka's procedure in the wage equation. In the continuation of this paper we refer to this method as Olsen method since it relies on assumptions of linearity of the error term first discussed in Olsen (1980). Both estimates are robust to misspecification of error distribution functions and allow us to minimize the distributional assumptions.

To our knowledge, this is the first paper dealing with both issues of self-selection and unobserved heterogeneity semiparametrically. Chen (2008) deals with both issues, but strictly parametrically, while the semiparametric methods 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 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.

Our empirical analysis is based on the National Longitudinal Survey of Youth (NLSY). We first obtain estimates of potential wages inequality robust to selection and truncation biases and then distinguish between the two components of inequality, heterogeneity and risk, semiparametrically. Results suggest that observed inequality and potential inequality are maximum for high school and college drop outs. Pure risk increases with college entry while unobserved heterogeneity accounts for a larger share of potential wage variability than previously thought.

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 inequality is not the correct quantification of real wage inequality due to self-selection and b) even if we were able to

---

[1]Gallant and Nychka (1987) refer to their method as "semi-non parametric". Since we use both semiparametric and semi-non parametric methods to estimate our model, we use only the term semiparametric to refer to it.

correct for self-selection the corrected wage inequality would still pool real risk and unobserved heterogeneity together. The Chen (2008) model, that we use as a starting point, offers a solution to both of these identification puzzles. Chen (2008) procedure is divided in two separate parts. First wage inequality corrected for self-selection is identified and then 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 would bias estimates if not accounted for. By differencing fixed effects out consistent estimates are obtained.

2.1 The model

The model presented in Chen (2008) 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) 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 for 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 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:

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

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 level. In other words, the potential wage is the hypothetical wage that the subject would earn if, instead of the educational choice he actually made, he had chosen any of the other three counterfactuals. It is determined by the regression model:

$$y_{sit} = \alpha_s + x_{it}\beta_s + \sigma_s e_{si} + \psi_{st}\epsilon_{it} \text{ if } s_i = s, \tag{1}$$

α_s is a schooling specific constant; β_s is a schooling specific vector of coefficients for the matrix of observable covariates x_{it}; the individual fixed effect is represented by the time invariant term ($\sigma_s e_{si}$) and it is allowed to correlate with x_{it}; the error term $\psi_{st}\epsilon_{it}$ denotes transitory shocks uncorrelated with personal characteristics and across time; e_{si} and ϵ_{it} are random variables uncorrelated with each other. 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 incorporating institutional or macroeconomic shocks uncorrelated with the individual effects.

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:

$$s_i = s \text{ if } a_{si} \leq \nu_i < a_{s+1,i} \text{ for } s = 0, 1, 2, 3 \tag{2}$$

In this expression ν_i is the unobserved schooling factor known to the individual that includes taste for educations, 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 all observable characteristics x_{it} plus an instrument for education[2]; θ is the vector of coefficients for z_i and κ_s is a constant with $\kappa_0 = -\infty$ and $\kappa_4 = \infty$, respectively. We assume ν_i to be uncorrelated with the transitory shocks ϵ_{it}, but to be correlated with the fixed effect e_{si}. The correlation coefficient is indicated by ρ_s.

In order to be able to disentangle the share of wage variance due to real uncertainty from that caused by unobserved heterogeneity, we will have to rely on some additional assumption regarding the disturbances in the primary and selection equation. Following Olsen (1980) and Maddala (1983) we assume linearity of the conditional expectations of e_{si} given ν_i so that:

$$\sigma_s e_{si} = \sigma_{e\nu s}\nu_i + \xi_s \tag{4}$$

with $E[\xi_s|\nu_i] = 0$, $Var[e_{si}|x_{it}, z_{it}] = \sigma_{es}^2$, $Var[\xi_s] = \sigma_\xi^2$ and $Cov[e_{si}, \nu_i] = \sigma_{e\nu s} = \rho_s\sigma_s\sigma_\nu$.

From the individual viewpoint the expected value of future wages is given by[3]:

$$E[y_{sit}|z_i, x_{it}, \nu_i] = \alpha_s + x_{it}\beta_s + \gamma_s\nu_i \tag{5}$$

where $\gamma_s \equiv \sigma_s\sigma_\nu\rho_s$. 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.: $\rho_s < 0$).

---

[2]Identification of true parameters β might be achieved even with $x_i = z_i$ through nonlinearity in the inverse Mills ratios. In practice the inverse Mills ratio is close to linear and the degree of identification weak, resulting in inflated second step standard errors and unreliable estimates of β (Vella, 1998; Cameron and Trivedi, 2005).

[3]See the appendix for derivation.

Since individuals posses a more accurate assessment of their own abilities then researchers, private information (ν_i) has to be accounted in order to build a true measure of risk. Wage uncertainty is the variance of permanent and transitory component from the individual viewpoint that has to be separated from unobserved heterogeneity. Following Chen (2008), we indicate wage uncertainty or risk with τ. Using equation (4) and our distributional assumptions on disturbances, given above, we obtain[4] an expression for risk as the variance of wage uncertainty given observed and unobserved heterogeneity:

$$\tau_s^2 = Var[\sigma_s e_{si} + \psi_{st}\epsilon_{it}|z_i, x_{it}, \nu_i] = \sigma_s^2(1 - \rho_s^2) + \psi_{st}^2 \tag{6}$$

Remembering that the extent of predictability of wage uncertainty from the personal standpoint is expressed by the correlation coefficient ρ_s equation (6) makes the formal link between uncertainty and private information explicit. In fact, if the correlation between unobserved schooling factor (ν_i) and permanent component of wage inequality (e_{si}) is perfect (i.e.: $\rho_s = 1$) the subject can predict exactly the permanent part of his wage variability and uncertainty is only caused by transitory shocks (ψ_{st}^2). On the other hand, if correlation is absent (i.e.: $\rho_s = 0$) the subject does not posses any additional information compared to the researcher and wage uncertainty is observed in the data.

