The Australian National University

Centre for Economic Policy Research

DISCUSSION PAPER

Are there Asymmetries in the Effects of Training on the Conditional Male Wage Distribution?

Wiji Arulampalam, Alison L Booth

and Mark L Bryan

DISCUSSION PAPER NO. 523
May 2006

ISSN: 1442-8636
ISBN: 0 7315 3593 6

Wiji Arulampalam, University of Warwick, and IZA Bonn

Alison L Booth, Australian National University, University of Essex, CEPR and IZA Bonn

Corresponding author: Mark L Bryan, University of Essex; email markb@essex.ac.uk

Acknowledgements
For helpful comments we thank the Editor, two anonymous referees, Marco Francesconi, Amanda Gosling, Kevin Hallock, Jennifer Smith, Ian Walker, and seminar participants at the Institute for Social and Economic Research at the University of Essex and the European Association of Labour Economists Conference 2004. We are also grateful for financial support from the Leverhulme Trust under the Award F/00213/H "Training in Europe: Its Causes and Consequences".

ABSTRACT

Recent studies have used quantile regression (QR) techniques to estimate the impact of education on the location, scale and shape of the conditional wage distribution. In our paper we investigate the degree to which work-related training – another important form of human capital – affects the location, scale and shape of the conditional wage distribution. Using the first six waves of the European Community Household Panel, we utilise both ordinary least squares and QR techniques to estimate associations between work-related training and wages for private sector men in ten European Union countries. Our results show that, for the majority of countries, there is a fairly uniform association between training and hourly wages across the conditional wage distribution. However, there are considerable differences across countries in mean associations between training and wages.

JEL Classification: J24, J31, C29

Keywords: private sector training, wages, education, quantile regression, unobservables

1. INTRODUCTION

The mean returns to various forms of human capital have been extensively investigated in the labour economics literature, especially the returns to formal education and work-related training.[1] Relatively recently, attention has shifted to exploring the degree to which education might be associated with more complex changes in the conditional wage distribution[2] Arias, Hallock and Sosa-Escudero (2001), Gonzales and Miles (2001) and Martins and Pereira (2004) estimate the returns to education across the conditional wage distribution using quantile regression (QR) techniques. Martins and Pereira (2004) use cross-sectional data from a variety of different data sources covering 15 European countries plus the USA and find that "returns to schooling increase over the wage distribution". Martins and Pereira, as well as Arias *et al.*, point out the implications of these results, that increased education may be associated with a widening of the (conditional) wage distribution, and may not always improve the prospects of low-earning workers as much as hoped by policy makers.

Our purpose in the present paper is to see if there is also an upward sloping profile for training across the conditional wages distribution. Education and work-related training are complementary, as numerous studies attest (see Arulampalam, Booth and Bryan, 2004 and references therein). Hence we might expect to observe an upward sloping profile when we graph the training association across quantiles of the conditional wage distribution.[3] We also wish to document how observed training associations differ across the European Union

---

1 For surveys of articles estimating the mean returns to training, see Ashenfelter and Lalonde (1996) and for the mean returns to education, see Card (1999) .

2 While ordinary least squares (OLS) techniques allow one to estimate the association between the regressors and the conditional mean of the distribution, quantile regression (QR) methods allow the regressors to be associated with change to the scale and shape as well.

3 Although there has been a recent surge in the estimation of wage equations using quantile regression techniques (see Fitzenberger, *et al*, 2001, for some applications), to our knowledge there are no studies investigating the association between work-related training and the conditional wage distribution.

(EU) countries for which we have harmonised data.[4] Using the first six waves of the European Community Household Panel (ECHP), we carry out this analysis for private sector men in ten European Union countries.

In the next section we describe our data source, estimating sub-samples, and the principal variables used in our analysis. In Section 3 we outline the econometric model, while in Section 4 we first present the OLS results and then the QR estimates. The final section draws some conclusions.

2. THE DATA AND EXPLANATORY VARIABLES

2.1. The Data Source and the Sample

Our data are from the first six waves of the European Community Household Panel (ECHP), a large-scale survey collected annually since 1994 in a standardised format that facilitates cross-country comparisons. The ECHP was specifically designed to be harmonised at the input stage: in most countries a standard questionnaire was used, with harmonised definitions and sampling criteria. Although a standardised questionnaire does not overcome the nuances of interpretation and meaning between different languages, the harmonised format greatly facilitates cross-country comparisons.

We include in our analysis the ten European countries listed in Table 1. We have only five waves for Austria and four waves for Finland, as they joined the ECHP after 1994. For Britain we use only the first five waves because the format of the training question altered from 1998 onwards (for further details, see Booth and Bryan, 2004). We omit Greece and Portugal from our estimation owing to apparent gaps in the training data and because of

---

4 There is an extensive literature on the evaluation of particular labour market programmes, using a variety of techniques. For example, Heckman *et al.* (1994) estimate the average effects of training on the treated, Heckman *et al* (1997) look at the distribution of treatment effects using a non instrumental variable (IV) framework, and Abadie *et al* (2002) examine the training effect on different quantiles of the wage distribution using the IV framework. Our interest here is in work-related training and not in a labour market program. Since we do not have a suitable instrument, we do not treat training as endogenous.

the smaller estimating sub-samples with usable information. We also omit Germany because the training variables are not comparable to the other countries.[5] The ECHP data for Britain were adapted from existing national household surveys, while the other countries used the full harmonised questionnaire. Sample sizes are reported in Column [1] of Table 2 and in Column [7] of the Data Appendix Table.

In earlier work using the ECHP, we found that training incidence is typically significantly higher in the EU public sector than the private (Arulampalam, Booth and Bryan, 2004). This finding came as no surprise, since private sector firms are more likely than the public sector to be constrained by the need to make profits, and so they may be less willing to finance training through fears of losing trained workers to rival non-training firms. Our preliminary testing showed that it is inappropriate to pool private and public sector workers, since the coefficients across the sub-groups differ significantly, as might be expected given that public and private sector employers typically have different objective functions. We therefore only focus on the private sector in this study. We also consider only men, although in a separate study we investigate the gender wage gap using QR techniques and the ECHP data (Arulampalam, Booth and Bryan, 2005).

