On the uncertain nature of human capital investments
Mazza, J.

Citation for published version (APA):
Mazza, J. (2012). On the uncertain nature of human capital investments Amsterdam: Universiteit van Amsterdam

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 5

Separating risk in education from heterogeneity: a semiparametric approach

5.1 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 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. Neglecting to disentangle risk and heterogeneity will cause an overestimation of risk and, in turn, an underestimation of the risk premium offered in the labor market in the form of higher wages. The sum of risk and unobserved heterogeneity forms what, in the reminder of this Chapter, we refer to as *wage dispersion*.

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 normality on the distribution of disturbances in both the schooling choice and wage 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 Chapter we apply a two stage semi parametric estimation method. Similar to the two step Heckman method we first estimate a selection equation from which we build four selection correction terms, one for each level of schooling, 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 hetero-

---

\(^1\)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 when referring to it.
5.2. Econometric specification

geneneity obtained by Gallant and Nychka’s procedure in the wage equation. In the continuation of this Chapter we refer to this method as the linear method since it relies on assumptions of linearity of the error term. 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 study 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 wage inequality robust to selection and truncation biases and then distinguish between the two components of inequality, heterogeneity and risk, semiparametrically. Results suggest that observed wage dispersion and potential wage dispersion are maximum for high school and college drop outs. Risk increases with college entry while unobserved heterogeneity accounts for a larger share of potential wage dispersion than previously established.

5.2 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 real 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.
The Chen (2008) model, that we use as a starting point, offers a solution to both these identification puzzles. Chen (2008)'s 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.

### 5.2.1 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:

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

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 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:

\[
y_{sit} = \alpha_s + x_{it}\beta_s + \sigma_s\epsilon_{si} + \psi_{st}\epsilon_{it} \text{ if } s_i = s, \tag{5.1}
\]

\(\alpha_s\) is a schooling specific constant; \(\beta_s\) 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 \((\sigma_s e_{si})\); the error term \(\psi_{si}\epsilon_{it}\) denotes transitory shocks uncorrelated with personal characteristics and across time; \(e_{si}\) and \(\epsilon_{it}\) are random unit root variables uncorrelated with each other. Inequality in potential wages within schooling levels equals \(\sigma_s^2 + \psi_{si}^2\): 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:

\[
s_i = s \text{ if } a_{si} \leq \sigma_v v_i < a_{s+1,i} \text{ for } s = 0, 1, 2, 3
\]  

In this expression \(v_i\) 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. \(v_i\) is unobservable to the researcher. \(a_{si}\) is the minimal or maximum level of \(v_i\) for individuals that choose schooling level \(s\) and it is determined by the relation:

\[
a_{si} = \kappa_s - z_i \theta.
\]  

The vector \(z_i\) contains observable characteristics \(x_{it}\) plus an instrument for education\(^2\); \(\theta\) is the vector of coefficients for \(z_i\) and \(\kappa_s\) (with \(s = 0, 1, 2, 3\)) are constants with \(\kappa_0 = -\infty\) and \(\kappa_4 = \infty\), respectively. We assume \(v_i\) to be uncorrelated with the transitory shocks \(\epsilon_{it}\), but to be correlated with the fixed effect \(e_{si}\). The correlation coefficient is indicated by \(\rho_s\).

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

\(^2\)Identification of true parameters \(\beta\) might be achieved even with \(x_i = z_i\) through non linearity 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 \(\beta\) (Vella, 1998; Cameron and Trivedi, 2005).
with \( E[\xi_{si}|\nu_i] = 0 \), \( \text{Var}[\sigma_s e_{si}|x_{it},z_{it}] = \sigma_s^2 \) and \( \text{Cov}[\sigma_s e_{si},\sigma \nu_i] = \gamma_s = \rho_s \sigma_s \sigma \nu_i \).

From the individual viewpoint the expected value of future wages is given by:\(^3\)

\[
E[y_{sit}|z_{it},x_{it},\nu_i] = \alpha_s + x_{it} \beta_s + \gamma_x \nu_i \quad (5.5)
\]

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 then researchers, private information \( (\sigma \nu_i)\) 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 (2008), we indicate this risk with \( \tau_{st}^2 \). Using equation (5.4) and our distributional assumptions on disturbances, given above, we obtain\(^4\) an expression for risk as the variance of the error term in (5.1) given observed and unobserved heterogeneity\(^5\):

\[
\tau_{st}^2 = \text{Var}[\sigma_s e_{si} + \psi_{st} \epsilon_{it}|x_{it},z_{it},\nu_i] = \sigma_s^2 (1 - \rho_s^2 \sigma^2) + \psi_{st}^2 \quad (5.6)
\]

Remembering that the extent of predictability of wage dispersion from the personal standpoint is expressed by the correlation coefficient \( \rho_s \) equation (5.6) makes the formal link between wage dispersion and private information explicit. In fact, if the correlation between unobserved schooling factor \( (\nu_i) \) and permanent component of wage dispersion \( (e_{si}) \) is perfect (i.e.: \( \rho_s = 1 \)) the subject can predict exactly the permanent part of his wage variability and risk is only caused by transitory shocks \( (\psi_{st}^2) \). On the other hand, if correlation is absent (i.e.: \( \rho_s = 0 \)) the subject does not possess any additional information compared to the researcher and risk is observed in the data.