Rearranging equation (6) as $\sigma_s^2 + \psi_{st}^2 = \gamma_s^2 + \tau_s^2$ helps us visualizing how potential wage inequality ($\sigma_s^2 + \psi_{st}^2$) is the sum of two elements: variance of unobserved heterogeneity (γ_s^2) and wage uncertainty (τ_s^2). Note also that if correlation between schooling and unobserved tastes for education exists (i.e.: $\rho_s \neq 0$) potential wage inequality overstates the real degree of risk ($\tau_s^2 < \sigma_s^2 + \psi_{st}^2$).

3 Semiparametric estimation

In the previous section we have discussed the potential source of self-selection. In presence of self-selection the zero conditional mean of the error term in the outcome equation is violated leading to inconsistent parameters estimates if an OLS regression is used. The first solution to correct for sample selection bias has been introduced by Heckman(1974; 1976; 1979). Heckman's approach restores the zero conditional mean of errors in the outcome equation via the inclusion of a selection correction term λ_i. Under normality, the selection correction term is proportional to the hazard rates and depends only on parameters of the selection equation: $\lambda_i = \frac{\phi(z_i'\theta)}{\Phi(z_i'\theta)}$ with ϕ and Φ denoting the probability density and cumulative distribution functions of the standard normal distribution, respectively.

The wide success that Heckman's estimator has encountered in the literature can be explained with the readiness of application. The procedure provides consistent estimates

---

[4]Details of derivation in the appendix.

given a valid exclusion restriction of one variable in z_i from x_i; additionally, the error terms in the selection and outcome equation need to have a bivariate normal distribution. However, if the true joint distribution of the error terms is not correctly specified, the proceadure produces inconsistent estimates. This lead to a criticism of the Heckman model to be too restrictive to be able to eliminate the selection bias (Goldberger, 1983; Hussinger, 2008).

A fertile line of research (Cosslett, 1983; Robinson, 1988; Powell, 1989; Cosslett, 1991; Ahn and Powell, 1993; Dahl, 2002; Newey, 2009) offers new 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:

$$E[\nu_i | a < \nu_i < b; x_i, z_i] = g(z_i'\theta)$$

with g 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 thus no assumption can be exploited to estimate θ; b) it is not possible to use distributional relationships between the error terms in the selection and outcome equation to estimate $E[\sigma_s e_{si} + \psi_{st}\epsilon_{it} | a \leq \nu_i < b]$.

To overcome the first complication we adopt the semiparametric estimation strategy proposed by Gallant and Nychka (1987) (GN). This estimator does not require assumptions on 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:

$$f_K(\nu) = \frac{1}{\pi} \sum_{k=0}^{2K} \iota_k^* \nu^k \phi(\nu) \tag{7}$$

where:

$$\pi = \int_{-\infty}^{\infty} \left(\sum_{k=o}^{K} \iota_k \nu^k \right)^2 \phi(\nu) \tag{8}$$

and

$$\iota_k^* = \sum_{i=a_k}^{b_k} \iota_i \iota_{k-i} \tag{9}$$

7

with $a_k = max(0, k - K)$ and $b_k = min(k, K)$. $\phi(\nu)$ is the standard normal density. The inclusion of π assures a proper approximation of the density function (i.e.: a function integrating to 1).

The corresponding cumulative function is then given by:

$$F_K(u) = \frac{1}{\theta} \sum_{k=0}^{2K} \iota_k^* \int\limits_{-\infty}^{u} \nu^k \phi(\nu) d\nu \qquad (10)$$

Gallant and Nychka (1987) show that estimates of θ are consistent provided that the order of polynomials K increases with sample size. 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. 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).

We overcome the second complication caused by the non reliance on normality in two different ways. First we apply the method proposed by Cosslett (1991). With this procedure, after having estimated $\widehat{\theta}$ via the GN estimator in the first step, the selection correction term to be included in the primary equation is approximated by J indicator variables $\{I(z_i\hat{\theta} \in \hat{I}_j)\}$. The correction term in our case assumes the form:

$$g(z_i\theta) = \sum_{j=1}^{J} b_j I_{ij}(\widehat{z_i\theta}) \qquad (11)$$

the unknown parameters b_j can be estimated by OLS. Again, consistency requires the order of approximation (J) to increase with sample size. Inclusion of the correction term in the primary equation reestablishes the zero conditional mean on the error term. The estimated equation takes the form:

$$y_{it} = \alpha_s + x_{it}\beta_s + \gamma_s \sum_{j=1}^{J} b_j I_{ij}(\widehat{z_i\theta}) + \omega_{it} \qquad (12)$$

In this equation, by construction, we have that $E[\omega_{is}|s_i = s; x_{it}, z_i] = 0$ and we can apply OLS to obtain consistent estimates of the vector β.

For comparison, we adopt an alternative strategy in the second step as well. Gallant and Nycha's method allows us to produce an estimate of $E[\nu_i|a \leq \nu_i < b]$. The inclusion of these estimates in the wage equation is sufficient to reestablish the zero conditional mean on the error term and the wage equation can be safely estimated by OLS. The exact specification of the alternative estimation of the primary equation is:

$$y_{it} = \alpha_s + x_{it}\beta_s + \gamma_s E[\nu_i | a \le \nu_i < b] + \omega_{it} \tag{13}$$

The variance of observed wage is expressed by[5]:

$$Var[\sigma_s e_{si} + \psi_{st}\epsilon_{it} | a \le \nu_i < b] = \sigma_s^2(1 - \rho_s^2\delta_{si}) + \psi_{st}^2 \tag{14}$$

δ_{si} is referred to as the truncation adjustment. Its expression is: $\delta_{si} = Var[\nu_i | a \le \nu_i < b] - \sigma_\nu^2$ with σ_ν^2 indicating the unconditional variance of ν_i. The sign of δ_{si} determines whether observed wage inequality will understates or overstates potential wage inequality. In fact, in case $\delta_{si} > 0$ (i.e.: $Var[\nu_i | a \le \nu_i < b] > \sigma_\nu^2$) observed wage inequality is greater than potential wage inequality while the opposite occurs when $\delta_{si} < 0$.