It is well-known that EU countries differ with regard to their vocational training and education systems. We wish to avoid conflating work-related training with initial vocational education or training. We therefore exclude from our analysis individuals under the age of 25 years, paid apprentices, and those on special employment-related training schemes.[6] We also exclude workers aged 55 years or more. This is because, among older workers, there

[5] The ECHP includes two datasets for Germany: the six-wave dataset (derived from the GSOEP survey), which excludes many shorter training spells (communication from DIW), and the original three-wave dataset. In the three-wave dataset, interview dates are treated as confidential, so it is not possible to construct job tenure or know whether training was before or after the previous interview.

[6] Apprentices and those on special training schemes account for only 1.1% of the sampled age group.

may be differential withdrawal from the labour force depending on, for example, how early retirement schemes operate.

For each country, our estimating sub-sample therefore comprises employed private sector men who are: (i) between the ages of 25 and 54 years and working at least 15 hours per week; (ii) not employed in agriculture; (iii) with valid observations for the principal variables used in the wage equations;[7] and (iv) with sequences of continuous observations starting from the first wave in the sample in order to have a complete training record (see also Data Appendix). Individuals can be present for a minimum of two waves (including the first wave) and a maximum of six waves for all countries except for Austria and Britain (where the maximum is five) and Finland (where the maximum is four).[8]

The restriction of working at least 15 hours per week was necessary because of the nature of the ECHP data, where – in the first two waves – we were unable to distinguish individuals regularly working fewer than 15 hours from those out-of-the labour force. In addition, some important variables like firm size and tenure are only available for individuals working 15 hours or more. Thus our estimating sub-samples will under-represent low-hours part-timers (though for most countries these represent only a tiny fraction of male workers).[9]

We include in our analysis the ten European countries listed in Table 1 and estimate the models using pooled person-year observations. Because of the definition of the training variable, individuals stay in the sample continuously until they fail to give an interview, which results in an unbalanced panel with different individuals contributing different numbers of observations.

---

[7] We construct some missing industry dummy variables as detailed in Column [5] in the Data Appendix Table. For example, a number of cases had missing information for industrial classification. Rather than drop these cases from estimation, we constructed a dummy variable "missing industry" and included this as a control in the regression.

[8] As explained in Section 3, since we require a complete record of training for each individual, we drop any observations, which follow a break in the data. Therefore, if an individual is observed in waves 1, 2, 3, 5 and 6, we use waves 1, 2 and 3 only.

[9] Exceptions are Britain (6.2% of the sub-sample), the Netherlands (8.8%) and Ireland (4.0%). In all other countries the proportion of low-hours part-timers is under 3%.

2.2. Hourly Wages

The dependent variable is the log of the average hourly wage, including overtime payments, in the respondent's main job.[10] The characteristics of each country's unconditional log wage distribution, deflated to 1999 prices, are reported in Table 1. The deflators are the European Union's harmonised indices of consumer prices (HICP; see *Eurostat Yearbook*, 2002). To allow cross-country comparisons of consumption wages, the log wage figures were converted to purchasing power parity (PPP) units, using the scaling factors supplied with the ECHP. The first column shows substantial variation in mean wages across countries, from a high of 2.77 log points in the Netherlands down to 2.15 log points in Spain (with 2.18 log points in Britain). But there are also differences in the dispersion of wages, as shown by the standard deviations in the second column. By this measure, the country with the lowest dispersion (0.30) is Denmark, while Ireland has the highest dispersion (0.53). It is notable that Spain has the lowest mean and one of the highest standard deviations (0.50). The remaining columns show the median, the 10th and 90th percentiles, and in the last column the difference between the 90th and the 10th percentiles. This measure of dispersion shows a similar pattern to the standard deviation: Spain, Britain, Ireland and France stand out as countries with high hourly wage dispersion.

2.3. The Training Variable

The form of the training question, harmonised across countries in the ECHP, is as follows: "Have you at any time since January (in the previous year) been in vocational education or training, including any part-time or short courses?" Although separate training courses within the reference period are not identified, respondents are asked for the *overall* duration and start/end dates of the training. Since the reference period may overlap with the reference

---

[10] The log wage was calculated from the ECHP variables as log (wage) = log (PI211MG * (12/52) / PE005A) = log (normal gross monthly earnings from main job including overtime * (12/52) / hours in main job including overtime). No specific information is provided on overtime hours and premia.

period of the previous wave, to avoid double counting, where possible we use the start and end dates to identify training events specific to each wave.[11] We then construct our training variable D_{it} as the cumulative count of *completed* events since the first wave of the sample. Most studies simply examine the impact of training incidence (and sometimes intensity) on wages, but not the number of events.[12] We follow Lillard and Tan (1992) in using the accumulated sum of all training events, where there is only one event measured at each wave owing to the nature of our data.

The framing of the training question suggests that the training responses should be interpreted as more formal courses of instruction, rather than informal on-the-job training (for which we control – at least in part – using job tenure). A separate question asks about "general or higher education". Participation in these more general courses is very low (average annual take-up by 25-54 year olds is less than 1%) so we are confident that our results are not affected by interactions with countries' differing formal educational systems.

Our measure of work-related training is based on a harmonised questionnaire and there are two additional reasons why it is likely to be robust across countries. First, there is typically much less regulation of work-related training than initial training and education. Second, as noted above the incidence of general education after age 25 is very low (typically less than 2%), so there is little danger of confusing training and education.

Table 2 reports information about completed training courses for private sector men by country. The first column gives the number of observations for each country, while the

[11] The modal interview month is October, corresponding to a reference period of 22 months. The British data do not include training dates. However they are derived from the British Household Panel Survey (BHPS), where the reference period only slightly exceeds one year. Since events are generally very short in Britain, there should be little chance of double counting. For France, we do not use training dates as they are missing for the majority of events. For the Netherlands, the end dates of training are not available so we use start dates only to identify events begun since the previous interview.

