

Is Training More Frequent When the Wage Premium Is Smaller ? Evidence from the European Community Household Panel[§]

Andrea Bassanini (OECD and CEPN-Paris XIII)[†]

Giorgio Brunello (Padova, IZA and CESifo)[‡]

[§] The views expressed in this paper cannot be attributed to the OECD or its member countries. We are grateful to Daron Acemoglu, Giuseppe Bertola, Federico Biagi, Alison Booth, Michael Burda, Eve Caroli, Daniele Checchi, Maria De Paola, Martine Durand, John Martin, Raffaele Miniaci, Hessel Oosterbeek, Enrico Rettore, Jean-Marc Robin, Muriel Roger, Paolo Sestito, Ken Swinnerton, Alain Trannoy, Tanguy van Ypersele, Guglielmo Weber, Lüdger Wössmann, two anonymous referees and seminar participants in Aix, Evry, Lyon (Journées AFSE), Marseille (GREQAM), Milan (Cattolica), Munich (CESifo), Orléans (T2M), Novara, Padova, Paris (CEPN, ESPE and EUREQUA), Pisa (Sant'Anna) and Tokyo (Keio) for comments and suggestions. The usual disclaimer applies. The data used in this paper are from the December 2001 release (contract 14/99).

[†] Andrea Bassanini: OECD and CEPN, University of Paris XIII ; Address : OECD, 2 rue André Pascal, 75775 Paris Cedex 16, France; Ph.: +33 1 45249032. e-mail: andrea.bassanini@oecd.org .

[‡] Giorgio Brunello: Department of Economics, University of Padova, IZA, Bonn and CESifo, Munich; Address : Department of Economics, University of Padova, via del Santo 33, 35100 Padova, Italy; Ph.: +39 044 8274223; e-mail: giorgio.brunello@unipd.it .

Abstract

According to Becker [1964], when labour markets are perfectly competitive, *general* training is paid by the worker, who reaps all the benefits from the investment. Therefore, *ceteris paribus*, the greater the training wage premium, the greater the investment in general training. Using data from the European Community Household Panel, we compute a proxy of the training wage premium in clusters of homogeneous workers and find that smaller premia induce greater incidence of off-site training, which is likely to impart general skills. Our findings suggest that the Becker model provides insufficient guidance to understand empirical training patterns. Conversely, they are not inconsistent with theories of training in imperfectly competitive labour markets, in which firms may be willing to finance general training if the wage structure is compressed, that is, if the increase in productivity after training is greater than the increase in pay.

Keywords: general training, off-site training, training wage premia, wage compression, ECHP

JEL Classification: J24, J31, J41

1. Introduction

Human capital is a key determinant of economic growth. The amount of training individuals receive during their working life has a significant impact on their career prospects, wages and employability. Moreover, improving workers' competencies is crucial in the face of rapid technological change. In spite of the broad consensus on the importance of training, there is a large debate in the economic literature and in policy circles concerning whether the current level of investment in training is efficient and which agent (employers or employees) has the incentive to invest.

According to Becker [1964], when labour markets are perfectly competitive, *general* training — that is training which raises productivity at other employers to the same extent as at the employer who provides it — is fully paid by the worker, who reaps all the benefits from the investment. However, although few surveys have information on the generality of skills, in those that do, most of the reported job-related training appears to be employer-paid, at least partially, even when it is viewed by respondents as general (Barron, Berger and Black [1999], Loewenstein and Spletzer [1999], Booth and Bryan [2005]). Furthermore, in many surveys, employers appear to pay for most of off-site training (OECD [2003]), which is found by Loewenstein and Spletzer [1999] to be essentially general. This evidence is difficult to reconcile with Becker's model unless we argue that the employee fully compensates the employer by accepting lower wages during (or before) training spells. Yet, no clear empirical support exists for this fact in the literature.¹

Recent theories of imperfect competition in the labour market can explain why employers have an incentive to pay for apparently general human capital. If the market for trained workers is less competitive on the demand side than the market for untrained workers, the ratio of wages to productivity is lower for trained than for untrained workers — that is, wages are compressed with respect to productivity along the training dimension. In these circumstances, the employer has an incentive to train because he can afford to pay a trained worker less than the marginal product while still retaining her (see *e.g.* Katz and Zidermann

¹ See among others, Loewenstein and Spletzer [1998], Barron, Berger and Black [1999] and Sicilian [2001], as well as Bishop [1997] for a survey of earlier studies.