We can then identify identify the transitory component via a fixed-effect model. The fixed-effect model allow us to filter out the individual permanent component $\sigma_s e_{si}$. In this way we identify the transitory component ψ_{st}^2. Defining $\zeta_{sit} \equiv \psi_{st}\epsilon_{it}$, our model takes the form:

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)\beta_s + (\zeta_{sit} - \bar{\zeta}_{si}) \text{ if } s_i = s, \tag{15}$$

$\bar{y}_i$, $\bar{x}_i$ and $\bar{\zeta}_{si}$ denote the average over time of the corresponding variables. The transitory component of wage inequality will be identified as the variance of the error term in equation (15).

The permanent component will be identified through a between individual model:

$$\bar{y}_i = \alpha_s + \bar{x}_i\beta_s + \gamma_s g(z_i\theta) + \bar{\omega}_i \tag{16}$$

Where $g(z_i\theta) = \sum_{j=1}^{J} b_j I_{ij}(\widehat{z_i\theta})$ in the Cosslett specification and $g(z_i\theta) = E[\nu_i | a \le \nu_i < b]$ in the alternative specification. Consequently, the permanent component of wage inequality is identified as[6]:

$$\hat{\sigma}_s^2 = \hat{\gamma}_s^2\bar{\bar{\delta}}_s + \widehat{Var}[\omega_{si} | a \le \nu_i < b] - \frac{\sum_t \widehat{\psi}_{st}^2}{\bar{T}} \tag{17}$$

The parameter $\hat{\gamma}_s^2$ is estimated as the coefficient for the correction terms distinguished by schooling level in an OLS regression.; $\widehat{Var}[\omega_{si} | a \le \nu_i < b]$ is estimated as the mean squared of the error term with the between individual estimator $\bar{T} \equiv (\sum_i T_i^{-1}/N)^{-1}$ and $\bar{\bar{\delta}}_s$ is the sample average of the truncation adjustment. We now have all elements to identify wage uncertainty as defined in equation (6): $\hat{\tau}_s^2 = \hat{\sigma}_s^2 - \hat{\gamma}_s^2 + \hat{\psi}_{st}^2$. To obtain a separate identification for the two components of γ_s we need to substitute equation (17) in the expression for

---

[5]See the appendix for derivation.
[6]See the appendix for derivation.

$Var[\sigma_s e_{si} | a \leq \nu_i < b]$ and we obtain: $\widehat{\rho}_s^2 = 1 - \frac{\widehat{\sigma}_{\xi s}^2}{\widehat{\sigma}_s^2}$.

As a result, all parameters of interest - σ_s, ψ_{st}, τ_s and ρ_s - are identified.

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 to 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, geographical residence and environmental characteristics.

We will restrict our analysis to male between the survey years 1991 and 2000 (calendar years 1990 to 1999). By selecting men only we do not have to worry about issues of labor market participation decisions that would rise if also women were included, while the wave restriction will allow us to focus on individuals already out of school and into the labor market. Additionally, we will exclude respondents who do not provide information about parental education, highest grade completed, exact work experience history, hourly rate of pay and ability index as defined below. After having selected the sample according to these guidelines, we have a balanced panel sample consisting of 3,373 individuals observed in 7 susequent 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.

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, North central, 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 this undesired effects we follow Kane and Rouse (1995) and Neal and Johnson (1996). First we regress the original test score on age dummies

Table 1: Summary statistics

Time Invariant Variables

(a) Schooling Variables		Number of siblings	3.63
Years of schooling	13.33		(2.52)
	(2.30)	Family income (1999 dollars)	23,320
Categorical education:			(16,941)
No high school	.15	Black	.25
	(.36)		(.42)
High school	.39	Hispanic	.14
	(.49)		(.35)
Some college	.23	(c) Geographic Controls at age 14	
	(.42)	Urban	.79
Four year college or beyond	.23		(.41)
	(.42)	Residence in Northeast	.19
(b) Ability and Family Background			(.39)
Armed Forces Qualifying Test score (adjusted)	43.30	Residence in South	.33
	(29.21)		(.47)
Highest grade mother	11.10	Residence in West	.19
	(3.20)		(.39)
Highest grade father	11.12	(d) Instrument for Schooling	
	(3.93)	Average unemployment rate (%)	7.32
			(.75)

Time Variant Variables

Calendar year	1990	1993	1995	1997	1999
Actual work experience	10.03	12.77	14.40	16.09	17.93
	(3.58)	(4.05)	(4.35)	(4.68)	(5.04)
Log hourly earnings	2.18	2.16	2.28	2.26	2.21
	(.98)	(1.06)	(1.06)	(1.14)	(1.25)
Unemployment rate (%)	5.81	8.70	6.01	5.59	4.12
	(2.13)	(2.51)	(1.74)	(1.87)	(1.05)

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

and quarter of birth, then we replace the original test score with the residuals obtained from this regression.

For family income we intend family income at age 17. If no measure for family income at 17 is recorded, we plug in the family income at the closest age to 17 available.

Besides the aforementioned controls common to choice and wage equation, the latter is augmented by the inclusion of experience in the labor market. 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.