[12] Exceptions are Lillard and Tan (1992), Arulampalam, Booth and Elias, (1997), Blundell et al. (1999) Arulampalam and Booth (2001), and Booth and Bryan (2002). Lillard and Tan (1992: p31) note that multiple training occurrences within a period are typically not known from US survey data.

second column reports the mean number of waves for each country.[13] The third column reports training incidence for completed courses only. For example, the first row of Table 2 shows that the Austrian sub-sample comprises 786 private sector men who are observed in three waves on average and of whom 15% have completed a training course in any year. The mean accumulated training count is simply the product of the second and third columns. The figures in the third column show that training incidence differs considerably across countries. We can identify three high-incidence countries – Britain, Denmark and Finland – where each year around 30% or more of individuals complete training courses. In contrast Austria, Belgium and France form a group of medium-incidence countries, where each year between 10% and 15% of men complete training courses. Finally, Ireland, Italy, the Netherlands and Spain have incidence below 10%.

Though our sample is limited to men in the private sector, the cross-country pattern summarised in Table 2 is similar to that found in analysis of overall training (Arulampalam *et al.*, 2004). The ranking also compares reasonably well (especially for the high incidence countries) with the cross-country comparisons using different data sources reported in OECD (1999); and with International Adult Literacy Survey (IALS) data on continuing training for several countries featured in OECD (2003).

2.4. Other Explanatory Variables

The education, industry and occupation variables in the ECHP are all coded according to standard, internationally comparable definitions. Education levels are defined according to UNESCO's International Standard Classification of Education (ISCED). ISCED was intended for education policy analysis and was designed to be invariant to differences in national education systems.[14] The ECHP distinguishes between education completed to the

---

[13] As explained earlier, individuals can be present for a minimum of two waves and a maximum of six waves for all countries except for Austria and Britain (where the maximum is five) and Finland (where the maximum is four).

[14] For details, see http://www.unesco.org/education/information/nfsunesco/doc/isced_1997.htm.

lower secondary stage (ISCED 0-2), upper secondary education (ISCED 3) and post-secondary or tertiary education (ISCED 5-7). ISCED 0-2 forms the base or omitted dummy variable in our regression results reported in Section 4.

The data on industrial sector are categorised according to the European Union's Classification of Economic Activities in the European Community (NACE), while occupation is defined using the International Standard Classification of Occupation (ISCO-88).

The other controls are demographic attributes and job characteristics expected to affect earnings. We include dummy variables for age and job tenure bands, any unemployment experienced since 1989, marital status, health problems affecting daily life, highest educational levels, fixed term or casual employment, part-time work, establishment size, one-digit occupation and industry, year and, where the data allow, region. We also include a separate control for training started in the current year but uncompleted at the survey date. Where there were non-trivial numbers of missing observations for variables like industry and region, we include these cases in the regression but control for the missing values using dummy variables (see Data Appendix for details).

3. THE ECONOMETRIC MODEL

There is an extensive literature that estimates the impact of training on expected wages using a linear regression framework (see *inter alia* references in Ashenfelter and Lalonde, 1996; and Arulampalam and Booth, 2001). Here, we deviate from this common practice by looking at the associations of training and other covariates with wages at different quantiles of the log wage distribution.[15] The main advantage of a quantile regression (QR) framework is that it enables one to model the effects of the covariates on the location, scale and shape of the

[15] The linear conditional quantile regression model was first introduced by Koenker and Bassett (1978). For a recent survey of these models, see Buchinsky (1998), Koenker (2005).

conditional wage distribution, unlike the linear regression model (least squares) that only allows one to look at the effect on the location (the conditional mean).

We specify the θth $(0<\theta<1)$[16] conditional quantile of the log wage (w) distribution for the i-th individual ($i=1,.., n$) in wave t ($t=2,..,T_i$) as

$$\text{Quant}_\theta(w_{it}|\mathbf{x}_{it}, \mathrm{D}_{it}, e_{i1}) = \alpha(\theta) + \mathbf{x}_{it}'\beta(\theta) + \mathrm{D}_{it}\,\gamma(\theta) + e_{i1}\delta(\theta) \quad (1)$$

implying

$$w_{it} = \alpha(\theta) + \mathbf{x}_{it}'\beta(\theta) + \mathrm{D}_{it}\,\gamma(\theta) + e_{i1}\delta(\theta) + u_{\theta it} \quad (2)$$

with $\text{Quant}_\theta(u_{\theta it}|\,\mathbf{x}_{it}, \mathrm{D}_{it}, e_{i1}) = 0$.

In the specifications of equations (1) and (2), D_{it} is defined as the cumulative count of *completed* events since the first wave. Therefore D_{it} increases by one for every year that the individual is in receipt of training between two waves. The vector $\mathbf{x}_{it}$ contains the human capital and job characteristics listed above.

Training can have cumulative effect on wages and our training dummy D_{it} enables us to allow for this. Unfortunately, we do not have information on the complete history of training received by individuals in the sample. In panel data models estimating the effect of training on the conditional mean wage, it is customary either to first difference the equation prior to estimation or to use within-group deviations to account for individual specific unobservables, which also allows one to control for training effects prior to the start of the spell. In order to account for initial unobservables in our QR model, we have included the residual (e_{i1}) from an OLS wage regression estimated at wave 1 (just prior to our observation period). The purpose of this wage residual is to control for the role of unobservable skills (such as previous training) acquired before the sample period. We do not work with the first differenced equation here since our interest is in charting the effects of training on different parts of the conditional wage distribution, and not the effects on annual wages growth.

[16] θ=0.5 refers to the Median.

Moreover, the first differenced equation is hard to interpret in the QR case because the difference of the wage quantile is not the same as the quantile of the differenced wage. Hence, with our QR model in levels, we interpret our results as the effects of training on the distribution of wages, conditional on initial unobservables.

