Unobserved Heterogeneity and Risk in Wage Variance: Does More Schooling Reduce Earnings Risk?

Jacopo Mazza[1] Hans van Ophem[§] Joop Hartog[§]

August 2013

Abstract

We apply a recently proposed method to disentangle unobserved heterogeneity from risk in returns to education to data for the USA, the UK and Germany. We find that in residual wage variation, uncertainty by far dominates unobserved heterogeneity. The relation between uncertainty and level of education is not monotonic and differs among countries.

JEL classifications: C01, C33, C34, J31

Keywords: wage inequality, wage uncertainty, unobserved heterogeneity, selectivity, education, replication

[1] Corresponding author. University of Manchester, Address: University of Manchester, Arthur Lewis Building, Oxford Rd., M13 9PL Manchester, UK. E-mail: jacopo.mazza@manchester.ac.uk.

[§] Universiteit van Amsterdam.

All data and computer programs are available on request. We gratefully acknowledge comments by two referees and by the editor, Ian Walker.

I. Introduction

Empirical information on the extent of risk in schooling choice is very important. With uncertain schooling benefits a fact of life, we need to know the extent of risk as an input for realistically modeling schooling choice as a choice under risk (Levhari and Weiss, 1974). Knowing the extent of risk is particularly relevant for policy issues. Education is often promoted as an insurance against earnings risk, but we have no solid evidence that it really is. While realised earnings variances for individuals with given levels of schooling are well documented, such data are not informative on risk as they also include unobserved heterogeneity that may govern potential students' choice[2].

A recent paper by Chen (2008) recognizes the potential bias in ex post earnings data and suggests a method to correct for it. Individuals are endowed with a factor ν that rules their choice of education: a single parameter reflecting their taste, abilities etc, known to the individual, unobserved by the outsider. Educational choice is modeled as an ordered probit on this taste factor, with interval boundaries depending on individuals' characteristics. Potential wage after completing an education has three components: rewards for individual characteristics, a permanent individual fixed effect and an annual transitory shock. Each component is education specific. The rewards for individual characteristics are known, the transitory shock is fully relegated to uncertainty. The fixed effect is partly known: only to the extent that it correlates with the schooling taste factor. The remaining part, the extent of imperfect correlation, is an element of the uncertainty faced by an individual.

To expand empirical knowledge on the magnitude of earnings risk associated with different levels of schooling, we apply Chen's method to data for the US, the UK and Germany[3]. Chen reported two main conclusions. First, for men in the US, risk does not increase with educational level as previous research on the topic suggested. Second, Chen finds evidence of pervasive underestimation of differences in potential wages by observed wage inequalities. Our results deviate from Chen's in several respects and we find no uniform relationship between uncertainty and level of

[2] Realised earnings variance has no robust relationship with length of education: depending on time and country, it may increase, decrease or stay constant. See Hartog, Van Ophem and Bajdechi (2004) and Hartog and Diaz Serrano (forthcoming)

[3] A different method to reach the same goal has been proposed by Cunha, Heckman and Navarro (2005). Also Belzil and Leonardi (2007) take endogeneity into account to establish how risk aversion is affecting educational choices.

education. However, a key conclusion stands firmly, both in Chen's results and in our own estimates: the contribution to wage inequality of unobserved heterogeneity is negligible relative to the contribution of uncertainty.

We intended our study as a replication of Chen's analysis, to generate internationally comparable information on the relationship between schooling and risk. However, we were unable to use the same instrument for schooling as Chen, because the data,on local tuition cost, were not available for the countries we selected. But we were able to use two identical instruments for each country (unemployment during schooling age and country GDP growth in the same period) and this makes our results comparable across countries. By applying two instruments we can apply standard tests for overidentifying restrictions and gain confidence on the validity of the exclusion restrictions selected. When we estimated Chen's model on her data with our instruments, we got different results. We cannot apply a pure replication of Chen's analysis, as observations on the instrument for schooling (local tuition cost) are only available for researchers residing in the US. As we were also unable to reconstruct Chen's dataset perfectly, we cannot exactly assess the effect of using different instruments. However, our interest centered on the relative magnitude of uncertainty and unobserved heterogeneity and in this respect all results point in the same direction: uncertainty by far dominates.

We proceed as follows. In section II we set forth Chen's model. Section III presents results for the US, Section IV for the UK and Section V for Germany. In Section VI we compare our results to the original Chen results and Section VII concludes.

II. Chen's model

A. The theoretical model

We present Chen's model in detail, for convenience of the reader, to define concepts and to point out how we dealt with obscurities in the original presentation. The model in Chen (2008) has been constructed to exploit the data in the NLSY79. Consider a panel dataset of N workers observed over T time periods indexed by subscripts i and t respectively. In the first period, worker i's schooling level is determined; it will not change over the following periods. The schooling level chosen by the individual will be indicated with *s*. Chen classifies the possible choices in the NLSY79 in four intervals: no high school diploma (s_i =0), high school graduate (s_i =1), some college (s_i =2) and four years college or beyond (s_i =3). y_{it} indicates the observed log wage in period *t* for person *i*. The worker's potential wage is obviously observed only in one educational level, therefore, the worker's *observed wage* is:

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

where *I{ }* is the indicator function taking value 1 if the subject belongs to that specific schooling category and 0 otherwise. The link between schooling level s_i and *potential wage* (y_{sit}) is given by the following regression model:

$$y_{sit} = \alpha_s + x_{it}\beta_s + \sigma_s e_{si} + \psi_{st}\varepsilon_{it} \text{ if } s_i = s. \tag{2}$$

α_s is the intercept for schooling level *s*, β_s the vector of coefficients of the observable characteristics x_{it}, e_{si} and ε_{it} are zero mean, unit variance random variables uncorrelated with each other[4]. The time invariant individual fixed effects are denoted by $\sigma_s e_{si}$. This term measures the unobserved earning potential at schooling level *s* which is allowed to be correlated with observable characteristics x_{it}. $\psi_{st}\varepsilon_{it}$ denotes the transitory shock, assumed to be uncorrelated with observables. The potential wage variation is $\sigma_s^2 + \psi_{st}^2$ for subjects' schooling choices *s* and covariates at time *t*. The permanent component σ_s^2 is created by variations in the individual specific effects which are supposed to vary across educations, but to be constant in time. The temporary shocks emerging from macroeconomic conditions or institutional changes are incorporated in ψ_{st}^2 which can vary with both time and schooling level. The variables of interest in this model are the variances of both components in potential wages.

The selection problem is formalized in a latent-index schooling assignment rule:

[44] We follow Chen in giving a general specification of the model. In the empirical implementation, the beta's are constrained to be equal across education, with education dummies added.

$$s_i = s \text{ if } \nu_i \in A_s \text{ for } s\text{=0,1,2 or 3,} \tag{3}$$

where the unobserved schooling factor ν_i summarizes the private information such as taste for education, ability and so on, which influences the subjects' educational choices. $A_s \equiv \{\nu_i : a_{si} \le \nu_i \le a_{s+1,i}\}$ is the group of individuals who chose educational level *s*. $a_{si} \equiv \kappa_s - z_i\theta$ is the minimal level of the unobserved schooling factor in A_s. The vector z_i contains time invariant covariates and an instrument for education whose coefficients are contained in θ. $\kappa_0 = -\infty$ and $\kappa_4 = \infty$. The structure of error terms is known to all agents and summarized by:

$$\begin{bmatrix} e_{si} \\ \varepsilon_{it} \\ \nu_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & \rho_s \\ & 1 & 0 \\ & & 1 \end{bmatrix}\right). \tag{4}$$

As assumed, the unobserved schooling factor is correlated with the individual fixed effects e_{si} , but not with the transitory shocks ψ_{st}. The correlation coefficient (ρ_s) can assume either positive or negative values. In case of positive value we have positive selection, the opposite in case of negative values.

The parameter ν_i clarifies why it is important to distinguish between wage variability and risk. In fact, the private information, by definition unobservable to the econometrician, can be used to predict the distribution of potential wages accessible to the subject for each schooling level. The expected value of potential wage at time *t* and schooling level *s*, from a personal point of view, is given by:

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

where $\gamma_s\nu_i$ represents the unobserved heterogeneity component at schooling level *s* and $\gamma_s \equiv \sigma_s\rho_s$. Equation (5) follows from the distributional assumptions in (4) and $E[e_{si} \mid s_i = s, x_{it}, \nu_i] = \rho_s\nu_i$.

Since the agent knows his own ability and tastes and uses the information to select the appropriate level of schooling, the degree of wage uncertainty can not exceed the degree of potential wage inequality. The *wage uncertainty* at schooling level *s* is measured by:

$$\tau_{st}^2 \equiv Var[\sigma_s e_{si} + \psi_{st}\varepsilon_{it} \mid s_i = s, x_{it}, \nu_i] = \sigma_s^2(1-\rho_s^2) + \psi_{st}^2 \le \sigma_s^2 + \psi_{st}^2. \qquad (6)$$

The second equality follows from the distributional assumptions described in (4): $Var(e_{si} \mid s_i = s, x_{it}, \upsilon_i) = 1 - \rho_s^2$. This equation makes explicit that potential wage variability ($\sigma_s^2 + \psi_{st}^2$) is the sum of two components: inequality created by wage uncertainty τ_{st} and inequality from unobserved heterogeneity $\gamma_s^2 = \sigma_s^2\rho_s^2$. In fact, if we rewrite equation (6) we obtain: $\sigma_s^2 + \psi_{st}^2 = \sigma_s^2\rho_s^2 + \tau_{st}^2$ and remembering $\gamma_s \equiv \sigma_s\rho_s$ we see that $\sigma_s^2 + \psi_{st}^2 = \gamma_s^2 + \tau_{st}^2$.

This equation also shows the three sources of uncertainty that each individual has to face: the earnings potential σ_s of the individual fixed effect (e_{si}); imperfect correlation between potential wages and private information (ρ_s); transitory shocks (ψ_{st}).