The instrument for schooling that we exploit for identification is the average national unemployment rate in the economy during the years that each respondent spent in school after mandatory schooling age. The last year of mandatory schooling is also included in the average. 205 individuals dropped out of school before mandatory schooling years were completed, therefore we exclude them from our analysis reducing the sample to 3,168 individuals. The intuition behind this instrument is that the unemployment rate that an individual would have

to face in the labor market influences his outside option making the possibility to drop out of school more or less attractive. To our knowledge, the only paper exploiting the same instrument is Arkes (2010), but with state level information on youth unemployment. Information about unemployment rates are taken from the Current Population Survey (CPS). The CPS is freely accessible from the Internet[7] and 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. Since it is possible that past level of unemployment affect current levels and thus wages and that labor market conditions at entry might carry over in subsequent years, we include the unemployment rate for the same years wages are observed directly in the wage equation. The assumption is then that conditional on present unemployment rates, past unemployment rates have no effect on present wages.

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.

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: *permanent component* (σ_s^2); *transitory component* (ψ_{st}^2); *unobserved heterogeneity* (ν_i) and *wage uncertainty* (τ_s^2).

5.1   Selection of the preferred model and first stage

In the first stage we estimate the choice equation via the Gallant and Nychka (1987) method discussed in section 3. In this way we obtain estimates for the density function of the unobserved heterogeneity component and we can substitute these estimates in the wage equation reestablishing the zero conditional mean of the error term. In this method it is essential for the degree of polynomial K to increase with sample size. To select the best approximation we apply two standard methods for selection: AIC and BIC[8]. 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.

We start from the 3rd degree polynomial since this is the first model generalizing the ordered probit (OP) to the semiparametric case. We can see that two out of three tests conducted - likelihood ratio and AIC criterion - select the 8th degree polynomial. BIC on the other hand prefers the 6th degree polynomial. We choose the former as our preferred model

---

[7]http://data.bls.gov/data/ accessed the 15/06/2010.
[8]See Cameron and Trivedi (2005) for a textbook discussion of the two criteria.

Table 2: Model comparison

K	log likelihood	LR-test of OP	p-value	LR-test of K-1	p-value	AIC	BIC
OP	-20,573.45					41,202.91	41,424.66
3	-20,544.56	57.78	.000	57.78	.000	41,149.13	41,386.72
4	-20,504.04	138.82	.000	81.04	.000	41,070.08	41,315.6
5	-20,502.54	141.83	.000	3.01	.083	41,069.07	41,322.51
6	-20,493.57	159.78	.000	17.94	.000	41,053.13	41,314.48
7	-20,492.92	161.06	.000	1.29	.257	41,053.85	41,323.12
8	**-20,489.37**	**168.17**	**.000**	**7.11**	**.008**	**41,048.73**	**41,325.92**

and use this specification to estimate the first stage of our selection model. We have not carried on our test on higher order polynomials since the maximization of the log-likelihood function with $K = 9$ does not converge[9].

Results of ordered Probit model and for the Gallant and Nychka (GN) procedure at 3rd and 8th degree polynomial are presented in table 3. It has to be noted that estimates of θ_s cannot be compared directly across models because the variances differ (Stewart, 2004). What we can compare are ratios of different coefficients from different models. For our purposes the most relevant result concerns our instrument and its very strong impact on schooling choices irrespective of the selected model. In the OP model, the *t-statistic* for the instrument is a reassuring 29.08. Even stronger is the impact that average unemployment rate has in the GN(8) model, which is our favorite model. In this model the *t-statistic* is 49.60. Remember that as a rule of thumb an *F-statistic* bigger than 10 is deemed to be sufficient to rule out concerns of weak instruments. We can safely assume that average unemployment rate in the labor market correlate with schooling decision. The sign of the effect of our instrument over schooling length is, as expected, positive. Our estimates confirm the assumption that pupils take current job market conditions into account when deciding whether to continue or terminate their educational path.

The other covariates all show the expected signs. Mother education, ability and family income are positively correlated with educational achievements while number of siblings negatively so. Father education is not correlated with educational choice. The only surprising results regards the coefficients for African-American and Hispanic students. Belonging to a ethnic minority encourages education at least in our sample. The surprising result is also encountered by Cameron and Taber (2004) and Chen (2008) on the same data.

5.2 Identification of selection correction term

Separate identification of wage uncertainty (τ_s^2) and unobserved heterogeneity (ν_i) requires the estimation of four selection correction terms, one for each schooling level[10]. The Cosslett procedure produces as many 'correction' dummies as intervals for which the distribution is

---

[9]The process was stopped after the 200th iteration. Data and programs are available on request.

[10]See section (2.1) and the appendix.

Table 3: First stage estimates for different values of K

	OP		GN(3)		GN(8)	
Avg. unemp. rate	.889***	(.018)	.975***	(.012)	.934***	(.019)
Mother attended college	.329***	(.028)	.328***	(.028)	.322***	(.029)
Father attended college	.534***	(.027)	.497***	(.029)	.485***	(.031)
Highest grade mother	.017***	(.004)	.018***	(.004)	.015***	(.004)
Highest grade father	-.002	(.004)	.002	(.003)	.001	(.003)
Number of siblings	-.024***	(.004)	-.019***	(.004)	-.018***	(.004)
Family income bottom quartile	-.015	(.029)	.023	(.029)	.024	(.027)
Family income second quartile	-.008	(.027)	.013	(.026)	.013	(.025)
Family income third quartile	.002	(.028)	.001	(.024)	-.004	(.023)
Family income top quartile	.187***	(.025)	.182***	(.024)	.181***	(.024)
AFQT score (adjusted)	.025***	(.000)	.023***	(.001)	.023***	(.001)
Black	.515***	(.023)	.487***	(.025)	.468***	(.028)
Hispanic	.331***	(.030)	.319***	(.029)	.298***	(.030)
Cut-off point (κ_1)	6.621***	(.162)	6.621		6.621	
Cut-off point (κ_2)	8.246***	(.165)	8.187***	(.030)	8.185***	(.046)
Cut-off point (κ_3)	9.192***	(.166)	9.158***	(.046)	9.122***	(.071)
Polynomial:						
1			-.591***	(.064)	-.306***	(.053)
2			.185***	(.043)	.169***	(.064)
3			.019***	(.008)	-.086	(.083)
4					-.121***	(.035)
5					.026	(.016)
6					.020***	(.005)
7					-.001	(.001)
8					-.001***	(.000)
Wald χ^2	8,693.37		9,911.69		3,785.80	