Note that, if the underlying model were truly a location model - in the sense that the changes in explanatory variables causing only a change in the location of the distribution of *w* and not in the shape of the distribution - then all the slope coefficients would be the same for all θ.[17] We use Stata 8 to estimate the coefficients of our QR model. To account for the use of repeated observations on individuals, the standard errors are calculated using a block-sampling method involving 500 replications (Fitzenberger, 1998; Fitzenberger and Kurz, 2003). As a benchmark for our QR results, we also present OLS estimates of the wage equation in Section 4.1 below before discussing the QR estimates in Section 4.2. The conditional mean model estimated using OLS also includes our residual (e_{i1}) from the wave 1 wage equation.

4. RESULTS

4.1. OLS Estimates

The first column of Table 3 reports the OLS estimates of accumulated work-related training events D_{it} and the two highest educational qualification dummy variables, with the base for education being lower secondary education (ISCED 0-2). Other controls included in the estimation are listed in the notes under the table. As explained above, we define a training event as a wave in which training was received, and these are summed across waves for each observation.

[17] Quantile regression models are more general than simple linear regression model allowing for heteroskedastic errors, since the QR model allows for more general dependence of the distribution of *w* (the dependent variable) on the **x**s instead of just the mean and the variance alone.

Table 3 reveals an interesting pattern of partial correlations between training and wages. The highest estimated association between wages and each training event is 8.8% in Ireland, while for some countries the coefficients are not statistically different from zero (Belgium, Italy and the Netherlands). Significant associations of about 3-5% higher wages per event are found for Austria, Finland, France, and Spain. A small statistically significant association of 0.7% per training event was found in Denmark. Note that two countries with the highest training incidence – Britain and Denmark– are also amongst the countries with lower training-wage associations, of approximately one percent per event for Denmark and just under two percent per event for Britain. It is interesting to compare our estimates with those of Bassanini et al (2005), who pooled all countries in the ECHP dataset and who included as controls country dummies, gender, age, age-squared, year, education, marital status and industrial dummies. In our specifications we included more explanatory variables (in particular more job characteristics) and allowed the impact of these to vary across countries. This has resulted in a much smaller estimated training effect than in Bassanini et al (2005). Of course our sample is also different, since we focus only on private sector men and so our estimates are not strictly speaking comparable.[18]

In summary, our OLS estimates reveal that there are quite considerable differences in *ceteris paribus* associations between wages and training across our sample of EU countries. These OLS results highlight the potential importance of cross-country heterogeneity in employer-provided and vocational training systems, discussed in Lynch (1994) and Booth and Snower (1996), *inter alia*.[19] We next turn to investigating whether there are *intra-*

---

[18] In an interesting paper also using the ECHP, Bassanini and Brunello (2003) utilise data for 7 countries from the 1996 wave to explore the degree to which training incidence is correlated with wage compression. Our estimates are not directly comparable with theirs, since they pool cross-country data and partition full-time men aged 30-60 years into clusters (by country, education level, occupation and sector) in order to test the degree to which cluster-specific measures of the training wage premium are correlated with general training incidence.

[19] There is very little comparative work investigating the extent and economic impact of work-related training, in part because harmonised data facilitating comparisons became available only recently (OECD, 1999).

country differences in such associations across the conditional wage distributions, and whether or not these intra-country differences vary across countries.

4.2. QR Estimates of Work-related Training

Table 4 shows the quantile regression estimates of the training effects (the partial derivative of the conditional quantile or mean wage with respect to training), as measured by the coefficient on the training receipt dummy, for five different values of θ (0.1, 0.25, 0.5, 0.75 and 0.90). We also control for residuals from a wave 1 wage regression, as described in Section 3. The coefficients on the residuals (not reported) are highly significant in all equations (with typical coefficients in the range 0.4–0.7), suggesting that it is important to account for unobserved skills acquired before the observation period. Figure 1 also presents the estimated effects for each of the quantiles of the log wage distribution, along with the 95% confidence band around the estimates. Superimposed on the plots is a dotted horizontal line representing the OLS estimates of the effect of training on expected log wages. If the QR estimates of the association of training with wages are the same across the conditional wage distribution, the implication is that training only affects the location of this conditional wage distribution.

Inspection of Figure 1 reveals first, that the QR estimates are fairly uniform across the conditional wages distributions in five countries and there are noticeable slopes in the other five countries: Belgium, France and Ireland (downward sloping), and Britain and Denmark (upward sloping). For example, in Britain the QR estimates range from 1.7% at the 10^{th} percentile to 2.6% at the 90^{th} percentile (see also Table 3), while the OLS estimate of the training-wages association is 1.9%, as already reported. Differences in the training coefficients across quantiles suggest that training may be associated with expanded or compressed conditional wage distributions. The larger coefficients at higher quantiles in Britain and Denmark indicate that training is associated with increased dispersion of the conditional wage distribution, *ceteris paribus*. The reverse result is found for Belgium, France and Ireland, where the estimates are smaller at the higher quantiles. Effectively this suggests that training is associated with a reduced dispersion of the conditional wage distribution in these countries, *ceteris paribus*.

However, a second feature of Figure 1 is that the confidence bands are quite wide, and the OLS estimates lie within the QR confidence bands for all of our countries. This is evidence against significant differences in the estimates across the wages distribution.[20] Furthermore, inspection of Table 3 reveals that, for four countries, there are no statistically significant associations between training and wages across the entire wages distribution. This is the case for Belgium, Italy and the Netherlands (whose OLS coefficients are also not statistically different from zero) and also for Denmark.

In summary, we find that the training effect is fairly uniform across the conditional wage distribution within a country. Moreover, this finding is repeated for the vast majority of the EU countries we investigated. This is an interesting result, and one that is counter to the results found in other studies for another important form of human capital, education. However, our results do suggest that there are considerable differences in *mean returns* to training across countries.