\(^3\)See the appendix for derivation.
\(^4\)Details of derivation in the appendix.
\(^5\)It is easy to note that if we would impose normality on the error term for the schooling equation, \( \tau_{st}^2 \) would assume the same specification as in Chen (2008).
Rearranging equation (5.6) as \( \sigma^2_s + \psi^2_{st} = \gamma^2_s + \tau^2_{st} \) helps visualizing how potential wage dispersion \( (\sigma^2_s + \psi^2_{st}) \) is the sum of two elements: variance of unobserved heterogeneity \( (\gamma^2_s) \) and risk \( (\tau^2_{st}) \). 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 \( (\tau^2_{st} < \sigma^2_s + \psi^2_{st}) \).

5.3 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 parameter 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 \( \lambda_i \). Under normality, the selection correction term is proportional to the hazard rates and depends only on parameters of the selection equation: \( \lambda_i = \frac{\phi(z_i')}{\Phi(z_i')} \) with \( \phi \) and \( \Phi \) 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 given a valid exclusion restriction of one variable in \( z_i; \) 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 normal, the procedure produces inconsistent estimates.

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:

\[
E[\sigma_i v_i | a_{si} \leq \sigma_i v_i < a_{s+1}; x_i, z_i] = g(z_i')
\]
with \( g(\cdot) \) some unknown function. The semiparametric approach differs from the parametric one in two important dimensions: a) no distributional assumption on \( \nu_i \) is specified and thus no assumption can be exploited to estimate \( \theta \); b) no imposition on the joint distribution of the error terms in the selection and outcome equation is used when estimating \( E[\sigma s e_{si} + \psi s t \epsilon_{St} | a_{si} \leq \sigma_i \nu_i \leq a_{si+1}] \).

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 \( \nu_i \) in the selection equation to estimate \( \theta \). 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:

\[
f_K(\nu) = \frac{1}{\eta} \sum_{k=0}^{2K} \lambda_i^* \nu^k \phi(\nu)
\]

where:

\[
\eta = \int_{-\infty}^{\infty} \left( \sum_{k=0}^{K} \lambda_i^k \nu^k \right)^2 \phi(\nu)
\]

and

\[
\lambda_k^* = \sum_{i=a_k}^{b_k} \lambda_i \lambda_{k-i}
\]

with \( a_k = \max(0, k - K) \) and \( b_k = \min(k, K) \); \( \phi(\nu) \) is the standard normal density and \( \lambda_i = (\lambda_0, \lambda_1, ..., \lambda_K) \) are parameters to be estimated that characterize the density of \( \nu_i \) with \( \lambda_0 = 1 \). The inclusion of \( \eta \) assures a proper approximation of the density function (i.e.: a function integrating to 1).

The corresponding cumulative function is then given by:

\[
F_K(u) = \frac{1}{\eta} \sum_{k=0}^{2K} \int_{-\infty}^{u} v^k \phi(v) dv
\]

Gallant and Nychka (1987) show that estimates of \( \theta \) 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. In this case, standard properties of (parametric) simulated maximum likelihood apply and, in this context, bootstrapped standard errors are appropriate. 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).

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 $\hat{\theta}$ 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 $\{I_{ij} = \ell(z_i \hat{\theta} \in \hat{I}_j)\}$. The correction term in our case assumes the form:

$$\gamma_s g_s(z_i \theta) = \sum_{j=1}^{J} b_{js} I_{ij}(\hat{z}_i \theta)$$

the unknown parameters $b_{js}$ 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:

$$y_{sit} = \alpha_s + x_{it} \beta_s + \sum_{j=1}^{J} b_{js} I_{ij}(\hat{z}_i \theta) + \omega_{it}$$

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 vectors $\beta_s$ and $b_{js}$.

For comparison, we adopt an alternative strategy in the second step as well. Gallant and Nychka’s method allows us to produce an estimate of $E[\sigma_v v_i|a_{si} \leq \sigma_v v_i < a_{s+1,i}]$. The inclusion of these estimates plus a coefficient related to the correlation in the wage equation is sufficient to reestablish the zero conditional
Separating risk in education from heterogeneity: a semiparametric approach

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:

\[ y_{it} = \alpha_s + x_{it}\beta_s + \gamma_s E[\sigma_v v_i | a_{si} \leq \sigma_v v_i < a_{s+1,i}] + \omega_{it} \] (5.13)

The variance of observed wage is expressed by:

\[ \text{Var}[\sigma_s \epsilon_{si} + \psi_s \epsilon_{it} | a_{si} \leq \sigma_v v_i < a_{s+1,i}] = \sigma_s^2(1 - \rho_s^2 \sigma_v^2 \delta_{si}) + \psi_s^2 \] (5.14)

\( \delta_{si} \) is referred to as the truncation adjustment. Its expression is:

\[ \delta_{si} = 1 - \text{Var}[\sigma_v v_i | a_{si} \leq \sigma_v v_i < a_{s+1,i}] \] (5.15)

The sign of \( \delta_{si} \) determines whether observed wage dispersion understate or overstate potential wage dispersion. In fact, in case \( \delta_{si} > 0 \) (i.e.: \( \text{Var}[\sigma_v v_i | a_{si} \leq \sigma_v v_i < a_{s+1,i}] > 1 \)) observed wage dispersion is greater than potential wage dispersion while the opposite occurs when \( \delta_{si} < 0 \).

