

Volume 29, Issue 1

Changes in the Characteristics and Skills of British Youth

Geraint Johnes
Lancaster University Management School

Abstract

Changes in the characteristics and skills of British youths between the mid-1980s and mid-2000s are evaluated using a method recently developed by Altonji et al. The main finding is that skills have increased over time in successive cohorts of young people. The improvement is, however, uneven, and those at the bottom end of the skills distribution have benefitted less than others. This implies, other things being equal, that the distribution of earnings will widen over the coming years.

1. Introduction

In recent work, Altonji *et al.* (2008) have shown that, while recent cohorts of young people entering the American labour market have greater levels of skills than their predecessors, the gains in skills have not been uniformly distributed. Between 1979 and 1997, the skills of youths in the middle 60 per cent of the skills distribution rose by around 5 per cent. But for those in the top two deciles the gains were larger, while for those in the bottom two deciles the gains were smaller. This implies that, other things being equal, the distribution of earnings will become wider over the coming years as young workers who entered the labour market around the turn of the century reach their peak lifetime earnings.

The method used by Altonji *et al.* is prospective. It uses information about the characteristics of workers when they were young along with data on their earnings at a later stage in their lives, when earnings typically reach a peak. It then reweights the distribution of characteristics, using data on a later cohort of young people, and uses the mapping of characteristics onto earnings (from the earlier cohort) to evaluate what one might expect the earnings of the later cohort to be at some date in the future.

In the United Kingdom there have been many changes to the education system that one might expect should have an impact on the skills of young people. Amongst these are the replacement of the General Certificate of Education Ordinary Level (GCE O level) and the Certificate of Secondary Education (CSE) examinations by the General Certificate of Secondary Education (GCSE) in 1987 and the introduction of a national curriculum, Grant Maintained schools (funded directly by central rather than local government), parental choice of school and school league tables in the wake of the 1988 Education Reform Act. The competitive forces introduced by the last of these reforms, in particular, might be expected to impact upon students' skills (Bradley *et al.*, 2001). It is therefore of interest to evaluate whether or not the evolution of young people's skills in the UK is similar to that observed by Altonji *et al.* in the USA.

In the next section I discuss data issues and methodology. Results are reported in Section 3, and the paper ends with a short concluding section.

2. Data and Methodology

The US data used by Altonji *et al.* come from the National Longitudinal Survey of Youth. This is a particularly attractive dataset in the present context because it gathers a long series of longitudinal data on several successive cohorts of youths. The length of the panel allows information to be gathered about a cohort's earnings once they have reached maturity. In the UK, there is no single dataset that can offer both this long panel and multiple cohorts over an extended period.

There is, however, a close substitute. The British Cohort Study (BCS) is a long panel comprising people who were born during the week beginning 5 April 1970. More recently, a suite of Youth Cohort Studies (YCS), each made up of a short panel based on more recent cohorts of young people, has gathered information about respondents' backgrounds. In the present study, I use the BCS to provide information about an

earlier cohort of individuals, and the 2004-07 YCS (cohort 12) to give data about the later cohort. The latter are the most recent data currently available.

The match between the BCS and YCS is not perfect. The first covers the whole of the UK while the second covers only England and Wales. There are differences in the wording of questions. There are differences in environment (such as the change from O level and CSE to GCSE) that make comparisons across time hazardous. Nevertheless the two data sources appear to offer the best opportunity to conduct for the UK an analysis that is similar to that provided by Altonji *et al.* for the USA.

A further difference between the BCS and YCS that can be corrected for in the analysis concerns the nature of the sample in these two surveys. The BCS, by design, is a wholly random sample. The YCS, on the other hand, oversamples some groups. In particular, there is an oversampling procedure for ethnic minorities. This is allowed for in the present study by appropriate weighting of the data during estimation.