So far we have focussed on the training associations, but it is also interesting to examine the coefficients associated with the controls for upper secondary and tertiary education. These are also presented in Table 3. The estimates reveal that only for Denmark is the association between upper secondary education and wages increasing across the wages distribution. By contrast, relative to the base of lower secondary education, the association between tertiary education and wages is clearly increasing across the log wage distribution in

[20] Note that this does not mean that the association between training and wages is best quantified by OLS, since with OLS the effects of *all* covariates are assumed to have only location shifts.

Austria, Belgium, Denmark, France, the Netherlands and Spain. In two other countries – Ireland and Italy – the association between tertiary education and wages is essentially flat over the distribution, while it actually declines in Britain and Finland.[21]

5. CONCLUSIONS

In this paper we used quantile regression techniques to investigate the degree to which work-related training affects the location, scale and shape of the conditional wage distribution. Using the first six waves of the European Community Household Panel, we investigated these issues for private sector men in ten European Union countries. Our results for training suggest that, for the majority of countries, associations between training and wages are similar across the conditional wage distribution. We also controlled for highest educational qualification, using harmonised measures. Consistent with the results of earlier papers, we found that the association between post-secondary education and wages is increasing across the wages distribution.

Overall, our results provide support to previous findings that education is associated with increased dispersion of the conditional wage distribution, although the effect appears to operate through tertiary education rather than upper-secondary education. By contrast, while finding positive associations between training and wages, we did not observe an upward sloping profile across quantiles of the conditional wage distribution. This suggests that there may be different forces at work in the relationship between training and wages. For example,

[21] Martins and Pereira (2004), using different non-harmonised data sets and employing years of schooling as their measure of education, found that returns to schooling increased over the wage distribution for their 16 different countries. Our results for tertiary education provide some support for their findings, although our specification includes many more explanatory variables along with a control for pre-sample unobservables. As is typical for studies that estimate the returns to schooling, Martins and Pereira used very few controls (experience and its square). Such a parsimonious specification is inappropriate for our purposes, where the focus is on training received after entering the labour market. Their estimating samples also included men in both public and private sectors aged 15-65 working at least 35 hours per week.

there could be unobserved heterogeneity with regard to training content or training costs.[22] However, since such explanations can only be speculative in this context, we do not pursue them further here.

Finally, our OLS results were also of interest in their own right. We found considerable cross-country differences in *mean associations between* training and wages. The highest was found for Ireland, with around 9% higher wages per event. The smallest associations were found in Belgium, Italy and the Netherlands, where the associations were statistically insignificant. Significant associations of about 3-5% higher wages per event were found for Austria, Finland, France, and Spain. Note that two of the countries with the highest training incidence – Britain and Denmark – are also amongst the countries with the lowest wage associations, of approximately one percent per event for Denmark and just under two percent per event for Britain.

[22] For example Almeida and Carneiro (2006) find, using data for Portugal, that direct costs represent the bulk of training costs and that foregone productivity accounts for less than 25% of the total costs of training. They suggest that the estimated coefficient to training in a wage equation is therefore unlikely to be an estimate of the true return to training. Such direct training costs are unobservable in the ECHP survey data.

References

Abadie, A., Angrist, J. and Imbens, G. (2002) “Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings’, *Econometrica*, 70(1), p.91-117.

Almeida R and P Carneiro (2006). ‘The Return to the Firm Investment in Human Capital.’ IZA Discussion Paper No. 1937, January.

Arias, O., Hallock, K. and Sosa-Escudero, W. (2001) ‘Individual heterogeneity in the returns ot schooling: instrumental variables quantile regression using twins data’, in Fitzenberger, B., Koenker, R., and Machado, J.A.F. (Eds.) (2001) *Economic Applications of Quantile Regression*, Physica-Verlag.

Arulampalam, W. and Booth, A.L. (2001) ‘Learning and earning: do multiple training events pay? A decade of evidence from a cohort of young British men’, *Economica*, 68, 379-400.

Arulampalam, W., Booth, A.L. and Bryan, M.L. (2004) ‘Training in Europe’, *Journal of the European Economic Association*, Volume 2, Issue 2/3, 346-360.

Arulampalam, W., Booth, A.L. and Bryan, M.L. (2005) ‘Is There a Glass Ceiling over Europe? An Exploration of Asymmetries in the Gender Pay Gap across the Wages Distribution’. Revised version of IZA Discussion Paper No. 1373 (April 2006).

Arulampalam, W., Booth, A.L. and Elias, P. (1997) ‘Work-related training and earnings growth for young men in Britain’, *Research in Labor Economics*, 16, 119-47.

Ashenfelter, O. and Lalonde, R.J. (eds) (1996) *The Economics of Training*, Edward Elgar.

Bassanini, A., A.L. Booth, G. Brunello, M. De Paola, and E. Leuven (2005). ‘Workplace Training in Europe’. IZA Discussion Paper No.1640, June. Forthcoming as Part II of *Education and Training in Europe*, Oxford University Press.

Bassanini, A. and G. Brunello (2003). “Is Training More Frequent When Wage Compression is Higher? Evidence from the European Community Household Panel”. IZA Discussion Paper No.839, June.

Blundell, R., Dearden, L. and Meghir, C. (1999), *The determinants and effects of work-related training in Britain*. London: Institute of Fiscal Studies.

Booth, A.L. and Bryan, M.L. (2004) ‘Who Pays for General Training in Private Sector Britain?”, mimeo, ISER, University of Essex. Forthcoming 2006 in *Research in Labour Economics.*

Booth AL and DJ Snower (1996) eds. *Acquiring Skills: Market Failures, their Symptoms, and Policy Responses.* Cambridge: Cambridge University Press.

Buchinsky, M. (1998) ‘Recent advances in quantile regression models’, *Journal of Human Resources*, 33, 88-126.

Buchinsky, M (2001) ‘Quantile regression with sample selection: estimating women’s return to education in the U.S.’, in Fitzenberger, B., Koenker, R., and Machado, J.A.F. (Eds.) (2001) *Economic Applications of Quantile Regression*, Physica-Verlag.

Card, D. (1999) ‘The causal effect of education on earnings’, in Ashenfelter. O. and Card, D *Handbook of Labor Economics, Volume 3A.* Elsevier, Amsterdam.