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

\[ (y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)\beta_s + (\zeta_{sit} - \bar{\zeta}_si) \text{ if } s_i = s, \] (5.16)

\( \bar{y}_i \), \( \bar{x}_i \) 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 (5.15).

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

\[ \bar{y}_i = \alpha_s + \bar{x}_i\beta_s + \gamma_s g_s(z_i \theta) + \bar{\omega}_i \] (5.17)

Where \( g(z_i \theta) = \sum_{j=1}^J b_j L_{ij}(\hat{z}_i \theta) \) in the Cossetti specification and \( g(z_i \theta) = E[\sigma_v v_i | a_{si} \leq \sigma_v v_i < a_{s+1,i}] \) in the alternative specification. The new error term \( \bar{\omega}_i \equiv \sigma_s \epsilon_{si} + \bar{\epsilon}_{it} - \gamma_s g(z_i \theta) \) has, by construction, expected value equal to zero. Consequently, a consistent estimator for the permanent component of wage

---

6See appendix for derivation.
7See appendix for derivation.
8See equation 5.13.
inequality is identified as:

$$\hat{\sigma}^2_s = \hat{\gamma}^2_s \delta_s + \overline{\text{Var}}[\omega_{si} | a_{si} \leq v_i < a_{s+1,i}] - \sum_i \frac{\hat{\psi}^2_{st}}{T}$$ (5.17)

The parameter $\hat{\gamma}^2_s$ is estimated as the coefficient for the correction terms distinguished by schooling level in an OLS regression; $\overline{\text{Var}}[\omega_{si} | a_{si} \leq v_i < a_{s+1,i}]$ is estimated as the mean squared of the error term with the between individual estimator $T \equiv (\sum_i T_i^{-1} / N)^{-1}$ and $\overline{\delta}_s$ is the sample average of the truncation adjustment. We now have all elements to identify wage uncertainty as defined in equation (5.6): $\hat{\tau}^2_{st} = \hat{\sigma}^2_s - \hat{\gamma}^2_s + \hat{\psi}^2_{st}$. To obtain a separate identification for the two components of $\gamma_s$ we need to substitute equation (5.17) in the expression for $\text{Var}[\sigma^2 e_{si} | a_{si} \leq v_i < a_{s+1,i}]$ and we obtain $\hat{\rho}^2_s = \frac{\hat{\gamma}^2_s}{\hat{\sigma}^2_s} \hat{\sigma}^2_s$.

As a result, all parameters of interest - $\sigma^2_s$, $\psi^2_{st}$, $\tau^2_{st}$ and $\rho_s$ - are identified.

## 5.4 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 males 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 subsequent waves.

Our dependent variables are two: schooling for the choice equation and earnings for the outcome equation. Schooling is measured as highest schooling

---

9 See appendix for derivation.
10 See appendix for derivation.
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 5.1, 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 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
Table 5.1: Summary statistics

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

<table>
<thead>
<tr>
<th>Time Variant Variables</th>
<th>1990</th>
<th>1993</th>
<th>1995</th>
<th>1997</th>
<th>1999</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual work experience</td>
<td>10.03</td>
<td>12.77</td>
<td>14.40</td>
<td>16.09</td>
<td>17.93</td>
</tr>
<tr>
<td></td>
<td>(3.58)</td>
<td>(4.05)</td>
<td>(4.35)</td>
<td>(4.68)</td>
<td>(5.04)</td>
</tr>
<tr>
<td>Log hourly earnings</td>
<td>2.18</td>
<td>2.16</td>
<td>2.28</td>
<td>2.26</td>
<td>2.21</td>
</tr>
<tr>
<td></td>
<td>(.98)</td>
<td>(1.06)</td>
<td>(1.06)</td>
<td>(1.14)</td>
<td>(1.25)</td>
</tr>
<tr>
<td>Unemployment rate (%)</td>
<td>5.81</td>
<td>8.70</td>
<td>6.01</td>
<td>5.59</td>
<td>4.12</td>
</tr>
<tr>
<td></td>
<td>(2.13)</td>
<td>(2.51)</td>
<td>(1.74)</td>
<td>(1.87)</td>
<td>(1.05)</td>
</tr>
</tbody>
</table>

Note: Standard deviations in parentheses. Unemployment rates calculated from CPS data.
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\(^{11}\) 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 5.1. 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.

### 5.5 Empirical results

As described in section 5.3, 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 \((\sigma_s^2)\); transitory component \((\psi_{st}^2)\); unobserved heterogeneity \((\nu_i)\) and risk \((\tau_{st}^2)\).