Equations (4) and (5) imply that potential wages are composed of observed heterogeneity ($x_{it}\beta_s$), unobserved heterogeneity ($\gamma_s\nu_i$) and an unforeseeable component (τ_{st}) plus an error term (u_{it}):

$$y_{sit} = \alpha_s + x_{it}\beta_s + \gamma_s\nu_i + \tau_{st}u_{it}, \qquad (7)$$

where u_{it} is a normalized random variable, uncorrelated with observable and unobservable characteristics. $\tau_{st}u_{it}$ is called the unforeseeable component of wage error, that is to say, risk. The first three terms of equation (7) are a direct consequence of the value of potential wage expected (by the individual) as explained in equation (5). The last term is describing uncertainty as modeled in equation (6) corrected by a normally distributed error term.

From this discussion it should be clear that the targets of identification are: wage uncertainty (τ_{st}) and the permanent and transitory component of potential wage inequality (σ_s^2 and ψ_{st}^2).

B. Model estimation and parameter identification

Equations (5) and (6) can not be used for regression analysis since v_i is unobserved; what is observed is the educational choice of an agent. The mean and variance of observed wages are determined by the following equations:

$$E[y_{it} \mid s_i = s; x_{it}, z_i] = E[y_{sit} \mid v_i \in A_s; x_{it}, z_i] = \alpha_s + x_{it}\beta_s + \sigma_s \rho_s \lambda_{si,} \tag{8}$$

$$\begin{gathered} Var[y_{it} \mid s_i = s; x_{it}, z_i] = Var[\sigma_s e_{si} + \psi_{st}\varepsilon_{it} \mid v_i \in A_s; x_{it}, z_i] = \\ \sigma_s^2(1-\rho_s^2\delta_{si}) + \psi_{st}^2 \le \sigma_s^2 + \psi_{st}^2 \end{gathered} \tag{9}$$

These equations specify the requirement for the construction of adjustments for truncation (δ) and selection (λ), explained below. Equation (8) shows that observed wages overstate or understate the mean potential wages depending on the sign of the correlation term ρ_s. To estimate selectivity adjustment λ_{si} , Chen starts by estimating an ordered probit on observed educational choice and calculates a generalisation of the inverse Mill's ratio:

$$\lambda_{si} \equiv E[v_i \mid v_i \in A_s] = [\phi(a_{si}) - \phi(a_{s+1,i}) / \Phi(a_{s+1,i}) - \Phi(a_{si})]. \tag{10}$$

Equation (9) shows how regardless of the sign of selection bias, observed wages understate the degree of potential wage inequality for each educational level. The degree of understatement is called by Chen truncation adjustment (δ_{si}), which also follows from the ordered probit:

$$\delta_{si} \equiv 1 - Var[v_i \mid v_i \in A_s] = \lambda_{si}^2 - [a_{si}\phi(a_{si}) - a_{s+1,i}\phi(a_{s+1,i})] / [\Phi(a_{s+1,i}) - \Phi(a_{si})], \tag{11}$$

where ϕ and Φ denote standard normal density and distribution function, respectively.

Transitory variance is estimated from a fixed-effect model based on equation (8) as a transformation of the variance of its residuals.[5] The fixed effect model is expressed as:

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)\beta_s + (\xi_{sit} - \bar{\xi}_{si}) \text{ if } s_i = s, \tag{12}$$

where $\bar{y}_i$, $\bar{x}_i$ and $\bar{\xi}_{si}$ denote the averages over time of the corresponding variables over the survey years. This identifies ψ_{st}^2 as the variance of the disturbances across individuals.

Next, a between-individuals model identifies the schooling intercept α_s and the unobserved heterogeneity coefficient γ_s:

$$\bar{y}_i = \alpha_s + \bar{x}_i\beta_s + \gamma_s\lambda_{si} + w_{si} \tag{13}$$

The error term $w_{si} \equiv \sigma_s u_{si} + \bar{\varepsilon}_{si} - \gamma_s\lambda_{si}$ satisfies by construction $E[w_{si} \mid s_i = s; \bar{x}_i, z_i] = 0$ and $Var[w_{si} \mid s_i = s; \bar{x}_i, z_i] = \sigma_s^2 - \gamma_s^2\delta_{si}^2 + \sum_t \psi_{st}^2 / T_i$. Thus, the consistent estimator for the permanent component of potential wage inequality is:

$$\hat{\sigma}_s^2 = \hat{V}ar(w_{si} \mid s_i = s; \bar{x}_i, z_i) + \hat{\gamma}_s^2\bar{\hat{\delta}} - \sum_t \hat{\psi}_{st}^2 / \bar{T} \tag{14}$$

The first term of this equation is the mean squared error of the between-individuals model; the second is the interaction between the consistent estimate of the unobserved heterogeneity term (γ_s) and the sample average of the truncation adjustment ($\bar{\hat{\delta}}$); the third equals the ratio between the transitory component of wage inequality ($\hat{\psi}_{st}^2$) and

[5] The complete process leading to the identification of the transitory component is discussed in Chen (2008) note 9 p. 278.

$\bar{T} \equiv (\sum_i T_i^{-1} / N)^{-1}$. Finally, the correlation coefficients are identified from $\rho_s = \gamma_s / \sigma_s$[6].

Let's recollect the concepts that have been introduced so far:

- *Observed wages* ($y_{it} \mid s_i = s; x_{it}, z_i$): wages observed in the data.
- *Potential wages* (y_{sit}): wages obtained by individual *i* if he had chosen schooling level *s*. Potential wage is the sum of observed heterogeneity ($x_{it}\beta_s$ - known to individuals and econometrician); unobserved heterogeneity ($\gamma_s \nu_i$ - known only to the individuals); unforeseeable component ($\tau_{st}\ u_{it}$ - unknown to everyone).
- *Observed wage inequality* ($Var[y_{it} \mid s_i = s; x_{it}, z_i]$): within educational category variation in wages. It is decomposed as the sum of transitory volatility (ψ_{st}^2 - estimates shown in panel B of the tables below) and the mean squared errors of the between individual-model (estimates shown in Panel A, the permanent component).
- *Potential wage inequality* ($\sigma_s^2 + \psi_{st}^2$): wage inequality that would have been experienced for each educational category if education was not chosen, but randomly assigned. It is the sum of the transitory volatility as defined above (ψ_{st}^2) and the permanent component (σ_s^2). The permanent component here accounts for selection and truncation biases (Panel C in the following tables).
- *Unobserved heterogeneity* ($\gamma_s \nu_i$): includes all the characteristics known to the individuals, but unknown to the econometrician that influence the schooling decisions and biases the OLS wage estimates.
- *Wage uncertainty* (τ_{st}^2): proper measure of risk in educational category *s*, equal to the sum of transitory component as defined above and a permanent component accounting for the unobserved schooling factor ν_i .

[6] For the sake of comparison, instead of using the Heckman selectivity correction, Chen has also estimated equation (12) with correction only for heteroscedasticity, by using GLS. We will also present GLS results.

III. Results for the US

In this section, we will present estimates for the US in the same specification that we use for the UK and for Germany, to obtain international comparability. In Section VI we will present a comparison with Chen's results for the US. Chen's instrument for schooling, college tuition fees in the county of residence, was not available to us for the US and is neither available for the UK and Germany.

For the US, we use the same data as Chen (2008): the NLSY 1979-2000. The original sample consist of 12,686 respondents aged 14-22 in 1979. We focus on men only in survey years 1991-2000, which correspond to calendar years 1990-1999, so that all the respondents should have terminated their studies. As sampling became bi-annual after 1994, we have data from 7 waves. Standard sampling weights provided with the NLSY79 are used to calculate all estimates. We exclude respondents that do not provide any information about parental education or the particular ability index that Chen utilizes. Exactly as in the original paper, 4,302 individuals remain in our sample[7]. Tables A1 and A2 in the Appendix report summary statistics (we will discuss the deviations from Chen in section VI)

Following Chen (2008), schooling is defined by the highest grade completed according to the 1990 survey when all respondents were at least 25 years old, and measured with four dummy variables: no high school (Years of Schooling YOS<12); high school (YOS=12); some college (12<YOS<16); college (YOS≥16). The ability index is the Armed Force Qualifying Test (AFQT). It was conducted in 1980 for all respondents of all ages and schooling levels; original scores are regressed on age dummies and quarter of birth and residuals are included in the choice and wage regressions. Quarters of birth capture schooling effects through compulsory schooling laws (Angrist and Krueger (1991). We use hourly pretax earnings, from wages, salary, commissions or tips from all jobs in the calendar year preceding the survey. The family income measure considers family income at age 17, or as close to 17 as possible: if family income at 17 is unavailable then the measure is taken at 16 or 18. For nearly half of the respondents the family income measure is unavailable. The work experience measure is constructed from the longitudinal work history in the

[7] 6,283 females, 824 individuals belonging to the supplemental military sample, 285 individuals with no information about ability index, 412 individuals with no information about mothers' and 580 about fathers' education were deleted.

NLSY79. Number of weeks worked in past calendar year is converted in number of full working years by dividing by 49.

We instrumented schooling choice with two instruments[8]. The first is the average national unemployment rate for the years spent in school after the mandatory schooling age. The second is the national GDP growth rate measured for the same period. The intuition behind these two instruments is that they will influence choice of education as they affect job finding probabilities and anticipated monetary benefits of education. A possible concern may be that unemployment rates during youth might correlate with present wages, as shown by Kahn (2010) and earlier by Beaudry and DiNardo (1991). We will therefore also include the national unemployment rate for the month of September of the year in which the individual enters in the labour market in the second stage of the selection model. The assumption is then that this controls for long run scarring effects and that, conditional on the labour market conditions for the year of labour market entry, past unemployment rates and GDP growth rates are uncorrelated with wages earned[9]. The evidence discussed by Kahn on the same sample of individuals exploited in our research, suggests that graduating in bad economic years causes wage losses between 1 and 20%. As long as the unemployment level at entry is able to capture the conditions of the labour market in that particular moment and the scarring effect, we can assume that current wages are related to unemployment rates experienced in the years preceding entry into the labour market only via this channel and thus consider our exclusion restriction as valid. We are forced to use the national unemployment rate and national GDP growth because residential information is not available to us. Data about unemployment rates are taken from the Current Population Survey (CPS)[10], while data on GDP growth are taken from the publicly available series published by the Bureau of Economic Analysis (BEA). By including two exclusion restrictions in the first stage we are able to test the validity of the combination of the two instruments via a standard Sargan-

---

[8] Instrumenting schooling does have an effect: OLS results are different. For sensitivity testing, we also used a third instrument, consumer sentiment; see section III, Table 5 and the discussion there. Consumer sentiment presumably also reflects expectations on the future state of the economy.