The method introduced by Altonji *et al.* draws heavily upon the work of DiNardo *et al.* (1996). This involves using wages at prime age as a unidimensional indicator of workers' skills, while recognising that such skills are a function of, *inter alia*, a multiplicity of characteristics possessed by individuals when young. In comparing skills across two cohorts, the method first evaluates the density of prime age wages enjoyed by the earlier cohort. To evaluate a corresponding ('counterfactual') density for the later cohort, a set of weights is applied to the wage density of the earlier cohort. These weights reflect the probability with which an individual is drawn from the second, rather than the first, cohort, based on their skills vectors.

To be precise, the data from the BCS and YCS are pooled in order to run a probit model (appropriately weighted to allow for the oversampling of minorities in the YCS) in which the dependent variable is a binary indicator of the survey from which the data are drawn. Observations from the BCS are assigned a zero value for this variable, while those from the YCS are assigned unit value. For the i th observation, denote the latent index underpinning this binary variable by θ_i . I then model this as

$$\begin{aligned} \theta_i = & \sum_{s,r} sex_{s,i} race_{r,i} \beta_{s,r} + \sum_{s,r} sex_{s,i} race_{r,i} paed_i \beta_{p,s,r} + \sum_{s,r} sex_{s,i} race_{r,i} maed_i \beta_{m,s,r} \\ & + \sum_{s,r} sex_{s,i} race_{r,i} together \beta_{t,s,r} + \sum_{s,r} sex_{s,i} race_{r,i} olevels \beta_{o,s,r} + \varepsilon_i \end{aligned} \quad (1)$$

where *paed* and *maed* respectively denote binary variables indicating whether or not the respondent's father or mother undertook post-compulsory education. The variable *together* takes unit value if the respondent's father and mother were living together (and with the respondent) when the respondent was aged 16. The *olevels* variable is likewise binary, taking unit value if the respondent achieved at least five GCSE passes at grades A*-C, or (in the case of the BCS) an equivalent number of O level passes and CSE grade 1 scores. The interpretation of the two *sex* (male and female) and two *race* (majority and minority) variables is obvious, and the ε term denotes regression error. Note that the parental and educational background variables are fully interacted with the gender and ethnicity indicators, and that the presence of the summation terms ensures that, in effect, the interactions and each individual interacted variable are all included as explanatory variables.

Denote by $p(t_1|x)$ the propensity score associated with an observation being drawn from the BCS, conditional on the characteristics vector x for that observation. Let $p(t_2|x)$ ($\equiv 1-p(t_1|x)$) be the corresponding propensity score associated with an observation being drawn from the YCS. Let $p(t_1)$ and $p(t_2)$ denote the number of observations in the BCS and YCS samples respectively. Then, following DiNardo *et al.* (1996) and Altonji *et al.* (2008),

$$\psi = p(t_2|x) p(t_1) / p(t_1|x) p(t_2) \quad (2)$$

is the vector of weights which, when applied to the wage density obtained from the BCS data, will provide the counterfactual wage density – that is, the density of wages that, given the skills of the YCS cohort, is expected to obtain for this group once they reach prime age. The values of $p(t_1)$ and $p(t_2)$ are 15554 and 14003 respectively. In contrast to Altonji *et al.* (2008) we do not cap propensity weights because we do not encounter any implausibly high values.

3. Results

The results of the probit are reported in Table A1 in the appendix. For each observation, this equation can be used to compute the propensity score following equation (2) above.

The wage data that are used for the BCS cohort are drawn from the sixth follow-up, which took place when respondents were aged 29.¹ Since this follow-up is taken at just one point in time, there is a variety of reasons why respondents may not have been observed. These include attrition over the lifetime of the study, and the fact that some respondents are unemployed or out of the labour force; moreover there is substantial non-response to the BCS question about gross wage. The BCS sample for which wage data are available is therefore small in relation to the sample upon which the calculation of the weights, ψ , is based. A small number of outliers who report hourly wages below £1 or above £25 is omitted. (The minimum wage at this time was £3.60 per hour.) This leaves 2825 in the sample.