#### 5.5.1 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 5.3. 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\(^{12}\). The two methods differ on how steeply they penalize model complexity. AIC tends to penalize


\(^{12}\)See Cameron and Trivedi (2005) for a textbook discussion of the two criteria.
5.5. Empirical results

Table 5.2: Model comparison

<table>
<thead>
<tr>
<th>K</th>
<th>log likelihood of OP</th>
<th>LR-test of K-1 p-value</th>
<th>LR-test of K-1 p-value</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>OP</td>
<td>-20,573.45</td>
<td></td>
<td></td>
<td>41,202.91</td>
<td>41,424.66</td>
</tr>
<tr>
<td>3</td>
<td>-20,544.56</td>
<td>57.78</td>
<td>.000</td>
<td>57.78</td>
<td>.000</td>
</tr>
<tr>
<td>4</td>
<td>-20,504.04</td>
<td>138.82</td>
<td>.000</td>
<td>81.04</td>
<td>.000</td>
</tr>
<tr>
<td>5</td>
<td>-20,502.54</td>
<td>141.83</td>
<td>.000</td>
<td>3.01</td>
<td>.083</td>
</tr>
<tr>
<td>6</td>
<td>-20,493.57</td>
<td>159.78</td>
<td>.000</td>
<td>17.94</td>
<td>.000</td>
</tr>
<tr>
<td>7</td>
<td>-20,492.92</td>
<td>161.06</td>
<td>.000</td>
<td>1.29</td>
<td>.257</td>
</tr>
<tr>
<td>8</td>
<td>-20,489.37</td>
<td>168.17</td>
<td>.000</td>
<td>7.11</td>
<td>.008</td>
</tr>
</tbody>
</table>

The model complexity less than BIC, thus if parsimony is important BIC should be the preferred criteria. In table 5.2 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 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\(^{13}\).

Results of ordered Probit model and for the Gallant and Nychka (GN) procedure at 3rd and 8th degree polynomial are presented in table 5.3. It has to be noted that estimates of \( \theta_s \) 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 correlates 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

\(^{13}\)The process was stopped after the 200th iteration. Data and programs are available on request.
Table 5.3: First stage estimates for different values of K

<table>
<thead>
<tr>
<th>OP</th>
<th>GN(3)</th>
<th>GN(8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. unemp. rate</td>
<td>0.889***(.018)</td>
<td>0.975***(.012)</td>
</tr>
<tr>
<td>Mother attended college</td>
<td>0.329***(.028)</td>
<td>0.328***(.028)</td>
</tr>
<tr>
<td>Father attended college</td>
<td>0.534***(.027)</td>
<td>0.497***(.029)</td>
</tr>
<tr>
<td>Highest grade mother</td>
<td>0.017***(.004)</td>
<td>0.018***(.004)</td>
</tr>
<tr>
<td>Highest grade father</td>
<td>-0.002(0.004)</td>
<td>0.002(0.003)</td>
</tr>
<tr>
<td>Number of siblings</td>
<td>0.024***(.004)</td>
<td>0.019***(.004)</td>
</tr>
<tr>
<td>Family income bottom quartile</td>
<td>-0.015(0.029)</td>
<td>0.023(0.029)</td>
</tr>
<tr>
<td>Family income second quartile</td>
<td>-0.008(0.027)</td>
<td>0.013(0.026)</td>
</tr>
<tr>
<td>Family income third quartile</td>
<td>0.002(0.028)</td>
<td>0.001(0.024)</td>
</tr>
<tr>
<td>Family income top quartile</td>
<td>0.187***(.025)</td>
<td>0.182***(.024)</td>
</tr>
<tr>
<td>AFQT score (adjusted)</td>
<td>0.025***(.000)</td>
<td>0.023***(.001)</td>
</tr>
<tr>
<td>Black</td>
<td>0.515***(.023)</td>
<td>0.487***(.025)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.331***(.030)</td>
<td>0.319***(.029)</td>
</tr>
<tr>
<td>Cut-off point (κ1)</td>
<td>6.621***(.162)</td>
<td>6.621</td>
</tr>
<tr>
<td>Cut-off point (κ2)</td>
<td>8.246***(.165)</td>
<td>8.187***(.030)</td>
</tr>
<tr>
<td>Cut-off point (κ3)</td>
<td>9.192***(.166)</td>
<td>9.158***(.046)</td>
</tr>
</tbody>
</table>

Polynomial:

1  | -.591***(.064)  | -.306***(.053)  |
2  | .185***(.043)  | .169***(.064)  |
3  | .019***(.008)  | -.086(0.083)  |
4  | -1.21***(.035) | 0.026(0.016)  |
5  | .020***(0.005) | -.001(0.001)  |
6  | .001***(0.000) | .001***(0.000) |

Wald $\chi^2$  | 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.
5.5.2 Identification of selection correction term

Separate identification of risk \( (\tau^2_{st}) \) and unobserved heterogeneity \( (v_i) \) requires the estimation of four selection correction terms, one for each schooling level\(^{14}\). The Cosslett procedure produces as many ‘correction’ dummies as intervals for which the distribution is 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 wage equation over the entire sample of individuals irrespective of educational level along with three educational dummies. These are the corresponding coefficients \( \gamma_s \) in table 5.4, Cosslett column.