[9] Arkes (2010) and Hausman and Taylor (1981) are two of the few studies that use unemployment during schooling years as instrument for schooling. The results in Arkes are more directly comparable to ours since Arkes utilizes unemployment rate during teenage in a 2SLS estimation and finds the same negative influence on schooling achievement that we encounter. Hausman and Taylor, instead, use the average of three time-varying covariates, among which teenage unemployment, as instrument for education, but they still find a negative relationship.

[10] The URL address is: http://data.bls.gov:8080/PDQ/outside.jsp?survey=ln. (Accessed 15/06/2010)

Hansen test. The result of the test shows that we cannot reject the null hypothesis of valid overidentifying restrictions with p-value 0.236.

[TABLE 1 ABOUT HERE]

Schooling equations are presented in Table 1. Our instruments for education have a significant effect for every level of education, even after controlling for ability, family background, racial and geographical origin and age. The effect of ability is positive, the effect of family background runs through parental education rather than income. The marginal effect of the unemployment rate is negative for the lower schooling levels and positive for the higher levels: higher unemployment rates during schooling age stimulates participation in higher education. Average GDP growth has

[TABLE 2 ABOUT HERE]

the reverse effect. Using national aggregates clearly does not leave us with weak instruments. Additional information on the performance of our instruments is given in Table 2, for an OLS on schooling measured in years. The effect of unemployment during school years is not sensitive to including GDP growth or average consumer sentiment[11], the effect of GDP growth changes if unemployment is included[12]. The effect of the other variables is not sensitive to instrument choice, except in the case of family income. Remarkable are the positive effects of being Black and being Hispanic on attained levels of education; Chen finds the same.

[TABLE 3 ABOUT HERE]

In the wage equation (Table 3), we find significant positive effects of the inverse Mill's ratios for no high school and some college. In these two drop-out categories, those with a strong taste for dropping out earn more than the average drop-out. The effect of schooling is sensitive to specification. From GLS to IV returns to years of

[11] Consumer sentiment data are taken from the Survey of Consumer computed by the Survey Research Center at the University of Michigan. (http://www.sca.isr.umich.edu/documents. Accessed on 18/12/2012).

[12] Unemployment and growth correlate at 0.63, consumer sentiment correlates 0.44 and 0.45 with unemployment and growth. .

education drop by one third, in the categorical specification the gap between high school and college increases from 0.221 for OLS to 2.002 with Heckman selectivity correction.

Using a categorical specification for schooling instead of years does have some effect on other estimated coefficients, but in both specifications the effects are mostly modest anyway. The switch to IV cq Heckman also has some effect on these coefficients. Wages are not significantly related to unemployment in the year of labour market entry, and in that sense do not support existence of a long time scarring effect.

In Table 4 we present our main results[13]. Observed wage inequality is decomposed in its two components. The first is the permanent component, identified by the mean squared residuals in the between-individuals model[14] (equation 14). The second is the transitory component ψ_{st}^2 identified by exploiting the mean-squared errors of the fixed-effects model as described in note 9 p. 278 in Chen (2008). Transitory volatility ψ_{st}^2 is consistently estimated by:

$$\hat{\psi}_{st}^2 = \hat{V}_{st} N^{-1} \sum_i T_i / (T_i - 2) - N^{-1} \sum_i \hat{\Omega}_{si} / (T_i (T_i - 2)) , \qquad (15)$$

where $\hat{V}_{st}$ is the mean squared errors of the fixed-effects model and $\hat{\Omega}_s = \sum \hat{V}_{st} / (1 - 1 / T_i)$.

Observed inequality fluctuates with level of education and is higher for the incomplete educations than for graduations. This pattern mostly reflects the permanent component, as the transitory component monotonically declines. Permanent inequality is substantially larger than transitory inequality. Transitory variance decomposed by age class (not reproduced here) increases above age 36 and age 45 is about double that at age 25.

Potential wage inequality is the sum of the permanent component after taking out the effects of selection and truncation, and the transitory component ($\sigma_s^2 + \psi_{st}^2$). Chen has shown analytically how observed wage inequality systematically understates potential inequality if education were randomly assigned (Chen, 2008, p 278). Chen

[13] In this section we focus on our own results; detailed comparison with Chen's results is discussed in section VI.

[14] Chen affirms on page 283 that the permanent component is defined as the variance in the individual fixed effect model. This would conflict with the definition given on page 278 and with equation 12. For this reason, we will adhere to the definition provided on page 278 and use the mean squared errors of the between-individuals model.

corrects this by incorporating a truncation adjustment term and a heterogeneity term (equations (10) and (11)). Observed inequalities are indeed smaller than potential inequalities. Potential inequality monotonically declines with level of education, and again, the permanent component strongly dominates over the transitory component.

Correlations between the schooling taste factor and the permanent component are positive and substantial for the educations below college, indicating positive selection. At college level, the correlation is not significant. A positive correlation between the schooling factor and the fixed effect in earnings is perhaps what is intuitively most plausible. Those with a taste for education and/or a relatively high unobserved ability might be thought of benefitting more than others from education. It's an intuition fed by the presumption of earnings maximization. But utility maximisation may lead individuals to intellectual endeavours that are not well paid (such as, for example, cultural studies). In the simple classification of education in four levels, much heterogeneity is hidden and it is hard to formulate a convincing anticipation of positive or negative selection. The correlations are not inconsistent with the effects of Heckman selectivity terms but not identical: the selectivity term for High School is not significant (as the sign of the Heckman selectivity term is determined by the correlation coefficient between unobserved heterogeneity in school choice and wages, the regression coefficient in the Heckman wage equation should have the same sign as the correlation coefficient in the Chen model). The permanent component in individuals' wage uncertainty jumps up after high school.

[TABLE 4 ABOUT HERE]

The last two rows bring out the decomposition we are mostly interested in: risk versus unobserved heterogeneity. Unobserved heterogeneity dominates for high school and below, risk dominates beyond high school, and strongly so for college. Risk is not monotonically related to level of schooling, but it is higher beyond high school than for high school and less than high school.

Anticipating the comparison, in section VI, of Chen's original results to ours, and acknowledging the common sensitivity of results to the selection of instruments, we have used five different instrumentations for schooling: unemployment, GDP growth, consumer sentiment (indicating the mood for future economic development), unemployment and GDP growth jointly (as applied above) and all three simultaneously. As the correlation between tastes for education and the permanent component in wages is pivotal, we only reproduce the correlation coefficients (Table 5). Clearly, in this choice set, the choice is immaterial: the outcome is not sensitive to the choice of instruments, the correlation coefficient is very robust. This is not due to lack of differentiation among instruments as the correlation is imperfect (see footnote 14). The results in Table 4 are based on column 4 of Table 5, GDP and unemployment as instruments.

[TABLE 5 ABOUT HERE]

IV. Estimation on British data[15]

The British Household Panel Survey (BHPS) is an annually collected survey, begun in 1991 with 5,500 households, containing approximately 10,000 individuals. Every year individuals of the original sample are interviewed; if a member of the original sample splits-off from his original family, he is followed in the new household and all adult members of the new family are interviewed as well. Also new members joining a selected family are added to the sample and children are interviewed once they reach age 16. Further extensions to Welsh, Scottish and Northern Irish families increased the sample size to 10,000 households across the UK. We could access the surveys until 2008; therefore 18 waves are included in our analysis.

From the initial sample of 42,567 individuals, we drop 21,593 females, 1,242 and 5,924 individuals that have no information on parents education and income respectively, 2,242 observations lacking information on ethnic background, 772 lacking information on educational attainments and 1,635 observations outside working age interval 18-65. Furthermore, since Eurostat records unemployment information starting only from 1983 we had to drop all 5,361 individuals who left school before then. Our final unbalanced sample counts 4,476 individuals observed on average for 6 periods. The BHPS does not provide any measure comparable to the AFQT score collected in the NLSY nor any other proxy plausibly related to ability. An additional difference between BHPS and NLSY is how earnings are recorded: monthly instead of hourly earnings. To make results comparable, we have rescaled the BHPS data to hourly pay, by dividing the monthly wage by 176 (assuming an 8 hours

[15] We are grateful to Martyn Andrews and Matt Dickson for clarifications on the UK data.

working day and 22 working days per month). Monthly pay, self-reported, is defined as the usual monthly wage or salary payment before tax and other deductions in current main job. Data on unemployment and GDP growth are taken from the publicly available Eurostat database[16]

A. British educational system

Compulsory education in the UK lasts for 11 years, from age five until age sixteen. It is divided in four key stages. The first two years (age five to seven) compose the first stage; the following four years (from seven to ten) the second and along with the first stage it constitutes primary education. The third (3 years from eleven to thirteen) and fourth (2 years from fourteen to fifteen) key stages form, altogether, the secondary education. At the end of secondary education the GSCE (General Certificate for Secondary Education) is awarded in specific subjects. Often, a good score in the GSCE is a requirement for access to further education.

A-levels (Advanced Level of General Education) are the first degree of non-compulsory education and are a prerequisite for access to academic courses in UK institutions. They take two years for completion, from age 16 to age 17.

University education is divided in two cycles. The first awards a Bachelor degree and generally lasts three years, while the second leads to a Master degree and takes in most cases one year. Along with the standard tertiary education, a number of other professional higher educations such as the Post Graduate Certificate in Education (PGCE) or the Bachelor of Education (BEd) or nursing degrees exist.

We use six categories of education: no qualification (no degree whatsoever), some qualification, GCSE (GCSE degree obtained) which is comparable to high school education, A-level qualification (A-levels degree obtained), First Degree education, comparable to college education and Further Degree which comprehends all postgraduate education.

B. Wage variance in British data

In Appendix Tables B1 and B2 we present sample characteristics. The modest increases (and even declines) in average experience and age over time, that may come as a surprise, can be explained from the addition of new families as young adults

[16] http://epp.eurostat.ec.europa.eu/ accessed on 23/04/2013.

leave their family home. As the schooling equation in Table 6 shows, the effect of family background (parental education) is positive. Just as in the US, ethnic minorities attain higher levels of schooling. In the UK, unemployment has a strong positive effect on schooling, GDP growth a negative effect, also just as in the US (the instruments correlate at 0.41; Sargan Hansen's J test for the instruments has value chi2(1) = .915633 (p = 0.3386) so the null hypothesis is not rejected).