Fitzenberer, B. (1998) ‘The moving blocks bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics*, 82, 235–87.

Fitzenberger, B., Koenker, R., and Machado, J.A.F. (Eds.) (2001) *Economic Applications of Quantile Regression*, Physica-Verlag.

Fitzenberger, B. and Kurz, C. (2003) ‘New insights on earnings trends across skill groups and industries in West Germany’, *Empirical Economics*, 28, 479-514.

Garcia, J., Hernandez, P. J., and Lopez-Nicolas, A. (2001) ‘How wide is the gap? An investigation of gender wage differences using quantile regression’, in Fitzenberger *et al* (Eds.) (2001) *Economic Applications of Quantile Regression,* Physica-Verlag.

Gonzales, X. and Miles, D. (2001), ‘Wage inequality in a developing country: decrease in minimum wage or increase in education returns’ in Fitzenberger *et al* (Eds.) (2001) *Economic Applications of Quantile Regression*, Physica-Verlag.

Heckman J., Smith, J. and Clements, N. (1997) ‘Making the most out of social experiments: accounting for heterogeneity in programme impacts’, *Review of Economic Studies*, 64, p.487-535.

Heckman, J., Smith, J. and Taber, C. (1994) ‘Accounting for dropouts in evaluations of social experiments’, NBER Technical working paper no. 166.

Koenker, R (2005) *Quantile Regression*, Cambridge University Press.

Koenker, R. and Bassett, G. (1978) ‘Regression quantiles’, *Econometrica*, 46, 33-50.

Lillard, L.A. and Tan, H.W. (1992) ‘Private sector training: who gets it and what are its effects?’, *Research in Labor Economics*, 13, 1-62.

Lynch, LM (1994) ed. *Training and the Private Sector: Intenational Comparisons.* University of Chicago Press.

Martins, PS and PT Pereira (2004). ‘Does education reduce wage inequality? Quantile regression evidence from 16 countries.’ *Labour Economics*, 11(3), 355-371, June.

OECD. (1999). ‘Training of Adult Workers in OECD Countries: Measurement and Analysis.’ Chapter 3 in OECD *Employment Outlook*, 134-175, Paris: OECD.

OECD (2003) ‘Upgrading workers’ skills and competencies’, Chapter 5 in OECD *Employment Outlook*, 237-296, Paris: OECD.

Figure 1: the association of training and wages across the wage distribution

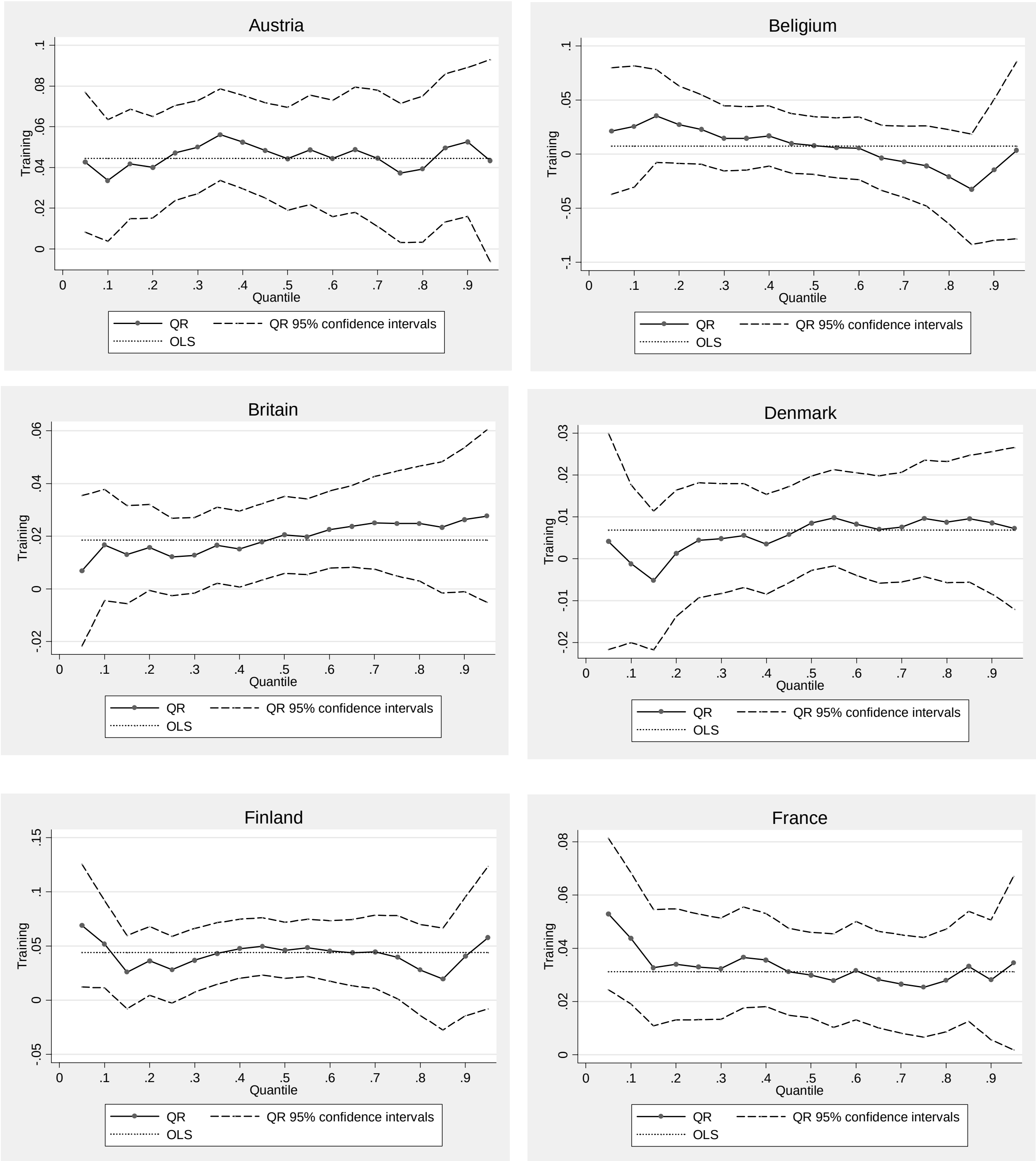

Figure 1 contd: the association of training and wages across the wage distribution