5.5.3 Wage equation

In table 5.4 we report estimates of equation (5.16) 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 \( g(z_i\theta) = \sum_{j=1}^{J} b_j I_{ij}(\hat{z}_i\theta) \) while what we refer to as Linear estimation in table 5.4 augments the wage equation with the inclusion of \( E[v_i|a_{si} \leq v_i < a_{s+1,i}] \) calculated from the Gallant and Nychka polynomial approximation in the first stage and assumes 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 Linear specification and significant only at a

\(^{14}\)See section (5.2.1) and the appendix.
Separating risk in education from heterogeneity: a semiparametric approach

Table 5.4: Wage equation

<table>
<thead>
<tr>
<th></th>
<th>GLS</th>
<th>Cossett</th>
<th>Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(I)</td>
<td>(II)</td>
<td>(III)</td>
</tr>
<tr>
<td>High school</td>
<td>.049***(.013)</td>
<td>-.071*(.015)</td>
<td>-.054(.097)</td>
</tr>
<tr>
<td>Some college</td>
<td>.091***(.014)</td>
<td>.166***(.027)</td>
<td>.243(.128)</td>
</tr>
<tr>
<td>4 yr. college or beyond</td>
<td>.351***(.017)</td>
<td>.502***(.034)</td>
<td>.453***(.079)</td>
</tr>
<tr>
<td>Experience</td>
<td>.132***(.005)</td>
<td>.133***(.005)</td>
<td>.132***(.005)</td>
</tr>
<tr>
<td>Experience^2</td>
<td>-.001***(.000)</td>
<td>-.001***(.000)</td>
<td>-.001***(.000)</td>
</tr>
<tr>
<td>AFQT score (adjusted)</td>
<td>.005***(.000)</td>
<td>.004***(.000)</td>
<td>.005***(.000)</td>
</tr>
<tr>
<td>Highest grade mother</td>
<td>-.002(.001)</td>
<td>-.003(.002)</td>
<td>-.002(.002)</td>
</tr>
<tr>
<td>Highest grade father</td>
<td>-.000(.001)</td>
<td>-.001(.002)</td>
<td>-.000(.001)</td>
</tr>
<tr>
<td>Number of siblings</td>
<td>.003(.002)</td>
<td>.003(.002)</td>
<td>.003(.002)</td>
</tr>
<tr>
<td>Family income bottom quartile</td>
<td>.011(.015)</td>
<td>.012(.017)</td>
<td>.010(.015)</td>
</tr>
<tr>
<td>Family income second quartile</td>
<td>-.041***(.014)</td>
<td>-.041***(.013)</td>
<td>-.041***(.014)</td>
</tr>
<tr>
<td>Family income third quartile</td>
<td>-.007(.013)</td>
<td>-.008(.013)</td>
<td>-.008(.013)</td>
</tr>
<tr>
<td>Family income top quartile</td>
<td>.065(.013)</td>
<td>.060***(.013)</td>
<td>.065***(.013)</td>
</tr>
<tr>
<td>Unemployment rate</td>
<td>-.001(.002)</td>
<td>-.001(.002)</td>
<td>-.001(.002)</td>
</tr>
<tr>
<td>Black</td>
<td>.047***(.016)</td>
<td>.035*(.017)</td>
<td>.049***(.016)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>.044(.017)</td>
<td>.039*(.017)</td>
<td>.045**(.017)</td>
</tr>
<tr>
<td>γ0</td>
<td>.860***(.186)</td>
<td>.089(.158)</td>
<td></td>
</tr>
<tr>
<td>γ1</td>
<td>.407***(.047)</td>
<td>.835*(.398)</td>
<td></td>
</tr>
<tr>
<td>γ2</td>
<td>.756***(.114)</td>
<td>-1.726(1.575)</td>
<td></td>
</tr>
<tr>
<td>γ3</td>
<td>.031(.080)</td>
<td>-.047*(.023)</td>
<td></td>
</tr>
<tr>
<td>R^2</td>
<td>.391</td>
<td>.394</td>
<td>.391</td>
</tr>
<tr>
<td>N</td>
<td>20,398</td>
<td>20,398</td>
<td>20,398</td>
</tr>
</tbody>
</table>

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.
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 Linear 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 Linear 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. 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 5.5.4 on these results.

5.5.4 Main results

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

In table 5.5 panel A we report measures for the observed wage dispersion. The observed wage dispersion is the sum of two factors \( \sigma^2_{st}(1 - \sigma^2_{st}\rho^2_{st}) \) and \( \psi^2_{st} \). 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 5.5 we can see that high school graduates are those showing
a lower dispersion both in the permanent (panel A) and transitory (panel B) component. On the other hand, college graduates have the highest variance. Observed wage dispersion monotonically increases after high school and the result 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 dispersion that each schooling level entails. In panel C we show our estimates of the permanent component corrected for selection and truncation biases. Is immediately evident the huge underestimation of wage dispersion if truncation and selection biases are not taken into account. Observed wage dispersion accounts for only 30% of potential wage dispersion in the case of high school and college drop outs and around 70% for high school graduates. These results abundantly exceed previous parametric estimations (Chen, 2008) for which underestimation varies between 2% and 30% depending on educational level.