Note: Geographic and cohort controls added. Geographic controls include the urban dummy and three regional dummies for residence at 14. Cohort controls include a full set of birth cohort dummies and age in the initial survey year. */**/*** indicate confidence levels of 10/5/1 percent respectively. Standard errors in parentheses.

partitioned into. In our case we have set each interval to contain 500 observations ordered according to the estimated score in the first stage. Dividing the entire sample in groups of 500 observations generates 43 intervals: 7 for high school drop outs; 15 for high school graduates; 9 for college drop outs and 9 for college graduates. In order to reduce 43 correction coefficients to four we adopt a two step estimation strategy. First we include the dummy variables in the wage equation differentiating for schooling level. In this way we obtain four correction terms one for each educational category. In a second stage the four correction terms are included in the a wage equation over the entire sample of individuals irrespective of educational level along with three educational dummies. These are the corresponding coefficients γ_s in table 4, Cosslett column.

5.3 Wage equation

In table 4 we report estimates of equation (16) with the two alternative specification for the function $g(.)$ in columns 2 and 3 and the simple uncorrected between-individuals effect model in column 1. The Cosslett specification expresses the correction function as $g(z_i\theta) = \sum_{j=1}^{J} b_j I_{ij}(\widehat{z_i\theta})$ while what we refer to as Olsen estimation in table 4 augments the wage equation with the inclusion of $E[\nu_i | a \leq \nu_i < b]$ calculated from the Gallant and Nychka polynomial approximation in the first stage and assunmes linearity in the error terms. Both estimates are based on a between-individual effect model and determine the causal impact of education on wages. The parameters of interest here are the schooling coefficients and the four correction terms.

Column 2 and 3 show that GLS estimation underestimates the schooling coefficients for all categories except high school graduates for whom overestimation occurs. These results are robust to the specification of correction function. It has to be noted that, even if small, the coefficient for high school graduates shows an unexpected negative sign; nevertheless the coefficient is not significantly different from zero in the Olsen specification and significant only at a 10% confidence level for the Cosslett estimator. The other two schooling coefficients in column 2 are very precisely estimated and show an underestimation of schooling benefit of 40% and 43% for college drop outs and college graduates respectively if self-selection issues are not addressed. Comparable level of underestimation is also detected by the Olsen estimator even though only the coefficient for college graduates is estimated with enough precision.

The four correction terms in column 2 highlight the importance of accounting for self-selection when studying the casual impact of education on wages. Coefficients for three out of four correction terms are positive, large and significant the only exception is represented by the coefficient for college graduates. The Olsen specification also reports large correction coefficients for high school graduates and college drop outs, but only the former is statistically significant and only at a 10% confidence level while the latter presents huge standard errors.

Table 4: Wage equation

	GLS (I)	Cosslett (II)	Olsen (III)
High school	.049***	-.071*	-.054
	(.013)	(.015)	(.097)
Some college	.091***	.166***	.243
	(.014)	(.027)	(.128)
4 yr. college or beyond	.351***	.502***	.453***
	(.017)	(.034)	(.079)
Experience	.132***	.133***	.132***
	(.005)	(.005)	(.005)
Experience^2	-.001***	-.001***	-.001***
	(.000)	(.000)	(.000)
AFQT score (adjusted)	.005***	.004***	.005***
	(.000)	(.000)	(.000)
Highest grade mother	-.002	-.003	-.002
	(.001)	(.002)	(.002)
Highest grade father	-.000	-.001	-.000
	(.001)	(.002)	(.001)
Number of siblings	.003	.003*	.003
	(.002)	(.002)	(.002)
Family income bottom quartile	.011	.012	.010
	(.015)	(.017)	(.015)
Family income second quartile	-.041***	-.041**	-.041**
	(.014)	(.013)	(.014)
Family income third quartile	-.007	-.008	-.008
	(.013)	(.013)	(.013)
Family income top quartile	.065	.060***	.065***
	(.013)	(.013)	(.013)
Unemployment rate	-.001	-.001	-.001
	(.002)	(.002)	(.002)
Black	.047***	.035*	.049**
	(.016)	(.017)	(.016)
Hispanic	.044	.039**	.045**
	(.017)	(.017)	(.017)
γ_0		.860***	.089
		(.186)	(.158)
γ_1		.407***	.835*
		(.047)	(.398)
γ_2		.756***	-1.726
		(.114)	(1.575)
γ_3		.031	-.047*
		(.080)	(.023)
R^2	.391	.394	.391
N	20,398	20,398	20,398

Note: Geographic controls include the urban dummy and three regional dummies for residence at 14. Cohort controls include a full set of birth cohort dummies and age in the initial survey year. */**/*** indicate confidence levels of 10/5/1 percent respectively. Bootstrapped standard errors on 200 replications in parentheses. Unemployment rate calculated on CPS data.

Interestingly enough, the only coefficient not significant in the Cosslett specification is now negative and significant albeit only at a 10% confidence level.