In the wage equations (Table 7), the Heckman specification does not give commonly anticipated results: the selectivity terms are mostly insignificant, the returns to schooling are not monotonically increasing and the other variables are also mostly

[TABLE 6 ABOUT HERE]

insignificant. In the IV specification for years of education, the rate of return to schooling is almost 8 percent, substantially higher than in the GLS estimation. There are some unexpected results (negative effect of father’s education, urban environment), but they disappear in the categorical specification. Unemployment at labour market entry has a significant positive effect on wages, contrary to the scarring hypothesis.

[TABLE 7 ABOUT HERE]

The results of prime interest are given in Table 8. Observed wage inequality has two peaks in relation to level of education, peaking first at GCSE and then at the highest degree; it is entirely dominated by permanent inequality, the contribution of the transitory component is negligible. Potential inequality jumps up between A levels and First degree, and is more or less stable before and after the jump. The correlation coefficient varies strongly with level of education. If significant, it is negative, indicating lower potential earnings for individuals with a stronger taste for education. The Heckman selectivity correction terms, while mostly insignificant, are also predominantly negative. As for the US, results are not inconsistent, as we do not find opposing signs which are both significant; but there is only one case of significant effect with the same sign (A levels). Uncertainty dominates unobserved heterogeneity, and strongly so in all but one case (First Degree). There is only modest variation in uncertainty, with peaks at GCSE and Higher Degree. As noted above, Chen has shown analytically that observed inequality should always be smaller than potential inequality. This is mildly violated for GCSE. However, the restriction only applies if normality in the error terms is imposed (see Mazza and Van Ophem, 2013).

[TABLE 8 ABOUT HERE]

V. Estimation on German data

For Germany we used data on men in the Socio-Economic Panel (SOEP), 1999 - 2008. The SOEP is a longitudinal study begun in 1984, conducted yearly by the German Institute for Economic Research (DIW, Berlin) and containing information on 11,000 households and 61,668 individuals. We exploit information about educational attainments of the respondents and their parents, income and demographics such as age, sex, region of residence and family composition (number of siblings). As for the BHPS, no proxy for ability is available. Additionally, we will not be able to control for parental family income since such information is not collected in the survey.

Applying the same sample selection procedure used for the previous two samples we drop 26,524 individuals with no income information; 406 with no educational attainment information, 16,190 females, 3,989 individuals outside the working age range, 539 for whom work experience cannot be reconstructed, 6,398 and 145 for lack of information on number of siblings and parental education respectively. We also have to drop an additional 3,356 individuals who were born and raised in former DDR since no information on youth unemployment levels is available for that country. We end up with an unbalanced sample of 4,121 individuals observed on average for 7 years.

As for the estimation on English data, unemployment and GDP figures come from the Eurostat database[17]. Sample characteristics are given in Appendix Tables C1 and C2. The German sample on average is somewhat older and more advanced in their career than the American and British samples.

A. German educational system

[17] http://epp.eurostat.ec.europa.eu/ accessed on 23/04/2013.

Education in Germany is compulsory from six until sixteen years of age. Compulsory education covers two separate cycles: primary education (*Grundschule*), common to everyone from age six to age ten[18], followed by a differentiated lower secondary education, from ten to sixteen.

The main secondary schools are *Hauptschule*, *Realschule*, *Gymnasium* and *Gesamtschule*. *Hauptschule* offers a lower secondary education and is immediately accessible after primary school irrespective of grades obtained in the preceding education cycle. It is addressed to students who are likely to terminate their education after the secondary cycle. *Realschule* and *Gymnasium* impose some minimum grades requirements for admission. *Realschule* offers a more applied type of education compared to *Gymnasium*. Its graduates typically pursue additional vocational education. *Gymnasium* places a strong emphasis on academic learning. Grade based admission is more selective than in *Realschule* and students graduating from it typically pursue university education. The nature of *Gesamtschule* varies somewhat between regions. It is generally comparable with English comprehensive school. Admission is unconditional on past performance as in the *Hauptschule* case, but for some students it provides a preparation for further academic careers in *Gymnasium* while others might end up graduating in a *Hauptschule*, depending on how well they perform. As a first approximation we could say that the best students usually attend *Gymnasium* while the weakest go to *Hauptschule* with *Realschule* and *Gesamtschule* being the middle ground.[19] Transition from primary to secondary education and choice among the different paths differs according to the specific regional laws. The main factors are performance at primary school and consultation with parents. The final decision is taken either by parents, by the school or by the school supervisory authority. Certain schools impose some pre-requirements such as a minimum ability level. In our sample we have grouped all students belonging to these schools in the intermediate school category.

After compulsory education students can access upper secondary education. The type of school entered depends on the qualifications and entitlements obtained at the end of lower secondary education. Upper secondary education is organized in two streams: general education (*Gymnasiale Oberstufe*) – high school category in our

---

[18] The Lander of Berlin and Brandenburg are an exception. In those two regions primary school lasts six years from six to twelve years of age.

[19] Some Lander offer additional types of schools, but the ones listed comprehend the majority of students and are common to the entire country.

classification – and vocational education which comprehends the *Berufsfachschule*, the *Fachoberschule*, the *Berufliches* Gymnasium or *Fachgymnasium*, the *Berufsoberschule* and other types of schools that exist only in certain Lander – taken as vocational high school category in our analysis.

Tertiary education is also divided in general education and vocational education. Next to the traditional universities there are the *Technische Universitäten* specialized in engineering and natural sciences. These institutions can award a doctoral degree (*Doktorgrad*). Vocational tertiary education is carried on in the colleges of art and music and in the universities of applied sciences (*Fachhochschulen*).

To summarize: we distinguish No high school (no diploma at all), Intermediate school (the four lower secondary schools), Vocational high school (four upper secondary vocational school types), High school (*Gymnasiale Oberstufe*), Higher vocational education (*Fachhochschulen*), Technical university and University.

B. Wage variance in German data

In the first-stage regression (see Table 9), we find the common result of parental education stimulating offspring's education. In Germany, ethnic minorities attain lower levels of education. This is in contrast to results for the US and the UK, and probably relates to the composition of the immigrant population (in Germany, the immigrant population is dominated by Mediterranean families who entered as unskilled workers). Just as in the UK, unemployment stimulates schooling, GDP

[TABLE 9 ABOUT HERE]

growth reduces schooling attainment (the two instruments correlate at -0.56, in contrast to the positive correlation in the American and British data; a negative correlation is indeed more plausible, although leads and lags related to labour hoarding may upset this; Sargan Hansen's J test for the instruments has value chi2(1) = 4.33267 (p = 0.0374), which means the null hypothesis is rejected at 5% level but not at 1%). Number of siblings reduces schooling length, age increases it.

In the GLS wage equation (Table 10), years of schooling have a rate of return of some 11%, and this is not affected by instrumenting. Experience has the usual concave effect on wages, parental education has no effect, there is no evidence of scarring when entering the labour market under adverse conditions. Instrumenting has not much effect on estimation results. Results are neither dramatically different in the categorical specification; significance levels may be affected, but most effects remain quite small. Quite remarkably, with Heckman selectivity correction all schooling effects are insignificant and so are the selectivity corrections.

[TABLE 10 ABOUT HERE]

The prime results are presented in Table 11. Observed inequality is highest for non-qualified workers, varies within secondary and tertiary education, and reaches approximately similar magnitudes at these levels. Permanent inequality strongly surpasses transitory inequality. Observed inequality is larger than potential inequality, except for high school drop-outs. Correlation coefficients vary strongly among educations and exhibit both positive and negative signs. The regression coefficient for Heckman selectivity correction and the estimated correlation coefficient in the Chen model are qualitatively identical in two cases (positive and significant); in all the other cases, the Heckman selectivity term is not significant, whereas the correlation coefficient is significant in two cases. The core conclusion is quite clear though: uncertainty is far more important than unobserved heterogeneity, and risk does not vary monotonically with level of education.

[TABLE 11 ABOUT HERE]

VI. Comparison with Chen's results

Pure replication of Chen's results was an essential element in our original research plan. Economists often praise the virtue of replication, but rarely attempt it. We strongly believe that putting empirical results under careful scrutiny is an important if not essential task per se. For two reasons, however, exact replication proved impossible: US restrictions prohibit access to exact locational information from abroad and Chen's data file did not survive her move from the US to Taiwan. The first reason made it impossible for us to use the same instrument for schooling as Chen did, the second reason made it impossible to create precisely the same dataset as Chen used[20].

Pure replication is commonly marred by small and big problems. Small problems are obscurities in definitions and procedures[21]. Chen does not specify her sampling weights, but applying the standard sampling weight provided with the NLSY79, as we did, seems fairly obvious. More problematically, strictly applying Chen's sample selection rules did not bring us to the reported sample size. This was ultimately solved when Chen sent us her Stata do file for sample selection. As shown in Appendix Table A1, that leads us to the same sample size, but leaves small differences in sample characteristics. With the original data file lost, Chen could not help us find the explanation.

A truly serious problem is that we are unable to apply exactly the same model specifications as Chen: we have no access to observations for her instrument for education. Chen uses average tuition fees in the county of residence for a public four-year college in the year when the respondent was 17. The estimate of fees is based on the geocode, which gives access to detailed information on the residence of respondents (it also allows to control for the population density in the county of residence). We do not have access to the geocode data since the use is limited to researchers at American institutions. For the sake of international comparability and data availability we used different instruments, youth unemployment and GDP growth at schooling ages. This rules out reliable measurement of the effect of different instruments on Chen's results.

[20] Full details of our initial attempts have been spelled out in a discussion paper (Mazza, Van Ophem and Hartog, 2011). All our doubts and queries raised in the replication have been submitted to Chen. Some but not all queries have been answered. Chen indicated that some questions could not be answered as her data files got lost when she switched jobs.

[21] Footnote 14 is a case in point. There are two more. Chen (2008) p. 279 claims to use annual earnings. This claim does not correspond with the earnings measure presented in her Table 1 which is the logarithm of hourly earnings in 1992 dollars. We tried both outcome variables and chose the latter. In fact, if annual earnings were used as dependent variable in the between-individuals model, the magnitude of the residuals as presented in her table 4 would seem to be too small. This is not the case if hourly earnings is the explained variable. Also, as is evident from her descriptive statistics, it seems that she included four dummies to characterize the entire quartile distribution of family income at age 17. Since it is evident that four dummies plus the constant would create a dummy trap, we suspect that she created a dummy variable for non-response to the family income question. This is the way we proceed.