It is then straightforward to evaluate the wage density using kernel methods; in this case I use the Epanechnikov (1969) kernel and the bandwidth is chosen to be that which would minimise the mean integrated squared error under Gaussian conditions. This is the default method used within the STATA software.

In Figure 1, the actual density of prime age wages is shown for the BCS sample by the solid line. The counterfactual density is shown by the dotted line. The latter clearly lies to the right of the former at all but the lowest levels of the wage. This indicates that the skills set possessed by young people has improved over time – albeit not uniformly.

¹ Data from a seventh follow-up, conducted at age 34, have recently become available, but are not used in the present study.

This can be seen also in Figure 2, which shows how the log wage has changed at each point along the skills distribution. I do not plot the first or last 5 centiles owing to the presence of outliers. The horizontal axis plots the centiles of the wage distribution, while the vertical axis shows the change in log wage. Typically young people's skills increased by about 7 per cent from the mid-1980s to the mid-2000s. For the bottom two deciles, however, the improvement in skills has not been so pronounced. Indeed, while results for those in the bottom 5 centiles need to be treated with caution, it is nonetheless worth noting a decline in measured skills for this group.

4. Conclusion

The results reported above suggest that young people are on average more skilled now than in the past, but also that the distribution of skills is becoming increasingly uneven. We should expect, therefore, a further widening of the income distribution over the next decade or so, other things being equal. The uneven nature of progress is affecting in particular the bottom 20 per cent of the skills distribution, and there is a clear need for policy measures to address this issue.

References

Altonji, Joseph G., Bharadwaj, Prachant and Lange, Fabian (2008) Changes in the characteristics of American youth: implications for adult outcomes, NBER working paper 13883.

Bradley, Steve, Johnes, Geraint and Millington, Jim (2001) The effect of competition on the efficiency of secondary schools in England, *European Journal of Operational Research*, 135, 545-568.

DiNardo, John, Fortin, Nicole M. and Lemieux, Thomas (1996) Labor market institutions and the distribution of wages, 1973-1992: a semi-parametric approach, *Econometrica*, 64, 1001-1044.

Epanechnikov, V.A. (1969) Nonparametric estimation of a multivariate probability density, *Theory of Probability and its Applications*, 14, 153-158.