[1990], Stevens [1994], Acemoglu and Pischke [1999a], Lazear [2003] and Booth and Zoega [2004]).

In this paper, we contribute to a growing body of literature attempting at discriminating empirically between these two alternative theories of training, by estimating how the probability of receiving general training is affected by the training wage premium — that we approximate with the difference between the median wage growth of trained and untrained employees in clusters of relatively homogeneous workers. Becker's model implies a positive relationship between the training wage premium and the incidence of general training. By contrast, a negative relationship falls within the possible outcomes of training theories based on imperfect competition in the labour market.

A few recent papers have provided empirical evidence that is consistent with models of firm-sponsored training based on imperfect competition in the labour market. In particular, panel data studies, which control for individual fixed effects and for job mobility, show that some of the benefits of general training are appropriated by workers with some lag and/or when they change employers. This evidence is consistent with the view that employers have some monopsony power over their own trained workers. For example, using data from the US National Longitudinal Survey of Youth (NLSY), Loewenstein and Spletzer [1999] find that, when training imparts general skills, the estimated effect of completed spells of employer-paid training on earnings is three times larger for training spells completed during previous jobs than during the current job. Similarly, Loewenstein and Spletzer [1998] find that completed spells of employer-provided off-site training in the current job have no effect on current wages. By contrast, off-site employer-paid training received at previous employers has a positive and persistent impact on wages. Using more waves of the same data, Lengermann [1999] finds that the latter effect increases over time. Booth and Bryan [2005] study three recent waves of the British Household Panel Survey (BHPS) and find that employer-provided training has a positive and persistent impact on wages, with evidence that the impact is larger for accredited training received at previous employers. Similar results are obtained by Blundell, Dearden and Meghir [1999], using three distant waves of the British National Child Development Survey (NCDS), and by Gerfin [2004] on Swiss data.

However, the presence of labour market imperfections is not the only possible explanation of the finding that wages after a training spell grow faster if the worker changes jobs. For instance, workers might undertake training to qualify for jobs in other firms that are more efficient at employing trained workers (see e.g. Moen and Rosen [2002]) and the main

empirical predictions of the Becker model are therefore not necessarily in contrast with these findings.

Another way of discriminating between alternative theories of training is to investigate the relationship between the minimum wage and training incidence. According to Becker, high minimum wages should decrease investments in human capital, as they would prevent minimum wage workers from accepting wage cuts to finance training. By contrast, in the alternative theories of training, the greater the minimum wage the greater the incentive for firms to pay for general training. The reason is that, in imperfectly competitive labour market, the minimum wage compresses the lower tail of the wage distribution without necessarily affecting individual productivity (see e.g. Acemoglu and Pischke [2003]). However, recent empirical studies both in the United States and in the United Kingdom report contradictory findings on the impact of the minimum wage on training.² There are several possible reasons why this strand of research is inconclusive. For instance, in countries where the minimum wage is high, it might be difficult to find a group which is not directly or indirectly affected by the minimum wage and qualifies as a genuine control. Conversely, in countries where the minimum wage is particularly low, the incidence of training in the treatment group is likely to be extremely small, since training is relatively infrequent at the bottom of the wage distribution. Moreover, most of these studies focus on changes of the minimum wage over time, but it is not clear what time horizon is appropriate to analyze the effect of institutional changes such as increments in the minimum wage. Last but not least, the degree of imperfection of the labour market might differ across countries.

The approach we follow in this paper is similar to that taken by studies of minimum wage and training, insofar as we focus on the implications of the different theories on the relationship between wage premia and training. We use cross-country data from the European Community Household Panel (ECHP) and partition workers into clusters of relatively homogeneous employees (in terms of country, education, occupation and sector). We then construct cluster-specific measures of the training wage premium – that we compute as the difference between the median wage growth rates of trained and untrained employees – and investigate whether these measures have a significant impact on general training, after dealing

² See Grossberg and Sicilian [1999], Neumark and Wascher [2001], and Acemoglu and Pischke [2003], for the United States, and Arulampalam, Booth and Bryan [2004], for the United Kingdom.

with the potential problems of endogeneity. We find a negative relationship between our measure of the training wage premium and the probability of taking general training. Our findings are inconsistent with the Becker model, which implies a positive relationship between the training wage premium and the incidence of general training. Conversely, although we cannot observe individual productivity gains from training, our findings can be interpreted as lower bound estimates of the impact of wage compression on training, to the extent that the (unobserved) productivity premium and the (observed) wage premium are not negatively correlated. Under this condition, our findings are consistent with theories based on imperfect competition. This approach seems to us particularly suitable to analyse the empirical relevance of different theories of training using European data: since the migration of labour between EU countries is still limited, we can meaningfully use the country dimension in the definition of clusters, which allows us to construct a sample with a large number of clusters and to obtain significant variation in our measures of the training premium.