Chen's original results have been included in Table 4, next to our own results[22]. Our estimates of observed in equality are smaller, mostly because of smaller estimated transitory inequality. Our estimates of potential inequality are larger than Chen's for high school and below, higher than Chen's for some college and college. Our estimated potential inequality monotonically rises with level of education, which is not the case for Chen's estimates. The estimated correlation coefficients are very different: negative but insignificant in Chen's case (in line with insignificance of all Heckman selection terms), significant, positive and substantial for all but college education in our case. As noted earlier, positive selection is probably more in line with most economists' intuition, but negative selection cannot be ruled out. As a consequence of the differences in correlations, the balance of heterogeneity and risk also differs between our and Chen's results. In Chen's results, risk strongly dominates heterogeneity, in our case heterogeneity dominates at the lower levels of education. In neither case is risk monotonically related to level of education, although one might say that in Chen's case it ís more or less constant for high school and beyond. In view of these differences it is highly regrettable that an exact replication exercise is not feasible. The difference in sample characteristics is probably too small for substantial impact. We can also rule out that discrepancies are the result of a misunderstanding of the estimation procedure. We have scrupulously tested our estimation routine through a Monte-Carlo simulation (results available on request) and our routine was able to retrieve all the parameters of the simulated dataset with good precision.

Logically remaining possible explanations are errors in Chen's estimation procedure and sensitivity to differences in instruments. Chen's instrument for education is tuition cost in the county of residence. One might speculate that low family income, low ability students are most sensitive to such cost measure. Youth unemployment and GDP growth may be more relevant for the average student, but we cannot rule out that the marginal student (low family income, low ability) is again relatively the most sensitive type of student. Thus, we see no straightforward explanation of the differences in estimation results in different specific sensitivities of the instruments. We only feel confident to draw the trivial conclusion that the results are not robust.

[22] In Chen (2008), the entry for Less than High school, E+B, is an obvious typing error, as Chen acknowledges on her website. We have inserted the correct number.

We have also estimated the model for American women[23]. The women have about the same average schooling level as men, but their AFQT score is lower, at 50.48 (compared to 57.88 for men). Inequality measures tend to be (somewhat) smaller for women than for men. As Table 12 shows, the estimated correlation coefficients are quite similar to those for men: significant, positive and substantial, except for college. Risk dominates over heterogeneity, except for high school graduates; the relationship of risk with level of education is not monotonic.

[TABLE 12 ABOUT HERE]

[23]We have chosen to ignore the problem of selective participation, no doubt more relevant for women than for men. If we estimate on a sample restricted to full-time women, correlation coefficients greater than 1 are found.

[FIGURE 1 ABOUT HERE]

VII. Conclusion

Variation in observed wages at given levels of education may be a misleading indication of the risk associated with investing in education. Conceptually, part of the variation will result from heterogeneity among students and may be foreseen by the potential student when deciding on schooling. In this paper, we have applied an econometric model developed by Chen (2008) to datasets for the US, the UK and Germany; we were able to use the same instruments for schooling for each country. We used the results to study two questions:

- Is wage variability by education dominated by risk or by unobserved heterogeneity?
- Is risk monotonically related to level of education?

We have graphed our results in Figure 1. The answer to the first question is almost unequivocal. In the large majority of cases, risk dominates over unobserved heterogeneity. In most of these cases, the dominance is overwhelming. In only 3 out of 21 cases, heterogeneity dominates; these 3 cases all relate to the US. If we would add Chen's results for US men, with a different instrument for education, the case would even be strengthened, as for each education in that study risk strongly dominates heterogeneity.

The answer to the second question is unequivocally negative. In none of the datasets is there a monotonic relationship. Increasing levels of education are not uniformly associated with lower (or higher) levels of risk.

The results with different instruments for schooling on the same dataset for US men are not identical. Unfortunately, we cannot exactly replicate the original study for the US, so we cannot unconditionally conclude that only different instruments are responsible for different estimation results.

Our conclusion is at variance with Cunha and Heckman (2007, p.892) who conclude from their survey of several contributions by Heckman and co-authors: "For a variety of market environments and assumptions about preferences, a robust empirical regularity is that over 50% of the *ex post* variance in the returns to schooling are foreseeable at the time students make their college choices". Chen found that for college drop-outs, unobserved heterogeneity is negligible while for college graduates, the ratio of uncertainty to heterogeneity is 2:1. The Heckman et al. model is econometrically involved and it is hard to pinpoint what exactly causes the difference from the Chen model. Our results with the Chen model are more in line with the picture that emerges from research on data from directly asking individuals about their wage expectations. The available literature from several countries suggests that (potential) students' expectations on their future earnings distributions are indeed simply anchored to observed wages for graduates already in the labour market and that deviations between their own expectations and the observed means are not systematically related to their own (perceived) qualities (Dominitz and Manski, 1996;Wolter and Weber, 2004; Schweri, Hartog and Wolter, 2011; Nicholson and Souleles, 2001; Betts, 1996; Brunello, Lucifora and Winter-Ebmer, 2004; Webbink and Hartog, 2004; Hartog, 2011). Survey results clearly suggest that risk dominates over heterogeneity.

Both the difference between Heckman et al. and the Chen model, and our experiences with replicating the Chen model indicate that the relationship between schooling and risk has not yet been reliably exposed. We have tried many different specifications, sometimes deliberately, sometimes erroneously, and we found the results to be quite sensitive. There is still a lot of work to be done, both in pure replication and application of the same model to different settings, as well as, no doubt, in new modeling.

Statistical Appendix A. Data and results for the US

[TABLE A1 ABOUT HERE]

[TABLE A2 ABOUT HERE]

Statistical Appendix B. Data and results for the UK

[TABLE B1 ABOUT HERE]

[TABLE B2 ABOUT HERE]

Statistical Appendix C. Data and results for Germany

[TABLE C1 ABOUT HERE]

[TABLE C2 ABOUT HERE]

References

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

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

Beaudry, P. and J. DiNardo (1991), "The Effect of Implicit Contracts on the Movement of Wages over the Business Cycle: Evidence from Microdata", *Journal of Political Economy*, 99 (4), 665-668.

Belzil C. and Leonardi M. (2007), Risk Aversion and Schooling Decision, IZA Discussion Paper 2994, August.

Betts (1996), "What Do Students Know About Wages?", *Journal of Human Resources*, 31 (1), 27-56.

Brunello, G., C. Lucifora and R. Winter-Ebmer (2004), "The Wage Expectations of European business and Economics Students", *Journal of Human Resources*, 39 (4), 1116-1142.

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

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

Cunha, F. and J. Heckman (2007), "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling", *Labour Economics, Special Issue: Education and Risk*, 14 (6), 870-893.

Dominitz, J. and C. Manski (1996), "Eliciting Student Expectations of the Return to Schooling", *Journal of Human Resources,* 31, 1-26.

Hartog, J. (2011), "A Risk Augmented Mincer earnings equation? Taking stock", *Research in Labor Economics,* .33, 129-173

Hartog, J., H. van Ophem and S. Bajdechi (2004), How Risky Is Investment in Human Capital?, CESifo Working Paper 1261.

Hartog, J. and L. Diaz Serrano (forthcoming), Schooling as a risky investment, NOW publishers

Hausman J.A. and Taylor W.E. (1981), "Panel Data and Unobservable Individual Effects", *Econometrica*, 49(6): 1377-1398.

Kahn L. B. (2010), “The Long-Term Labor Market Consequences of Graduating From College in a Bad Economy”, *Labour Economics*, 17(2): 303-316.

Levhari D. and Weiss Y. (1974), “The Effect of Risk on the Investment in Human Capital”, *American Economic Review*, 64(6): 950-963.

Mazza, J. and H. van Ophem (2013), Separating risk in education from heterogeneity: a non-parametric approach, Working Paper Manchester/Amsterdam

Mazza, J., H. van Ophem and J. Hartog (2011), Unobserved heterogeneity and risk in wage variance: does schooling provide earnings insurance?, Amsterdam: Tinbergen Institute TI 2011-045/3 and Bonn: IZA DP 5531.

Nicholson, S. and N. Souleles (2001), Physician income expectations and specialty choice, Cambridge: NBER Working Paper 8536.

Schweri, J., J. Hartog and S. Wolter (2011), Do Students Expect Compensation for Wage Risk? *Economics of Education Review*, 30 (2), 215-227.

Webbink, D. and J. Hartog (2004), “Can Students Predict Their Starting Salary? Yes!”, *Economics of Education Review*, 23 (2), 103-113.

Wolter, S. and B. Weber (2004), Returns to education: are students’ expectations rational? Working Paper Swiss Coordination Center for Research in Education, Aarau, Switzerland.