Figure 1 Wage densities

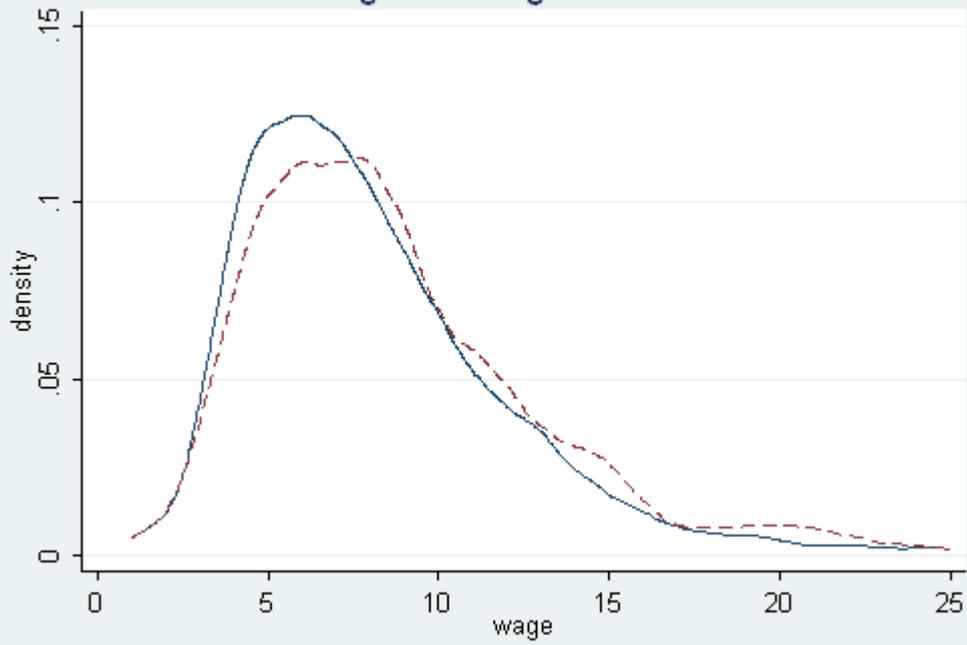
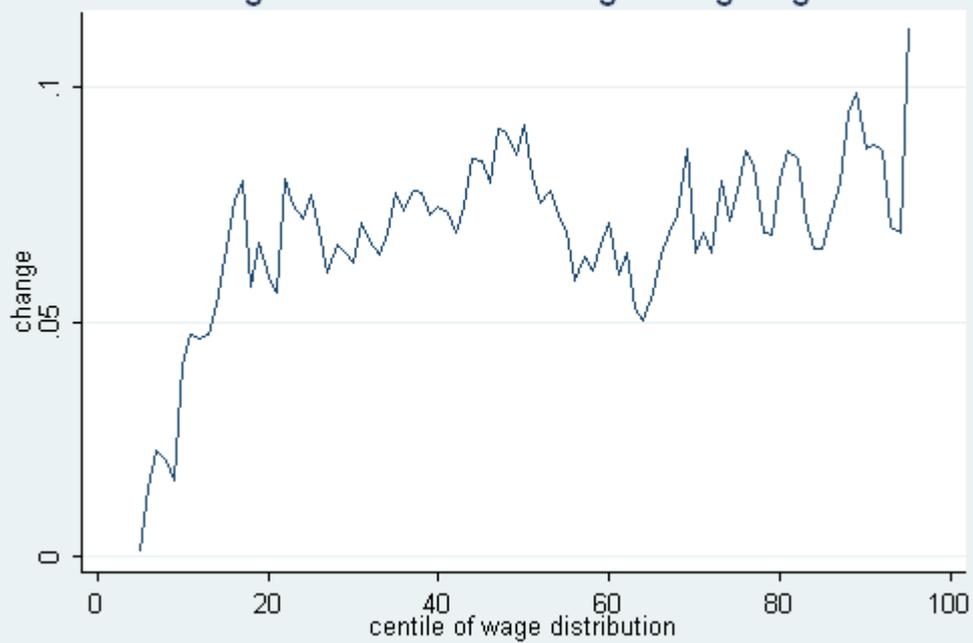


Figure 2 Predicted change in log wage



Appendix

Table A1 Probit results: dependent variable = 1 if observation is drawn from the YCS

variable	coefficient (st.err.)	variable	coefficient (st.err.)
mw	-0.189 (6.89)	fw*maed	0.728 (24.21)
mo	1.169 (12.50)	fo*maed	0.199 (1.62)
fw	0.066 (2.38)	mw*together	-0.349 (12.18)
fo	1.548 (15.06)	mo*together	-0.063 (0.63)
mw*paed	-0.021 (0.72)	fw*together	-0.533 (17.72)
mo*paed	-0.629 (5.76)	fo*together	-0.446 (3.82)
fw*paed	0.431 (14.05)	mw*olevels	0.470 (18.57)
fo*paed	0.005 (0.04)	mo*olevels	0.633 (6.08)
mw*maed	0.083 (2.88)	fw*olevels	0.336 (13.07)
mo*maed	-0.590 (5.51)	fo*olevels	0.382 (3.45)
log likelihood	-17870.797	number of observations	29557

Note: the gender and ethnicity terms are defined as follows: (i) mw is one for white males and zero for all other respondents; (ii) mo is one for other males and zero for all other respondents; (iii) fw is one for white females and zero for all other respondents; and (iv) fo is one for other females and zero for all other respondents.