The other covariates are very robust to specification and beside the coefficient for African-Americans and Hispanic all show the expected sign.

Overall regression based on the Cosslett method appear to offer the better fit: schooling coefficients are always significant and correction terms, beside the one for college graduates, as well. For this reason we base the identification of our parameters of interest shown in section 5.4 on these results.

5.4 Main results

The inequality measures for variance of wage residuals are: i) the observed wage inequality given the choice of schooling ($Var[y_{sit}|s_i = s, x_{it}, z_i]$); ii) the potential wage inequality purged of selection and truncation biases ($\sigma_s^2 + \psi_{st}^2$); and iii) the uncertainty in potential wages, after removing truncation and selection biases and incorporating unobserved heterogeneity factors (τ_s^2).

In table 5 panel A we report measures for the observed wage inequality. The observed wage inequality is the sum of two factors $\sigma_s^2(1 - \rho_s^2)$ and ψ_{st}^2. The first is the permanent component, identified by the mean squared residuals in the between-individuals model not corrected for selectivity. The second is the transitory component identified by exploiting the mean-squared errors of the fixed-effects model.

From table 5 we can see that high school graduates are those showing a lower variance both in the permanent (panel A) and transitory (panel B) component. On the other hand, college graduates have the highest variance. Observed wage inequality monotonically increases after high school and the results holds for both of its components. This result is also encountered by Chen (2008). College enrollment causes a 15% increase in inequality and college completion an additional 23%.

If individuals act upon private information about their personal tastes and inclinations to select the preferred risk/pay-off profile, estimates presented in panel A are biased measures of the real level of wage inequality that each schooling level entails. In panel C we show our estimates of the permanent component corrected for selection and truncation biases. It is immediately evident the huge underestimation of wage inequality if truncation and selection biases are not taken into account. Observed wage inequality accounts for only 30% of potential wage inequality in the case of high school and college drop outs and around 70% for high school graduates. These results abundantly exceeds previous parametric estimations (Chen, 2008) for which underestimation varies between 2% and 30% depending on educational level.

On the contrary, observed wage inequality overstates potential inequality in the case of

	Less than high school	High school	Some college	4 yr. college and beyond
I. Observed wage inequality				
A. Permanent component	.308	.304	.352	.421
B.Transitory component (ψ_{st}^2)	.156	.131	.142	.207
Age 25-30	-.064	-.052	-.043	-.105
Age 31-36	-.060	-.054	-.047	-.110
Age 37-42	-.025	-.023	-.020	-.055
Observed inequality (A+B)	.464	.435	.494	.628
II. Potential wage inequality				
C. Permanent component (σ_s^2)	1.007	.444	1.082	.311
D. Transitory component (same as B)				
Potential wage inequality (C+D)	1.163	.575	1.224	.518
III. Wage uncertainty				
E. Correlation coefficient (ρ_s)	.694	.494	.588	.045
F. Permanent component (C-CxE^2)	.521	.335	.707	.311
G. Transitory component (same as B)				
Degree of wage uncertainty (τ_s^2)	.365	.363	.613	.420
Unobserved heterogeneity (ν_i)	.486	.108	.374	.001

Note: Estimates of transitory volatility are obtained by regressing ψ_{st}^2 on age dummies and categorical educational variables.

college graduates[11]. This result suggest that if we were to assign education randomly, intra educational wage variability would be smaller than the observed variability for this category. At first sight the result is puzzling and former parametric estimates (Chen, 2008) do not allow for this possibility since in the parametric case with normal distribution of the error terms, observed wage variability always understates potential wage variability by construction. Our model does not impose such restriction. In reality we have no prior to suggest that observed wage variability systematically understates potential wage variability. Even if individuals do posses superior knowledge about themselves, we cannot assume that they will try to minimize risk. It is reasonable to imagine that other more compelling factors enter their utility function besides risk minimization and these other factors might offset considerations of future wage variability. If this is the case we cannot exclude *a priori* the apparently odd result that we encounter.

[11]The terminology observed and potential wage inequality is misleading in some respect. Using this terminology can generate the impression that potential wage inequality is the maximum wage inequality possible by definition. What we indicate with potential wage inequality, instead, is within educational category wage variance in case education was randomly assigned. We have decided to maintain this terminology since it has been already used in the literature.

The hierarchical order of the four educational categories regarding potential wage variability is changed once truncation and selection biases are corrected for. If we control for self-selection dropping out of college leads to the highest variance in the permanent component. College entry yields an enormous 112% increase in wage variability. College completion, on the other hand, diminishes variability by 64% the lowest level among the four groups. It is worth nothing that the transitory component is not influenced by self selection since it is modeled as exogenous shocks that individuals can not act upon by construction, therefore all relative changes have to be attributed to the permanent component.

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 F. It is interesting to compare the two estimates of panel C and panel F since from this comparison we can already understand the importance of unobserved heterogeneity for wage variability. Controlling for unobserved heterogeneity considerably diminishes all estimates of permanent component with the exception of college graduates who remain unaffected. These results are due to the strength of correlation between wages and schooling factor. The parameter ρ_s in panel E describes a positive and strong correlation between the two for three out of four categories, but a weak one for college graduates. A positive ρ_s also tell us that the labor market rewards people with a high unobserved schooling factor. This result deviates from previous parametric estimates (Chen, 2008).

Estimates for unobserved heterogeneity in panel G confirm the intuition. Unobserved heterogeneity is an important element in explaining potential wage variability for the three lowest educational categories. It explains a share varying between 42% and 18% of potential wage inequality. Again, the lone exception is given by college graduates for whom accounting for unobserved heterogeneity has no impact.