Tables and figures

Table 1. Schooling: Ordered Probit estimation - NLSY

	Coefficients	**Marginal effect at means**			
		Less than high school	High school	Some College	College
Highest grade mother	.050***	-.002***	-.018***	.006***	.013***
	(.013)	(.000)	(.005)	(.002)	(.003)
Highest grade father	.064***	-.002***	-.023***	.008***	.017***
	(.010)	(.000)	(.004)	(.001)	(.003)
Family income bottom quartile	-.126	.005	.045	-.018	-.032
	(.197)	(.008)	(.070)	(.030)	(.047)
Family income second quartile	.035	-.001	-.013	.005	.009
	(.096)	(.003)	(.035)	(.012)	(.026)
Family income third quartile	.037	.001	-.013	.005	.010
	(.063)	(.002)	(.023)	(.008)	(.017)
Family income top quartile	.029	.001	-.010	.004	.008
	(.048)	(.001)	(.017)	(.006)	(.012)
AFQT score (adjusted)	.025***	.001***	-.009***	.003***	.006***
	(.001)	(.000)	(.000)	(.000)	(.000)
Black	.510***	-.011***	-.187***	.039***	.159***
	(.073)	(.002)	(.026)	(.005)	(.024)
Hispanic	.452***	-.010***	-.166***	.034***	.141***
	(.087)	(.002)	(.031)	(.005)	(.030)
Number of siblings	-.023	.001	.008	-.003	-.006
	(.012)	(.000)	(.004)	(.001)	(.003)
Average unemployment rate during schooling years	1.454***	-.048***	-.532***	.192***	.387***
	(.083)	(.007)	(.033	(.019)	(.023)
Average GDP Growth	-.144***	. 005***	.053***	-.019***	-.038***
	(.034)	(.001)	(012)	(.004)	(.009)
Geographic controls	Yes	Yes	Yes	Yes	Yes
Cohort and age controls	Yes	Yes	Yes	Yes	Yes
κ	12.221***				
	(.661)				
	14.457***				
	(.685)				
	15.350***				
	(.693)				
Wald χ^2	1,190.73				

Note: Standard errors in parentheses. *p<0.05; **p<0.01; ***p<0.001

Table 2. Schooling: OLS - NLSY

	(I)	(II)	(III)	(IV)
Highest grade mother	.102***	.113***	.101***	.100***
	(.006)	(.006)	(.005)	(.006)
Highest grade father	.085***	.088***	.086***	.084***
	(.004)	(.005)	(.004)	(.004)
Family income bottom quartile	-.220**	-.140	-.269***	-.241***
	(.077)	(.080)	(.080)	(.070)
Family income second quartile	.084***	.125***	.073	. 061
	(.039)	(.041)	(.040)	(.036)
Family income third quartile	-.058***	-.020	-.068***	-.068***
	(.025)	(.026)	(.026)	(.023)
Family income top quartile	-.001	.053***	-.010	-.027
	(.019)	(.020)	(.020)	(.017)
AFQT score (adjusted)	.034***	.040***	.034***	.034***
	(.000)	(.000)	(.000)	(.000)
Black	.718***	.847***	.680***	.643***
	(.025)	(.027)	(.025)	(.026)
Hispanic	.767***	.841***	.766***	.745***
	(.033)	(.037)	(.033)	(.033)
Number of siblings	-.039***	-.057***	-.036***	-.032***
	(.005)	(.005)	(.005)	(.005)
Average unemployment rate during schooling years	1.618***		1.866***	1.616***
	(.026)		(.030)	(.038)
Average GDP Growth		.223***	-.161***	-.212
		(.012)	(.010)	(.010)
Average Consumer Sentiment				.061***
				(.004)
Geographic controls	Yes	Yes	Yes	Yes
Cohort and age controls	Yes	Yes	Yes	Yes
Constant	-4.039***	7.789***	-5.503***	-7.853***
	(.222)	(.107)	(.241)	(.236)
R^2	.427	.504	.511	.511

Note: Standard errors in parentheses. *p<0.05; **p<0.01; ***p<0.001

Table 3. Wages: GLS, IV and Heckman Estimations - NLSY

	(1) GLS	(2) IV	(3) Categorical GLS	(4) Heckman
Years of schooling	.074***	.050***		
	(.002)	(.006)		
Work Exp.	.091***	.039***	.094***	.111***
	(.005)	(.002)	(.005)	(.024)
Experience^2	-.001***	.000	-.001***	-.002
	(.000)	(.000)	(.000)	(.001)
Highest grade mother	.001	.010***	.008***	.006
	(.002)	(.002)	(.002)	(.004)
Highest grade father	.004**	.004***	.009***	.006
	(.002)	(.001)	(.002)	(.003)
Siblings	.002	-.007	-.002	-.008
	(.002)	(.002)	G(.002)	(.004)
Family income 1stq	-.094*	-.029	-.099**	-.162*
	(.037)	(.022)	(.038)	(.078)
Family income 2ndq	-.060	-.013	-.047***	-.050
	(.037)	(.011)	(.038)	(.037)
Family income 3rdq	-.071	.003	-.069	-.040
	(.037)	(.007)	(.037)	(.024)
Family income 4thq	.017	.016***	.015	-.013
	(.037)	(.005)	(.037)	(.018)
AFQT score (adjusted)	.004***	.005***	.007***	.007***
	(.000)	(.000)	(.002)	(.000)
Black	-.063***	-.040***	.013	.003
	(.013)	(.013)	(.013)	(.032)
Hispanic	.017	.060	.070***	.045
	(.015)	(.014)	(.015)	(.033)
Unemployment at entry	-.010**	-.005	-.006*	.000
	(.003)	(.002)	(.003)	(.008)
No high school			-.059	-.401
			(.031)	(.739)
Some college			.030	-.495**
			(.027)	(.220)
College			.123***	1.471**
			(.026)	(.426)
Selectivity adjustments				
No high school				.538**
				(.195)
High school				.731
				(.424)
Some college				.429*
				(.213)
College				-.166
				(.327)
Geographic controls	Yes	Yes	Yes	Yes
Cohort and age controls	Yes	Yes	Yes	Yes
Constant	4.037***	.831***	4.688***	1.405***
	(.060)	(.075)	(.058)	(.132)
N	21,763	21,763	21,763	21,763
R^2		.307		.332
Wald χ^2	8,636.61		7,433.03	

Note: Standard errors in parentheses. *p<0.05; **p<0.01; ***p<0.001 Bootstrapped standard errors for Heckman estimation based on 200 replications. Reference category for family income is “family income not observed”.

Table 4. Estimates of variances of wages, US men.

	Less than high school		High school		Some College		College	
Observed wage inequality	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Permanent component	.340***	.218	.301***	.214	.340***	.267	.287***	.292
	(.012)	(0.26)	(.006)	(0.14)	(.013)	(0.28)	(.006)	(.022)***
B. ($\hat{\psi}^2_{st}$)-Transitory component	.028***	.293	.020***	.197	.019***	.233	.015***	.221
	(.002)		(.001)		(.001)		(.001)	
Observed inequality (A+B)	**.368*** **	**.511**	**.320**	**.411**	**.359*** **	**.500**	**.302*** **	**.513**
	(.012)		**(.006)**		**(.013)**		**(.006)**	
Potential wage inequality								
C. (σ^2_s)-Permanent component	.613***	.284	.554***	.242	.458***	.274	.293***	.356
(Adjusted for selection and truncation biases)	(.049)		(.119)		(.089)		(.048)	
Potential wage inequality (C+B)	**.641*** **	**.577**	**.574*** **	**.439**	**.476*** **	**.507**	**.308*** **	**.577**
	(.049)		**(.119)**		**(.089)**		**(.048)**	
Wage uncertainty								
D. Correlation coefficient	.815***	-.568	.842**	-.371	.554**	-.190	-.328	-.534
	(.163)		(.313)		(.175)		(.386)	
E. Permanent component (C-C*D^2)	.206***	.192	.161**	.209	.317***	.264	.261***	.251
(Accounted for unobserved Schooling Factor)	(.022)		(.056)		(.014)		(.038)	
Degree of wage uncertainty (E+B)	**.234*** **	**.485**	**.181*** **	**.197**	**.336*** **	**.233**	**.277*** **	**.221**
	(.022)		**(.051)**		**(.014)**		**(.038)**	
γ **-Unobserved heterogeneity**(C-E)	**.407*** **	**.092**	**.393** **	**.033**	**.141**	**.010**	**.031**	**.105**
	(.066)		**(.175)**		**(.093)**		**(.086)**	

Note: Columns (1), (3), (5) and (7) are our estimates, columns (2), (4), (6) and (8) are taken from Chen (2008, E+B less than high school corrected). Standard errors based on 500 replications in parentheses. Standard errors in parentheses *p<0.05; **p<0.01; ***p<0.001.

Table 5. Correlation coefficients estimates by schooling level.

Instruments:	**Unemp.**	**GDP growth**	**Cons. Sent.**	**Unemp./GDP**	**All**
No high school	.756	.799	.768	.815	.811
High school	.908	.980	.970	.842	.822
Some college	.630	.715	.668	.554	.554
College	-.370	-.457	-.735	-.328	-.351
F-test	220.82	12.67	87.05	2,530.02	1,874.84
	(.000)	(.000)	(.000)	(.000)	(.000)
Hansen-Sargan test				(.236)	(.875)

Note: *p-values* in parentheses.

Table 6. Schooling: OLS and Ordered Probit estimation - BHPS

	OLS	Probit						
		Coefficient	Marginal effects at means					
			No qualification	Some qualification	GCSEs	A levels	First Degree	Higher degree
Father ed.	.141***	.130***	-.009***	-.013***	-.033***	.030***	.022***	.004***
	(.023)	(.010)	(.001)	(.001)	(.002)	(.002)	(.001)	(.001)
Mother ed.	-.001	.129***	-.008***	-.011***	-.028***	.025***	.019***	.003***
	(.025)	(.011)	(.001)	(.001)	(.002)	(.002)	(.001)	(.000)
Minority	1.368***	.292***	-.011***	-.017***	-.050***	.037***	.034**	.007*
	(.111)	(.050)	(.002)	(.003)	(.011)	(.006)	(.008)	(.002)
Urban	.269***	.184***	-.011***	-.015***	-.035***	.035***	.023***	.004**
	(.056)	(.022)	(.002)	(.002)	(.004)	(.005)	(.003)	(.000)
Siblings	-.022**	-.035***	.002***	-.003***	-.008***	-.007***	-.005***	-.001***
	(.007)	(.003)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Avg. unemp.	.632***	.232***	-.015***	-.022***	-.054***	.049***	.036***	.006***
	(.010)	(.006)	(.000)	(.001)	(.001)	(.001)	(.001)	(.000)
Avg. GDP growth	-.207***	-.094***	.005***	.007***	.017***	-.016***	-.012***	-.002***
	(.020)	(.011)	(.001)	(.001)	(.002)	(.002)	(.001)	(.000)
Age	.269***	.086***	-.005***	-.008***	-.020***	.018	.013	.002***
	(.003)	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)

Table 6. Schooling: OLS and Ordered Probit estimation - BHPS continued

	OLS	Probit						
		Coefficient	Marginal effects at means					
			No qualification	Some qualification	GCSEs	A levels	First Degree	Higher degree
κ_1	2.763***	2.658***						
	(.158)	(.125)						
κ_2		4.325***						
		(.085)						
κ_3		4.903***						
		(.085)						
κ_4		6.171***						
		(.090)						
κ_5		7.535***						
		(.095)						
N	14,538	14,538						
R^2	.429							
Wald χ^2		5,838.57						