On the contrary, observed wage dispersion overstates potential dispersion in the case of college graduates\(^\text{15}\). This result suggests 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 dispersion systematically understates potential wage dispersion. 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.

The hierarchical order of the four educational categories regarding poten-

\(^{15}\)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.
### Table 5.5: Parameters of interest

<table>
<thead>
<tr>
<th></th>
<th>Less than high school</th>
<th>High school</th>
<th>Some college</th>
<th>4 yr. college and beyond</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>I. Observed wage inequality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Permanent component</td>
<td>.308</td>
<td>.304</td>
<td>.352</td>
<td>.421</td>
</tr>
<tr>
<td>B. Transitory component ($\psi^2_{st}$)</td>
<td>.156</td>
<td>.131</td>
<td>.142</td>
<td>.207</td>
</tr>
<tr>
<td>Age 25-30</td>
<td>-.064</td>
<td>-.052</td>
<td>-.043</td>
<td>-.105</td>
</tr>
<tr>
<td>Age 31-36</td>
<td>-.060</td>
<td>-.054</td>
<td>-.047</td>
<td>-.110</td>
</tr>
<tr>
<td>Age 37-42</td>
<td>-.025</td>
<td>-.023</td>
<td>-.020</td>
<td>-.055</td>
</tr>
<tr>
<td>Observed inequality (A+B)</td>
<td>.464</td>
<td>.435</td>
<td>.494</td>
<td>.628</td>
</tr>
</tbody>
</table>

| **II. Potential wage inequality** |                       |             |              |                          |
| C. Permanent component ($\sigma^2_s$) | 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 ($\rho_s$) | .694                  | .494        | .588         | .045                     |
| F. Permanent component (C-CxE$^2$) | .521                  | .335        | .707         | .311                     |
| G. Transitory component (same as B) |                       |             |              |                          |
| Degree of wage uncertainty ($\tau^2_s$) | 365                   | 363         | .613         | .420                     |
| Unobserved heterogeneity ($v_i$) | 486                   | 108         | 374          | .001                     |

*Note: Estimates of transitory volatility are obtained by regressing $\psi^2_{st}$ on age dummies and categorical educational variables.*
tial wage dispersion 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 dispersion. College completion, on the other hand, diminishes variability by 64% to the lowest level among the four groups. It is worth noting 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 \( v_i \) 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 dispersion. 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 the schooling factor. The parameter \( \rho_s \) 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 \( \rho_s \) 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 dispersion for the three lowest educational categories. It explains a share varying between 42% and 18% of potential wage dispersion. 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 dispersion - 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 unobserved heterogeneity and not risk is the main contributor to wage dispersion. The contribution of unobserved heterogeneity for the identification of the potential wage dispersion 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 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 risk. 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 to would shed some light on the exact mechanism, but that exceeds the scope of the present Chapter.

Our results clearly show that potential and observed wage dispersion 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 dispersion for college graduates suggest that the use that individuals make of private information when selecting a university degree might not be at odds with assumptions of risk minimizing agents at least for this specific category.

5.6 Conclusion

In this chapter 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 risk that every education level entails.

Our results indicate that all decompositions of wage dispersion (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 as some previous research indicates (Moretti, 2000; Hartog, 2011) 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.
5.A Appendix: identification of $\rho_s$ and $\sigma_s$ in a linear model

To see how $\rho_s$ and $\sigma_s$ are identified consider the model as formalized by Chen (2008):

$$y_{sit} = \alpha_s + x_{it}\beta_s + \sigma_s e_{si} + \psi_{sit}\epsilon_{it}, \quad (5.18)$$

$$s_i^* = z_i\theta + \sigma_v v_i \quad (5.19)$$

where $s_i^* = s$ if $a_{si} \leq \sigma_v v_i \leq a_{s+1,i}$. This is the usual sample selection model with ordered censoring rules. We make the following assumption on a part of the error term:

$$\sigma_s e_{si} = \gamma_s \sigma_v v_i + \sigma_{\xi si}\bar{\zeta}_{si} \quad (5.20)$$

where $\gamma_s \equiv \sigma_s \rho_s$ 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 $v_i$ in the choice equation is correlated with $e_{si}$ but not with $\epsilon_{it}$. We also assume that $\bar{\zeta}_{si}$ is independent of $v_i$ and that the $\xi_{si}$ are uncorrelated across schooling levels. Additionally, $Var[\xi_{si}\bar{\zeta}_{si}] = \sigma_{\xi s}^2$. From these assumptions we obtain that:

$$E[\sigma_s e_{si} + \psi_{st}\epsilon_{it}| a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] = E[\sigma_s e_{si}| a_{si} \leq \sigma_v v_i \leq a_{s+1,i}]c = \gamma_s E[\sigma_v v_i| a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] \quad (5.21)$$

$$Var[\sigma_s e_{si}| a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] = \gamma_s^2 Var[\sigma_v v_i| a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] + \sigma_{\xi s}^2 \quad (5.22)$$