The decomposition of the two determinants of inequality - uncertainty and heterogeneity - shows that private information has an important effect on the identification of a causal relation between risk and wages for three out of four categories. Even if risk explains most of wage variability, or the totality of it in the case of college completion, for college drop outs and high school graduates, in the case of high school drop outs is unobserved heterogeneity and not risk the main contributor to wage variability. The contribution of unobserved heterogeneity for the indentification of the potential wage inequality measures encountered in our estimates is well above the level that previous parametric estimates detected. The other side of the coin is that the share of wage variability that can be attributed to uncertainty or risk is much lower than what emerges from previous estimates (Chen, 2008). Risk is maximum for college drop outs while heterogeneity peaks for high school drop outs.

With the exception of high school drop outs, the majority of wage variation that we encounter in the data is anyhow due to wage uncertainty. As for the other parameters of

interest risk increases with college entry. For risk-averse individuals this particular feature of wage variability might discourage investment into further education in the absence of compensating mechanisms. The most immediate compensation that comes to mind is via higher wages, thus, the increase in category wage inequality that has characterized the US economy in the past decades might be a compensation to superior risk. On the other hand, college enrollment and even better, college completion, opens up more opportunities both in terms of further education (i.e.: M.Sc. or PhD etc.) and in terms of possible careers. The increased uncertainty might simply reflect this larger choice set compared to high school students and high school drop outs. Controlling for the number of occupational choices that each education gives access too would shed some light on the exact mechanism, but that exceeds the scope of the present paper.

Our results clearly show that potential and observed wage variability are prominent in the two categories of drop outs, this is an interesting result, possibly related to some common features of these two group of individuals. The level of unobserved heterogeneity detected in our estimation is far from irrelevant and suggest that self-selection severely biases estimates of the causal impact of schooling on risk. The low impact of unobserved schooling factor and the overestimation of potential by observed wage inequality for college graduates suggest that the use that individuals make of private information when selecting a university degree might not be at odd with assumptions of risk minimizing agents at least for this specific category.

6 Conclusion

In this paper we apply two semiparametric techniques developed by Cosslett (1983) and Gallant and Nychka (1987) to distinguish various components of wage variance. We extend the original technique from the dichotomous to the polychotomous case providing consistent estimates of: within education potential wage variation, accounting for selection and truncation biases; degree of private information owned by the individuals and used to select their favorite level of education; magnitude of pure risk that every education level entails.

Our results indicate that all decompositions of wage uncertainty (observed, potential and the transitory component) are maximum for high school and college drop outs. Holding a degree, either from high school or college, is found to decrease wage variability by half. This is an aspect of nonlinear benefits of education rarely considered before.

Other important results are the much higher impact of unobserved heterogeneity on wage variability and the finding that in the case of college graduates, if education was randomly assigned to individuals, their within educational wage variability would be reduced compared to the *status quo*. We offer two interpretations to this finding: when selecting a college degree individuals might not use their private information to minimize their future wage

variability or they might not posses sufficient information on how their personal tastes and inclination for schooling match with different college degrees. This might not be the case for the other three categories since the different options in terms of educational paths that they offer are more limited. What is clear from our analysis is that investing in education has a significant impact on risk of future wages and this is especially true for college education if not finished. If compensation for risk exists in the labor market and if this compensation works via higher wages for riskier occupations, our findings might contribute to explain the increase of educational level inequality observed in the U.S. and other advanced economies in the past 30 years.

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, the increased uncertainty in pay-offs that we encounter in our estimates might be offset 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 let 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.

Arkes, J.: 2010, 'Using unemployment rates as instruments to estimate returns to schooling'. *Southern Economic Journal* **76**(3), 711–722.

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

Cameron, C. and P. Trivedi: 2005, *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

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

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.

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.

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

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.

Heckman, J.: 1974, 'Shadow Prices, Market Wages and Labour Supply'. *Econometrica* **42**(4), 679–694.

Heckman, J.: 1976, 'The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models'. *Annals of Economics and Social Measurement* **5**(4), 475–492.

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

Hussinger, K.: 2008, 'R&D and Subsidies at the Firm Level: an Application of Parametric and Semiparametric Two-Step Selection Models'. *Journal of Applied Econometrics* **23**(6), 729–747.

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.

Maddala, G. S.: 1983, *Limited-Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

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

Olsen, R. J.: 1980, 'A Least Squares Correction for Selectivity Bias'. *Econometrica* **48**(7), 1815–1820.

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.

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

A   Appendix: identification of ρ_s and σ_s

To see how ρ_s and σ_s are identified consider the model as formalized by Chen (2008):

$$y_{si} = x_{si}\beta_s + \sigma_s e_{si} + \psi_{st}\epsilon_{it}, \tag{18}$$

$$s_i^* = z_{si}\theta_s + \nu_i \tag{19}$$

where $s_i^* = s$ if $a \leq \nu_i \leq b$. This is the usual sample selection model with ordered censoring rules. Following Olsen (1980) we make the following assumption on a part of the error term:

$$\sigma_s e_{si} = \sigma_{e\nu s}\nu_i + \xi_s \tag{20}$$

where $\sigma_{e\nu s} \equiv \sigma_s \rho_s$ is the covariance coefficient between the error term in the outcome equation and the error term in the choice equation. As Chen (2008) does, we assume that the error term ν_i in the choice equation is correlated with e_{si}, but not with ϵ_{it}. We also assume that ξ_s is independent of ν_i and that the ξ_s are uncorrelated across schooling levels. Additionally, $Var[\xi_s] = \sigma_{\xi s}^2$. From these assumptions we obtain that:

$$E[\sigma_s e_{si} + \psi_{st}\epsilon_{it}|a \leq \nu_i < b] = E[\sigma_s e_{si}|a \leq \nu_i < b] = \sigma_s \rho_s E[\nu_i|a \leq \nu_i < b] \tag{21}$$

$$Var[\sigma_s e_{si}|a \leq \nu_i < b] = \sigma_{e\nu s}^2 Var[\nu_i|a \leq \nu_i < b] + \sigma_{\xi s}^2 \tag{22}$$