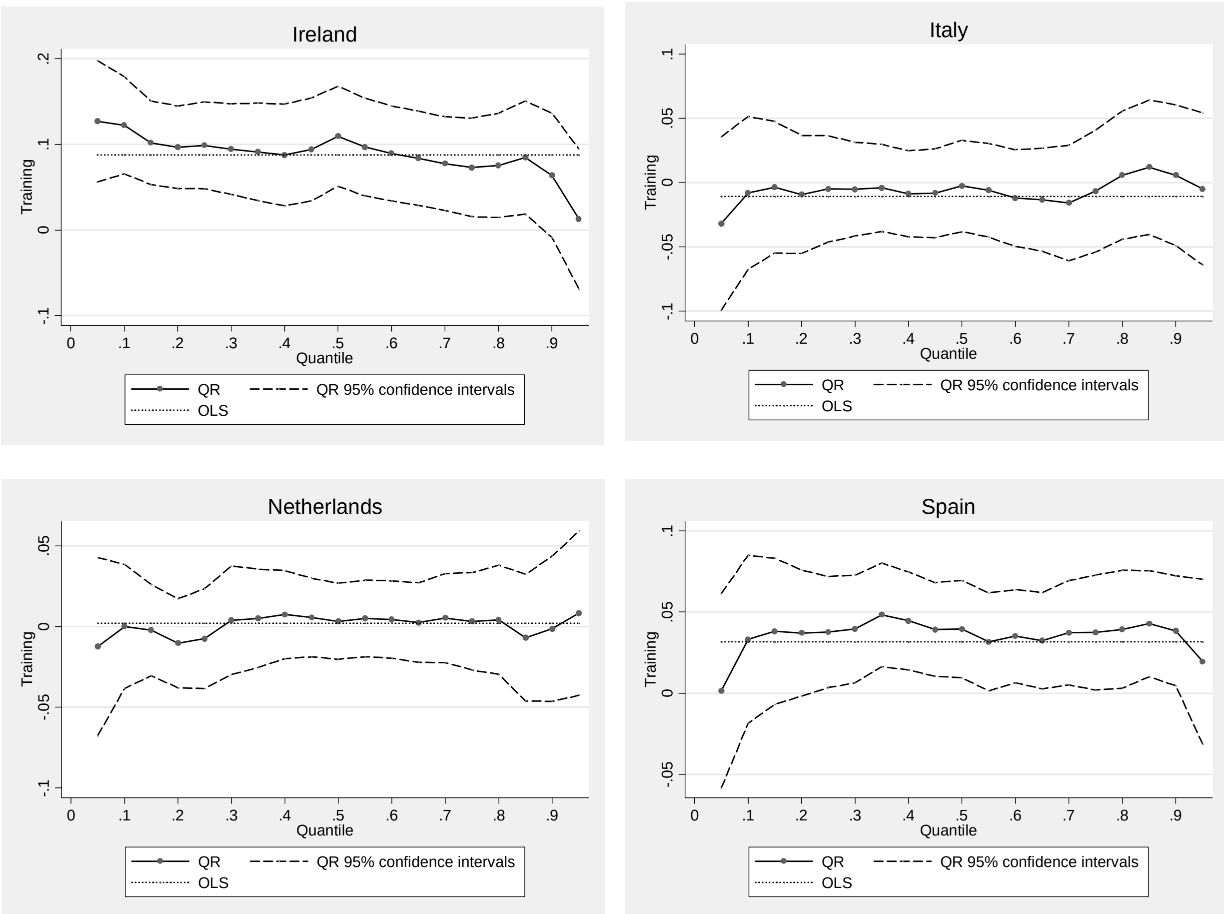

Table 1: Log hourly wage distributions in purchasing power parity (PPP) units

	Mean [1]	St dev [2]	Median [3]	10th percentile [4]	90th percentile [5]	90-10 differential [6]
Austria	2.348	0.385	2.311	1.947	2.807	0.859
Belgium	2.497	0.353	2.444	2.107	2.914	0.806
Britain	2.180	0.465	2.502	1.964	3.108	1.144
Denmark	2.743	0.302	2.715	2.397	3.180	0.783
Finland	2.342	0.397	2.298	1.918	2.848	0.931
France	2.340	0.450	2.277	1.862	2.925	1.063
Ireland	2.380	0.532	2.380	1.833	2.966	1.133
Italy	2.234	0.335	2.204	1.874	2.618	0.744
Netherlands	2.766	0.396	2.744	2.421	3.219	0.798
Spain	2.154	0.505	2.104	1.580	2.855	1.275

Notes: The log wage was calculated from the ECHP variables as log (wage) = log (PI211MG * (12/52) / PE005A) = log (normal gross monthly earnings from main job including overtime * (12/52) / hours in main job including overtime). It was then deflated to 1999 prices using harmonised indices of consumer prices (HICP) from the Eurostat Yearbook 2002, and converted to purchasing power parity (PPP) units using the ECHP variable PPPxx (where xx is the year). The above statistics are based on the intercept estimate in the regression of log wage on a set of wave dummies. The omitted category was wave 5. The reported numbers in [2] refer to the regression standard error.

Table 2: Training Participation across Europe for Private Sector Men in Employment Aged 25-54 Years

	Number of men observed [1]	Mean number of observed waves [2]	Annual training incidence (completed) [3]	Mean accumulated training count [4]
Austria	786	3.01	0.15	0.45
Belgium	492	3.10	0.10	0.31
Britain	986	3.41	0.39	1.33
Denmark	626	3.44	0.37	1.27
Finland	740	2.40	0.29	0.70
France	1448	3.41	0.11	0.38
Ireland	544	3.26	0.05	0.16
Italy	1092	3.44	0.03	0.10
Netherlands	908	3.93	0.05	0.20
Spain	1204	3.35	0.08	0.27

Note: column [3] reports the average proportion of men who have completed a training course since the previous interview; column [4] indicates the mean number of courses completed over the panel (which equals the product of columns [2] and [3]).