To re-establish the zero conditional mean of the error term in the outcome equation (5.16) in presence of self-selection, we need a correction term accounting for $E[\sigma_v v_i| a_{si} \leq \sigma_v v_i \leq a_{s+1,i}]$ and an estimate for $\gamma_s \equiv \sigma_s \rho_s$. In fact, the equation:

$$y_i = \alpha_s + x_i\beta_s + \gamma_s g_s(z_i\theta) + \omega_i \quad (5.23)$$

can be consistently estimated by OLS since $E[\omega_{is}|x_{it}, z_i] = 0$ by construction.
The function \( g(z_i\theta) = E[v_i|a_{si} \leq v_i \leq a_{s+1,i}] \) is unknown and is entered using the method of Cosslett (1983) or Gallant and Nychka (1987). We can obtain estimates for both \( E[v_i|a_{si} \leq v_i \leq a_{s+1,i}] \) and \( \text{Var}[v_i|a_{si} \leq v_i \leq a_{s+1,i}] \) by approximating the unknown density function by a Hermite series of \( K \) degrees polynomials.

\[
\hat{E}[\sigma_v v_i|a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] = \int_a^b \sigma_v v_i f_K(\sigma_v v_i) dv \quad \text{(5.24)}
\]

\[
\hat{\text{Var}}[\sigma_v v_i^2|a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] = \hat{E}[\sigma_v v_i^2|a_{si} \leq \sigma_v v_i \leq a_{s+1,i}^2] - \hat{E}[\sigma_v v_i|a_{si} \leq \sigma_v v_i \leq a_{s+1,i}]^2 \quad \text{(5.25)}
\]

According to our distributional assumptions the variance of the permanent component from an individual standpoint is given by:

\[
\text{Var}[\sigma_s e_{si}|x_it] = \text{Var}[\gamma_s \sigma_v v_i + \sigma_\xi \xi_{si}|x_it] = \sigma_s^2 \sigma_v^2 \rho_s^2 + \sigma_\xi^2 \quad \text{(5.30)}
\]

Remembering that \( \text{Var}[\sigma_s e_{si}|x_it] = \sigma_s^2 \)

we obtain:

\[
\sigma_\xi^2 = \sigma_s^2 (1 - \rho_s^2 \sigma_v^2) \quad \text{(5.26)}
\]

Substituting (5.26) into (5.22) and rearranging, we obtain the variance of observed wages corrected for truncation:

\[
\text{Var}[\sigma_s e_{si}|a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] = \sigma_s^2 (1 - \rho_s^2 \sigma_v^2 \delta_{si}) \quad \text{(5.27)}
\]

To see how \( \rho_s \) and \( \sigma_s \) are identified consider the model as formalized by Chen (2008):

\[
y_{sit} = a_s + x_{it} \beta + \sigma_s e_{si} + \psi_{st} \epsilon_{it}, \quad (5.28)
\]

\[
s_i^* = z_i \theta + \sigma_v v_i \quad (5.29)
\]

where \( s_i^* = s \) if \( a_{si} \leq \sigma_v v_i \leq a_{s+1,i} \). This is the usual sample selection model with ordered censoring rules. We make the following assumption on a part of the error term:

\[
\sigma_s e_{si} = \gamma_s \sigma_v v_i + \sigma_\xi \xi_{bsi} \quad (5.30)
\]

where \( \gamma_s \equiv \sigma_s \rho_s \) is the covariance coefficient between the error term in the
5.A. Appendix: identification of $\rho_s$ and $\sigma_s$ in a linear model

outcome equation and the error term in the choice equation. As Chen (2008) does, we assume that the error term $v_i$ in the choice equation is correlated with $\epsilon_{si}$, but not with $\epsilon_{it}$. We also assume that $\zeta_{si}$ is independent of $v_i$ and that the $\zeta_{si}$ are uncorrelated across schooling levels. Additionally, $\text{Var}[\sigma_s \zeta_{si}] = \sigma_{\zeta s}^2$. From these assumptions we obtain that:

$$E[\sigma_s e_{si} + \psi_s e_{it} | a_{si} \leq \sigma_s v_i \leq a_{s+1,i}] = E[\sigma_s e_{si} | a_{si} \leq \sigma_s v_i \leq a_{s+1,i}] = \gamma_s E[\sigma_s v_i | a_{si} \leq \sigma_s v_i \leq a_{s+1,i}] \quad (5.31)$$

$$\text{Var}[\sigma_s e_{si} | a_{si} \leq \sigma_s v_i \leq a_{s+1,i}] = \gamma_s^2 \text{Var}[\sigma_s v_i | a_{si} \leq \sigma_s v_i \leq a_{s+1,i}] + \sigma_{\zeta s}^2 \quad (5.32)$$