To re-establish the zero conditional mean of the error term in the outcome equation (16) in presence of self-selection, we need a correction term accounting for $E[\nu_i|a \leq \nu_i < b]$ and an estimate for $\lambda_s \equiv \sigma_s \rho_s$. In fact, the equation:

$$y_{is} = \alpha_s + x_{it}\beta_s + \lambda_s g(z_i\theta) + \omega_{is} \tag{23}$$

can be consistently estimated with OLS since $E[\omega_{is}|x_{it}, z_i] = 0$ by construction. $g(z_i\theta) = E[\nu_i|a \leq \nu_i < b]$ is an unknown function entered using the method of Gallant and Nychka (1987) or Cosslett (1983). We can obtain estimates for both $E[\nu_i|a \leq \nu_i < b]$ and $Var[\nu_i|a \leq \nu_i < b]$ by approximating the unknown density function by a Hermite series of K degrees polynomials.

$$\widehat{E}[\nu_i|a \leq \nu_i < b] = \frac{\int_a^b \nu_i f_K(\nu_i)d\nu}{Pr[a \leq \nu_i < b]} \tag{24}$$

$$\widehat{Var}[\nu_i|a \leq \nu_i < b] = \widehat{E}[\nu_i^2|a \leq \nu_i < b] - \widehat{E}[\nu_i|a \leq \nu_i < b]^2 \tag{25}$$

where K represent the degree of polynomial used and $f_K(\nu_i)$ is a density function at ν_i assuming the form:

$$f_K(\nu_i) = \frac{1}{\pi}\sum_{k=0}^{2K} \iota_k^* \nu_i^k \phi(\nu_i) \tag{26}$$

$\pi = \int \sum_{k=0}^{2K} \iota_k^* \nu_i^k \phi(\nu_i) d\nu = \sum_{k=0}^{2K} \iota_k^* \mu_k$ is a scaling factor ensuring a proper approximation for the density function (i.e.: a function integrating to 1), μ_k is the kth moment of the standard normal distribution and ϕ the standard normal density.

According to our distributional assumptions the variance of the permanent component from an individual standpoint is given by: $Var[\sigma_s e_{si}|x_{it}] = Var[\sigma_s \rho_s \nu_i + \xi_s|x_{it}] = \sigma_s^2 \sigma_\nu^2 \rho_s^2 + \sigma_{\xi s}^2$. Remembering that $Var[\sigma_s e_{si}|x_{it}] = \sigma_s^2$ we obtain:

$$\sigma_{\xi s}^2 = \sigma_s^2(1 - \rho_s^2 \sigma_\nu^2) \tag{27}$$

Substituting (27) into (22) and rearranging, we obtain the variance of observed wages corrected for truncation:

$$Var[\sigma_s e_{si}|a \le \nu_i < b] = \sigma_s^2(1 - \rho_s^2 \delta_{si}) \tag{28}$$

where we can estimate δ_{si} by: $\widehat{\delta}_{si} = \widehat{Var}[\nu_i|a \le \nu_i < b] - \widehat{\sigma}_\nu^2$. The parameter $\hat{\gamma}_s = \widehat{\rho_s \sigma_s}$ is estimated as the coefficients for the correction terms distinguished by schooling level in an OLS regression.

In this model the error term is composed by two elements, the permanent component $(\sigma_s e_{si})$, for which we have explicated the variance in (22), and the transitory shocks $\psi_{st} \epsilon_{it}$. The expression for the variance of the complete error term is:

$$Var[\sigma_s e_{si} + \psi_{st} \epsilon_{it}|a \le \nu_i < b] = Var[\sigma_s e_{si}|a \le \nu_i < b] + \psi_{st}^2 = \sigma_s^2(1 - \rho_s^2 \delta_{si}) + \psi_{st}^2 \tag{29}$$

$\hat{\gamma}_s = \widehat{\rho_s \sigma_{es}}$; $Var[\sigma_s e_{si} + \psi_{st} \epsilon_{it}|a \le \nu_i < b] = Var[\omega_{si}|a \le \nu_i < b]$ can be consistently estimated as the mean squared errors of the residuals in the between individual effects model of expression (23). ψ_{st}^2, as explained in Chen (2008), is identified from the fixed-effect model as the variance of residuals in equation (15). Substituting these elements into (29) and rearranging we identify the permanent component of wage inequality corrected for truncation as:

$$\widehat{\sigma}_s^2 = \widehat{Var}[\omega_{si}|a \le \nu_i < b] - (\widehat{\sigma_s \rho_s})^2 \overline{\widehat{\delta}}_s - \frac{\sum_t \widehat{\psi}_{st}^2}{\overline{T}} \tag{30}$$

With $\overline{T} \equiv (\sum_i T_i^{-1}/N)^{-1}$ and $\overline{\widehat{\delta}}_s$ is the sample average of the truncation adjustment. Substituting (30) into (27) we obtain estimates for $\sigma_{\xi s}^2$ and $\widehat{\rho}_s^2 = \frac{\widehat{\sigma}_s^2 - \widehat{\sigma}_\xi^2}{\widehat{\sigma}_s^2 \widehat{\sigma}_\nu^2}$.

Note that we have identified ρ_s^2 and σ_s^2 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:

linearity in the equation of the error term $(\sigma_s e_s = \omega_{\epsilon \nu s} \nu_i + \xi_s)$;

1. the distribution of ξ_s is independent of ν_i.