Standard errors in parentheses *p<0.05; **p<0.01; ***p<0.001

Table 7. Wages: GLS, IV and Heckman Estimations BHPS

	(1) GLS	(2) IV	(3) Categorical GLS	(4) Heckman
Years schooling	.046***	. .077***		
	(.001)	(.003)		
Work exp.	.085***	.109***	.083***	.989***
	(.001)	(.003)	(.001)	(.108)
Experience 2	-.001***	-.002***	-.001***	-.043***
	(.000)	(.000)	(.000)	(.008)
Mother ed.	-.052***	.051***	.027***	-.229
	(.004)	(.006)	(.004)	(.077)
Father ed.	.031***	.025***	.017***	.050
	(.004)	(.005)	(.003)	(.098)
Siblings	-.001	-.010***	.001	.075
	(.001)	(.001)	(.001)	(.115)
Minority	-.106***	-.140***	-.101***	-.498
	(.019)	(.029)	(.018)	(.371)
Urban	.050	-.075***	.030***	.150
	(.008)	(.013)	(.008)	(.037)
Unemp. at entry	.014****	.005***	.014***	.023
	(.005)	(.001)	(.000)	(.017)
Noqualifications			-.153***	-.101
			(.011)	(.168)
Some qualification			-.048***	-242*
			(.012)	(.110)
A-levels			.174***	.234***
			(.008)	(.038)
First degree			.493***	.648***
			(.010)	(.100)
Higher degree			.531***	.133
			(.018)	(.518)
Selectivity Adjustments				
No qual.				.015
				(.090)
Some qual.				-.132
				(.086)
GCSE				-.044
				(.047)

Table 7. Wages: GLS, IV and Heckman Estimations BHPS. Continued.

	(1) GLS	(2) IV	(3) Categorical GLS	(4) Heckman
A level				-.122**
				(.047)
First degree				-.060
				(.086)
Higher degree				.218
				(.239)
Constant	1.400*	5.027****	6.060***	6.547***
	(.485)	(.053)	(.023)	(.095)
N	3,197	3,1975	3,194	3,194
R^2		.302		
Wald χ^2	8,174.18		10,651.96	716.86

Standard errors in parentheses *p<0.05; **p<0.01; ***p<0.001.

Table 8. Estimates of variances of wages, UK men.

	No qualification	Some qualification	GCSE	A levels	First degree	Higher degree
Observed wage inequality						
A. Permanent component	.089***	.096**	.120***	.096***	.091***	.128***
	(.003)	(.006)	(.004)	(.003)	(.005)	(.014)
B. ($\hat{\psi}_{st}^{2}$)-Transitory component	.000***	.000	.000	.002***	.002***	.003
	(.000)	(.000)	(.000)	(.000)	(.000)	(.003)
Observed inequality (A+B)	**.089*** **	**.096** **	**121*** **	**.098*** **	**.093*** **	**.131*** **
	(.003)	**(.006)**	**(.004)**	**(.003)**	**(.005)**	**(.014)**
Potential wage inequality						
C. (σ_{s}^{2})-Permanent component	.091***	.109***	.119***	.098***	.152***	.139***
(Adjusted for selectivity and truncation biases)	(.004)	(.008)	(.004)	(.004)	(.013)	(.017)
Potential wage inequality (C+B)	**.090*** **	**.109*** **	**.119*** **	**.100*** **	**.153*** **	**.142*** **
	(.004)	**(.008)**	**(.004)**	**(.004)**	**(.013)**	**(.017)**
Wage uncertainty						
D. Correlation coefficient	.074	-.353***	.010	-.219**	-.665***	-.358***
	(.165)	(.081)	(.071)	(.065)	(.046)	(.115)
E. Permanent component (C-C*D ^{2})	.091***	.095***	.119***	.093***	.084***	.121***
(Accounted for unobserved schooling factor)	(.003)	(.006)	(.004)	(.003)	(.005)	(.014)
Degree of wage uncertainty (E+B)	**.091*** **	**.095** **	**.119*** **	**.095*** **	**.086*** **	**.125*** **
	(.003)	**(.006)**	**(.004)**	**(.003)**	**(.005)**	**(.014)**
γ **-Unobserved heterogeneity**(C-E)	**.000**	**.013*** **	**.000**	**.005**	**.067*** **	**018** **
	(.003)	**(.006)**	**(.001)**	**(.003)**	**(.013)**	**(.012)**

Note: standard errors based on 500 bootstrap replications in parentheses. Standard errors in parentheses
*p<0.05; **p<0.01; ***p<0.001.

Table 9. Schooling: OLS and Ordered Probit estimation - SOEP

	OLS	Probit							
		Coefficients	Marginal Effects at means						
			No high school	Intermediate School	Voc. High School	High School	Higher voc.	Technical Uni.	University
Mother ed.	.108***	.070***	-.000***	-.005***	-.022***	.001***	.011***	.005***	.010***
	(.013)	(.007)	(.000)	(.000)	(.002)	(.000)	(.001)	(.000)	(.001)
Father ed.	.276***	.127***	-.001***	-.008***	-.039***	.002***	.019***	.009***	.018***
	(.011)	(.006)	(.000)	(.000)	(.002)	(.000)	(.001)	(.000)	(.001)
German	.912***	.557***	-.006***	-.055***	-.130***	.015***	.086***	.033***	.056***
	(.054)	(.030)	(.001)	(.004)	(.004)	(.001)	(.004)	(.001)	(.002)
Siblings	-.157***	-.085***	.000***	.006***	.026***	-.002***	-.013***	-.006***	-.012***
	(.008)	(.005)	(.000)	(.000)	(.001)	(.000)	(.001)	(.000)	(.001)
Age	.102***	.049***	-.000***	-.003***	-.015***	.001***	.007***	.003***	.007***
	(.002)	(.001)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)	(.000)
Avg. Unemp.	.117***	.068***	-.000***	-.004***	-.021***	.001***	.010***	.005***	.009***
	(.011)	(.006)	(.000)	(.000)	(.002)	(.000)	(.001)	(.000)	(.001)
Avg. GDP growth	-2.033***	-.952***	.005***	.064***	.292***	-.018***	-.143***	-.066***	-.133***
	(.031)	(.022)	(.000)	(.001)	(.008)	(.001)	(.004)	(.022)	(.003)
Constant	14.485								
	(.259)								
κ_1		-3.434***							
		(.173)							

Table 9. Schooling: OLS and Ordered Probit estimation – SOEP. Continued

	OLS	Probit							
		Coefficients	Marginal Effects at means						
			No high school	Intermediate School	Voc. High School	High School	Higher voc.	Technical Uni.	University
κ_2		-2.330***							
		(.156)							
κ_3		-.162***							
		(.149)							
κ_4		-.026***							
		(.149)							
κ_5		.656***							
		(.147)							
κ_6		.969***							
		(.147)							
N	24,869	24,869	325	2,425	15,671	876	4,015	1,426	3,930
R^2	.468								
Wald χ^2		8,449.22							

Standard errors in parentheses *p<0.05; **p<0.01; ***p<0.001.

Table 10. Wages: GLS, IV and Heckman Estimations - SOEP

	(1) GLS	(2) IV	(3) GLS Cathegorical	(4) Heckman
Years School	.099***	.115***		
	(.002)	(.008)		
Work Exp.	.090***	.103***	.088***	.102***
	(.002)	(.003)	(.002)	(.007)
Work Exp.^{2}	-.001***	-.002***	-.001***	-.002***
	(.000)	(.000)	(.000)	(.000)
Mother ed.	-.003	-.004	-.002	.000
	(.003)	(.004)	(.003)	(.009)
Father ed.	.004	.000	.006*	.008
	(.003)	(.003)	(.003)	(.007)
German	-.053	-.064***	-.027*	-.003
	(.029)	(.016)	(.013)	(.029)
Siblings	-.006**	-.006*	-.008***	-.013**
	(.002)	(.002)	(.002)	(.005)
Age	-.018***	-.023***	-.015***	-.014***
	(.001)	(.003)	(.001)	(.003)
Unemp. level at entry	.002	-.004	.008***	.004
	(.001)	(.002)	(.001)	(.003)
No High School			-.480***	.476
			(.040)	(.261)
Intermediate School			-.309***	-.124
			(.024)	(.089)
Voc. High School			-.108***	-.003
			(.024)	(.054)
Higher voc.			.169***	.250***
			(.023)	(.053)
Technical Uni.			.311***	.515***
			(.026)	(.068)
University			.535***	.709***
			(.023)	(.088)
Selectivity Adjustments				
No high school				.237**
				(.073)
Intermediate School				.086
				(.057)

Table 10. Wages: GLS, IV and Heckman Estimations – SOEP. Continued.

	(1) GLS	(2) IV	(3) GLS Cathegorical	(4) Heckman
Voc. High School				-.022
				(.030)
High School				.086
				(.057)
Higher voc.				.081**
				(.035)
Technical Uni.				-.089
				(.068)
University				-.031
				(.066)
Constant	6.036***	5.982***	7.135***	7.856***
	(.031)	(.033)	(.043)	(.678)
N	4,121	4,121	4,121	4,121
R^2		.225		.318
Wald χ^2	7,743.34		7,490.87	

Standard errors in parentheses *p<0.05; **p<0.01; ***p<0.001.