To re-establish the zero conditional mean of the error term in the outcome equation (5.16) in presence of self-selection, we need a correction term accounting for $E[\sigma_s v_i | a_{si} \leq \sigma_s v_i \leq a_{s+1,i}]$ and an estimate for $\gamma_s \equiv \sigma_s \rho_s$. In fact, the equation:

$$\bar{y}_i = \alpha_s + \bar{x}_i \beta_s + \gamma_s g_s(z_i, \theta) + \bar{\epsilon}_i \quad (5.33)$$

can be consistently estimated with OLS since $E[\omega_{is} | x_{it}, z_i] = 0$ by construction. $g(z_i, \theta) = E[v_i | a_{si} \leq v_i \leq a_{s+1,i}]$ is an unknown function entered using the method of Cosslett (1983) or Gallant and Nychka (1987). We can obtain estimates for both $E[\sigma_s v_i | a_{si} \leq v_i \leq a_{s+1,i}]$ and $\text{Var}[v_i | a_{si} \leq v_i \leq a_{s+1,i}]$ by approximating the unknown density function by a Hermite series of $K$ degrees polynomials.

$$\hat{E}[\sigma_s v_i | a_{si} \leq v_i \leq a_{s+1,i}] = \int_a^b \frac{\sigma_s v_i f_K(\sigma_s v_i) dv}{\text{Pr}[a_{si} \leq v_i \leq a_{s+1,i}]} \quad (5.34)$$

$$\hat{\text{Var}}[\sigma_s v_i | a_{si} \leq v_i \leq a_{s+1,i}] = \hat{E}[\sigma_s^2 v_i^2 | a_{si} \leq v_i \leq a_{s+1,i}] - \hat{E}[\sigma_s v_i | a_{si} \leq v_i \leq a_{s+1,i}]^2 \quad (5.35)$$

According to our distributional assumptions the variance of the permanent component from an individual standpoint is given by: $\text{Var}[\sigma_s e_{si} | x_{it}] = \text{Var}[\gamma_s \sigma_s v_i + \sigma_{\zeta s i} x_{it}] = \sigma_s^2 \rho_s^2 + \sigma_{\zeta s}^2$. Remembering that $\text{Var}[\sigma_s e_{si} | x_{it}] = \sigma_s^2$ we obtain:
\[ \sigma^2_{\xi_s} = \sigma^2_s (1 - \rho^2_s \sigma^2_v) \]  

(5.36)

Substituting (5.36) into (5.32) and rearranging, we obtain the variance of observed wages corrected for truncation:

\[ \text{Var} [\sigma_s e_{si} | a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] = \sigma^2_s (1 - \rho^2_s \sigma^2_v \delta_{si}) \]  

(5.37)

where we can estimate \( \delta_{si} \) by:

\[ \hat{\delta}_{si} = 1 - \text{Var} [\sigma_v v_i | a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] \].

The parameter \( \hat{\gamma}_s = \hat{\sigma}_s \hat{\rho}_s \) 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 (5.32), and the transitory shocks \( \psi_{st} \epsilon_{it} \). The expression for the variance of the complete error term is:

\[ \text{Var} [\sigma_s e_{si} + \psi_{st} \epsilon_{it} | a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] = \text{Var} [\sigma_s e_{si} | a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] + \psi^2_{st} \]  

\[ = \sigma^2_s (1 - \rho^2_s \sigma^2_v \delta_{si}) + \psi^2_{st} \]  

(5.38)

\[ \hat{\gamma}_s = \hat{\sigma}_s \hat{\rho}_s \text{Var} [\sigma_s e_{si} + \psi_{st} \epsilon_{it} | a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] = \text{Var} [\omega_{si} | a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] \]

can be consistently estimated as the mean squared errors of the residuals in the between individual effects model of expression (5.33). \( \psi^2_{st} \), as explained in Chen (2008), is identified from the fixed-effect model as the variance of residuals in equation (5.15). Substituting these elements into (5.38) and rearranging we identify the permanent component of wage inequality corrected for truncation as:

\[ \hat{\sigma}_s^2 = \text{Var} [\omega_{si} | a_{si} \leq \sigma_v v_i \leq a_{s+1,i}] + (\hat{\sigma}_s \hat{\rho}_s)^2 \hat{\delta}_s - \frac{\sum t \hat{\gamma}_{st}^2}{T} \]  

(5.39)

With \( T = (\Sigma_i T_i^{-1} / N)^{-1} \) and \( \hat{\delta}_s \) is the sample average of the truncation adjustment. Substituting (5.39) into (5.36) we obtain estimates for \( \sigma^2_{\xi_s} \) and \( \rho^2_s = \frac{\hat{\gamma}_s}{\sigma_s \sigma_v} \).

Note that we have identified \( \rho^2_s \) and \( \sigma^2_s \) 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. linearity in the equation of the error term \( (\sigma_s e_s = \gamma_s \sigma_v v_i + \sigma_{\xi_s} \xi_{si}) \);
2. the distribution of $\xi_{si}$ is independent of $\nu_i$. 