Table 3 – Wage Effects of Training and Education

	OLS	10%	25%	50%	75%	90%
Austria						
Training	**0.045**	**0.034**	**0.047**	**0.044**	**0.037**	**0.053**
Upper secondary educ	**0.073**	0.041	**0.084**	**0.079**	**0.097**	***0.064***
Tertiary educ	**0.170**	**0.130**	**0.173**	**0.156**	**0.239**	**0.243**
Belgium						
Training	0.007	0.026	0.023	0.008	-0.011	-0.015
Upper secondary educ	**0.030**	-0.007	0.022	0.019	0.036	***0.072***
Tertiary educ	**0.152**	**0.133**	**0.129**	**0.142**	**0.138**	**0.177**
Britain						
Training	**0.019**	0.017	0.012	**0.021**	**0.025**	***0.026***
Upper secondary educ	**0.037**	**0.057**	**0.063**	**0.064**	**0.048**	0.031
Tertiary educ	**0.116**	**0.130**	**0.138**	**0.132**	**0.116**	**0.116**
Denmark						
Training	***0.007***	-0.001	0.004	0.009	0.010	0.009
Upper secondary educ	**0.072**	**0.054**	**0.066**	**0.068**	**0.109**	**0.112**
Tertiary educ	**0.129**	**0.108**	**0.113**	**0.117**	**0.156**	**0.198**
Finland						
Training	**0.044**	**0.052**	***0.028***	**0.046**	**0.040**	0.040
Upper secondary educ	**0.057**	**0.093**	**0.062**	0.020	0.012	***0.055***
Tertiary educ	**0.137**	**0.161**	**0.119**	**0.095**	**0.092**	**0.137**
France						
Training	**0.031**	**0.044**	**0.033**	**0.030**	**0.025**	**0.028**
Upper secondary educ	**0.106**	**0.114**	**0.112**	**0.105**	**0.105**	**0.118**
Tertiary educ	**0.268**	**0.261**	**0.251**	**0.243**	**0.265**	**0.289**
Ireland						
Training	**0.088**	**0.122**	**0.099**	**0.109**	**0.073**	***0.064***
Upper secondary educ	**0.102**	**0.093**	**0.078**	**0.111**	**0.132**	**0.103**
Tertiary educ	**0.215**	**0.214**	**0.207**	**0.230**	**0.236**	**0.208**
Italy						
Training	-0.011	-0.008	-0.005	-0.003	-0.006	0.006
Upper secondary educ	**0.037**	**0.056**	**0.055**	**0.046**	**0.043**	0.016
Tertiary educ	**0.266**	**0.290**	**0.297**	**0.257**	**0.238**	**0.254**
Netherlands						
Training	0.002	0.000	-0.007	0.003	0.003	-0.001
Upper secondary educ	**0.046**	0.014	***0.024***	**0.033**	0.024	***0.040***
Tertiary educ	**0.297**	**0.224**	**0.230**	**0.254**	**0.258**	**0.292**
Spain						
Training	**0.032**	0.033	**0.038**	**0.040**	**0.037**	**0.039**
Upper secondary educ	**0.088**	**0.085**	**0.069**	**0.082**	**0.081**	**0.094**
Tertiary educ	**0.205**	**0.187**	**0.173**	**0.211**	**0.202**	**0.215**

Notes: (i) Bold (bold and italics) type denotes significance at 5% (5-10%) level at least. (ii) Upper secondary education corresponds to International Standard Classification of Education (ISCED) level 3. Tertiary education corresponds to ISCED levels 5-7. The base for education is ISCED 0-2, as noted in the text. (iii) Other controls are dummies for age and job tenure bands, training started in the current year but uncompleted at the survey date, unemployment since 1989, marital status, health problems affecting daily life, fixed term or casual employment, part-time work, establishment size, one-digit occupation and industry, year and, where available, region. The residual from wave 1 regression is also included in all the models estimated see Section 3 for further details.

DATA APPENDIX: Selection of estimating samples

Unless otherwise stated, we applied the initial selection described in Section 2 of the text. We then dropped observations with missing or invalid data on the variables in the wage equations, that is principally: training, fixed term or casual contract, occupation, industry, region, establishment size, tenure, part-time status, education, health status and marital status. Where the number of missing values was non-trivial (typically where this would have necessitated a drop in sample size of 5% or more as a consequence), we also included a dummy variable for missing value observations in order to preserve the sample sizes. Finally, we kept only continuous sequences of observations from the first wave (ECHP wave 1) to ensure a complete record of training for each individual. The table details the number of observations remaining at each of these selection stages.

[1] Country	[2] Initial no. of obs after first selection	[3] No. of obs with valid data	[4] Addition al selections used	[5] Included missing value dummies	[6] Included waves	[7] No. of obs [ind] after selection of continuing spells	[8] Other comments
Austria	3189	3029			3-6	2366 [786]	.
Belgium	3406	2680		Size	2-6	1524 [492]	.
Britain	5569	4246	Wave 6 deleted.	Industry, Fixed Term/Casual contract	2-5	3366 [986]	Training not dated.
Denmark	3249	3126		Industry	2-6	2152 [626]	
Finland	2386	2282		Industry, Occupation	4-6	1775 [740]	
France	7483	6589		Fixed Term/ Casual, Size, Occupation, Industry	2-6	4936 [1448]	Training is not dated.
Ireland	2729	2643		Region	2-6	1774 [544]	
Italy	6944	6173			2-6	3757 [1092]	
Netherlands	7038	6719		Industry	2-6	3569 [908]	No training finish dates available.
Spain	6445	6296		Region	2-6	4032 [1204]	