Table 11. Estimates of variances of wages, German men

	No high school	Intermediate School	Voc. High School	High School	Higher voc.	Technical Uni.	University
Observed wage inequality							
A. Permanent Component	.481***	.202***	.158***	.362***	.240***	.180***	.274***
	(.083)	(.013)	(.006)	(.031)	(.012)	(.013)	(.017)
B. ($\hat{\psi}_{st}^2$)-Transitory Component	.022***	.013***	.012***	.007*	.025***	.019***	.011**
	(.005)	(.002)	(.002)	(.004)	(.003)	(.003)	(.005)
Observed Inequality (A+B)	**.503***	**.215***	**.170***	**.370***	**.265***	**.200***	**.285***
	(.085)	**(.014)**	**(.006)**	**(.032)**	**(.014)**	**(.014)**	**(.018)**
Potential wage inequality							
C. (σ_s^2)-Permanent component	.583***	.195***	.144***	.349***	.221***	.175***	.254***
(Adjusted for selection and truncation biases)	(.121)	(.014)	(.006)	(.032)	(.012)	(.015)	(.018)
Potential Inequality (C+B)	**.605***	**.207***	**.156***	**.357***	**.246***	**.194***	**.265***
	(.124)	**(.015)**	**(.006)**	**(.033)**	**(.013)**	**(.015)**	**(.019)**
Wage uncertainty							
D. Correlation Coefficient	.488**	.198*	-.032	.092	.159***	-.259**	-.096
	(.152)	(.074)	(.039)	(.064)	(.041)	(.088)	(.078)
E. Permanent Component	.444***	.187***	.144***	.346***	.216***	.163***	.252***
(Accounted for Unobserved. Schooling Factor)	(.078)	(.013)	(.006)	(.031)	(.012)	(.013)	(.017)
F. Degree of Wage Uncertainty (E+B)	**.466***	**.200***	**.156***	**.354***	**.240***	**.182***	**.263***
	(.081)	**(.013)**	**(.006)**	**(.032)**	**(.012)**	**(.013)**	**(.018)**
γ - Unobserved Heterogeneity	**.139***	**.008**	**.001**	**.003**	**.006***	**.012**	**.002**
	(.069)	**(.005)**	**(.000)**	**(.005)**	**(.002)**	**(.008)**	**(.005)**

Note: standard errors based on 500 bootstrap replications in parentheses. Standard errors in parentheses *p<0.05; **p<0.01; ***p<0.001.

Table 12. Key results for US women

	No qualification	High school	Some college	College
Corr. coeff.	.525**	.900***	.533**	-.287
	(.178)	(.201)	(.166)	(.331)
Risk	.459***	.167**	.296***	.285
	(.049)	(.053)	(.021)	(.031)
Unobs. Het.	.162*	.596***	.107	.023
	(.061)	(.153)	(.064)	(.051)

Note: Standard errors in parentheses *p<0.05; **p<0.01; ***p<0.001

Table A1. Mean and standard deviation time invariant variables NLSY79

	Our sample	Chen's sample
(a)Schooling variables		
Years of schooling	13.33	13.44
	(2.45)	(2.50)
Categorical education		
No high school	.11	.10
	(.32)	(.30)
High school	.42	.43
	(.49)	(.50)
Some college	.22	.21
	(.41)	(.41)
Four-year college or beyond	.27	.26
	(.41)	(.44)
(b) Ability and family background		
Armed forces qualifying test score (adjusted)	57.88	62.35
	(28.08)	(28.50)
Highest grade mother	11.81	11.85
	(2.66)	(2.61)
Highest grade father	11.93	12.01
	(3.63)	(3.53)
Number of siblings	3.19	3.16
	(2.18)	(2.17)
Family income	50,317*	50,321*
	(34,417)	(34,544)
Black	.11	.11
	(.31)	(.31)
Hispanic	.05	.05
	(.22)	(.22)
(c)Geographic controls at age 14		
Urban	.77	.77
	(.42)	(.42)
Northeast	.20	.21
	(.40)	(.41)
South	.30	.29
	(.46)	(.45)
West	.16	.15
	(.36)	(.36)
(d) Instrument for schooling		
Average unemployment during schooling years	7.09	
	(6.71)	
Average GDP growth during schooling years	2.51	
	(1.53)	
Average cons. sent. during schooling years	76.89	
	(4.85)	
Unemployment rate at entry	7.50	
	(1.42)	
N	4,302	4,302

*1999 dollars. Average unemployment rates calculated on CPS data. Consumer sentiment index calculated on Thomson Reuters / University of Michigan Surveys of Consumers data. Standard deviations in parentheses

Table A2. Mean and standard deviation time variant variables NLSY79, selected years

	Our Sample				
	Calendar year				
Labor market variables	1990	1993	1995	1997	1999
Actual work experience	9.93	11.94	13.70	15.40	17.22
	(3,35)	(3.80)	(4.06)	(4.32)	(4.58)
Log Hourly earnings	2.35	2.50	2.60	2.70	2.83
	(.67)	(.66)	(.68)	(.84)	(.84)
Age	28.50	31.50	33.50	35.50	37.50
	(2.26)	(2.26)	(2.26)	(2.26)	(2.26)
Sample Size	4,302	4,302	4,302	4,302	4,302
	Chen's sample				
	Calendar year				
Labor market variables	1990	1993	1995	1997	1999
Actual work experience	9.03	11.47	13.25	15.01	16.74
	(3.37)	(3.87)	(4.11)	(4.34)	(4.67)
Log Hourly earnings*	2.42	2.47	2.51	2.59	2.70
	(.68)	(.69)	(.70)	(.84)	(.85)

Note: *in 1992 dollars. Standard deviations in parentheses.

Table B 1. Mean and standard deviation time invariant variables – BHPS

Schooling variables		Highest grade father	3.22
Years of Schooling	13.38		(1.04)
	(3.05	Number of siblings	2.92
Categorical education			(1.02)
No qualification	.18		
	(.38)		
Some qualification	.09	Minority	.03
	(.29)		(.18)
GCSEs	25	Urban	.83
	(.43)		(.38)
A-levels	.35	Avg. unemp. rate	12.45
	(.48)		(2.31)
First degree	.10	Avg. GDP growth	1.64
	(.30)		(1.49)
Higher degree	.02	Unemp. rate at entry	5.56
	(.16)		(8.01)
Highest grade mother	2.70		
	(.90)		

Note: Standard deviations in parenthesis

Table B 2. Mean and standard deviation time variant variables – BHPS, selected years

	1991	1994	1997	2000	2003	2007
Work experience	20.84	20.46	21.69	22.77	25.00	27.36
	(13.64)	(13.25)	(.64)	(12.82)	(12.15)	(11.27)
Log monthly wage	6.99	7.06	7.14	7.28	7.42	7.59
	(.60)	(.67)	(.64)	(.58)	(.60)	(.60)
Age	39.39	38.63	39.19	40.05	41.63	43.90
	(13.12)	(13.17)	(13.00)	(13.04)	(13.04)	(12.25)
Sample Size	3,385	3,073	3,462	4,589	3,514	3,397

Note: Standard deviations in parenthesis. Work experience is measured as Age – Years of Schooling – 5.

Table C 1. Mean and standard deviation time invariant variables – SOEP

Years of Schooling	12.61	Highest grade mother	3.07
	(2.82)		(1.32)
Categorical education:		Highest grade father	3.32
No qualification	.01		(1.59)
	(.10)	Number of siblings	1.97
Intermediate school	.08		(1.77)
	(.26)	German	.90
Vocational High school	.55		(.29)
	(.50)	West German	.78
High school	.03		(.41)
	(.20)	Avg. unemp. rate	4.79
Higher vocational education	.14		(1.82)
	(.33)	Avg. GDP growth	2.14
Technical university	.05		(1.62)
	(.21)	Unemp. rate at entry	3.13
University	.14		(3.66)
	(.36)		

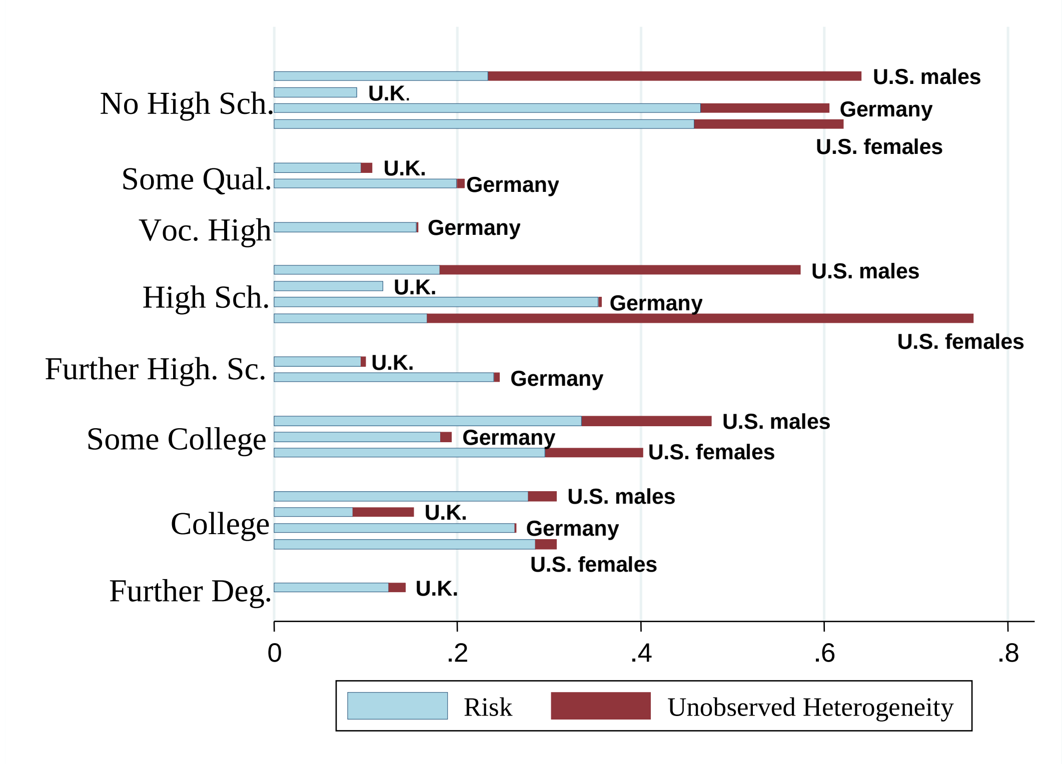
Note: Standard deviations in parenthesis

Table C 2. Mean and standard deviation time variant variables – SOEP, selected years

	1999	2001	2003	2005	2007	2008
Work experience	18.45	19.85	20.67	20.95	21.69	22.29
	(10.54)	(10.71)	(10.21)	(10.21)	(10.03)	(9.82)
Log monthly wage	7.29	7.38	7.49	7.49	7.52	7.56
	(.59)	(.59)	(.73)	(.79)	(.74)	(.77)
Age	41.13	42.61	43.96	44.39	45.37	46.02
	(9.76)	(9.83)	(9.97)	(10.01)	(9.71)	(9.48)
Sample Size	1,455	2,471	2,885	3,102	3,409	3,219

Note: Standard deviations in parentheses.

Figure 1. Risk and unobserved heterogeneity.



Note: The graph represents the last two rows of the tables that decompose wage variance ("Degree of wage uncertainty" and " Unobserved heterogeneity".