Labour Economics 24 (2013) 323–338

Contents lists available at ScienceDirect

Labour Economics

journal homepage: www.elsevier.com/locate/labeco

Unobserved heterogeneity and risk in wage variance: Does more schooling reduce earnings risk?☆

Jacopo Mazza [a,*], Hans van Ophem [b], Joop Hartog [b]

[a] University of Manchester, Arthur Lewis Building, Oxford Rd., M13 9PL Manchester, UK
[b] Universiteit van Amsterdam, The Netherlands

HIGHLIGHTS

- Is wage variability by education dominated by risk or by unobserved heterogeneity?
- Is risk monotonically related to level of education?
- We apply a model developed by Chen (2008) to datasets for US, UK and Germany.
- Risk dominates over unobserved heterogeneity.
- Increasing levels of education are not associated with lower levels of risk.

ARTICLE INFO

Article history:
Received 30 May 2012
Received in revised form 18 September 2013
Accepted 19 September 2013
Available online 2 October 2013

JEL classification:
C01
C33
C34
J31

Keywords:
Wage inequality
Wage uncertainty
Unobserved heterogeneity
Selectivity
Education
Replication

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.

1. 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.[1]

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

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

* Corresponding author.
E-mail address: jacopo.mazza@manchester.ac.uk (J. Mazza).

[1] 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 et al. (2004) and Hartog and Diaz Serrano (forthcoming).

0927-5371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.labeco.2013.09.007

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.[2] 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 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 2 we set forth Chen's model. Section 3 presents results for the US, Section 4 for the UK and Section 5 for Germany. In Section 6 we compare our results to the original Chen results and Section 7 concludes.

2. Chen's model

2.1. 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.[3] 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:

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

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:

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

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

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

[3] 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.