The paper is organized as follows. Section 2 presents our empirical approach, Section 3 describes the data and Section 4 is dedicated to the presentation of the empirical results. Conclusions follow.

2. General Training and Wage Premia

To fix the ideas and discuss our empirical specification, in this section we briefly sketch a simplified two-period model of general training, which essentially corresponds to the Becker model in the case of perfect competition, and to a simplified version of the model by Acemoglu and Pischke [1999a] in the case of imperfect competition.

In the first period a worker is matched with an employer and training can take place. In the second period, the worker can quit if she receives a better wage offer. Let us denote with $f(\tau)$, $w(\tau)$ and $v(\tau)$ the worker's second period labour productivity, wage and outside option, respectively, with $\tau = 1$ indicating that the individual takes general training in the first period (while $\tau = 0$ indicates that no training takes place). Let us assume that training is perfectly general, which is equivalent to assume that $f(\tau)$ and $v(\tau)$ do not vary across firms. In the case of perfect competition considered by Becker, $f(\tau) = w(\tau) = v(\tau)$ and $\Delta f = \Delta w = \Delta v$, where $\Delta w = w(1) - w(0)$ is the wage premium and Δf and Δv are similarly defined. Since $\Delta f - \Delta w = 0$, if training costs $c(1)$ are positive and $c(0) = 0$ the employer does

not invest. Conversely, and assuming no discounting for simplicity, the worker invests and pays the training costs as long as $\Delta w > c(1)$. If training costs vary across workers, the greater the wage premium, the larger the number of employees that are ready to pay for training. In other words, for a given individual i , the probability that she decides to take training can be written as:

$$\text{Prob}\{\tau_i = 1\} = \text{Prob}\{\Delta w_i > c_i(1)\} \quad [1]$$

If we remove the assumption of perfect competition on the demand side, the worker has to incur a cost upon quitting — that is, her outside option $v(\tau)$ is below her productivity $f(\tau)$ by the amount $D(\tau) > 0$. Assume that the current employer wants to retain the trained worker. He can offer the worker a wage w equal to her outside option v plus a fraction of the gap D , that is $w(1) = v(1) + \beta(f(1) - v(1)) = v(1) + \beta D(1)$, where $\beta < 1$ represent the worker's bargaining power. In the absence of training, the wage offered to retain her would have been $w(0) = v(0) + \beta(f(0) - v(0)) = v(0) + \beta D(0)$. Assuming that workers cannot bear any training cost (e.g. because of borrowing constraints), the expected profits from training and no training are $\pi(1) = f(1) - w(1) - c(1) = (1 - \beta)(f(1) - v(1)) - c(1)$ and $\pi(0) = f(0) - w(0) = (1 - \beta)(f(0) - v(0))$ respectively. In the decision to invest in general training, the employer compares the expected profits from training with the expected profits in the event of no training. The employer decides to bear the training costs if and only if $\Delta f - \Delta w > c(1)$ or, equivalently, $\Delta f - \Delta v > c(1)/(1 - \beta)$. This inequality will hold if training costs are sufficiently small and D is an increasing function of training, since when $D'(\tau) > 0$, $\Delta f - \Delta v > 0$. There are several types of labour market imperfections on the demand side that might make $D'(\tau) > 0$: firms may lack of information on previous training of job candidates; search costs might be greater for trained workers; general and firm-specific skills might be difficult to separate and might, therefore, be imparted together; labour demand might also be distorted by institutions such as coordinated industrial relations or pay-back clauses³ (see Acemoglu and Pischke [1999b] for a

³ For instance, contracts with pay-back clauses typically impose that the worker pays-back part of training costs incurred by the firm if she quits within a specified period and the firm can show that skills learned through training are useful for the worker in the new job. When applicable, these contracts

survey). Lazear [2003] suggests that this pattern might also emerge if skills are a multi-dimensional variable, with different firms using different combinations of general skills. If training costs vary across workers, the more compressed the wage structure (the higher $\Delta f - \Delta w$), the larger the number of employees for which employers are ready to pay for training. For a given individual i , the probability that she receive employer-paid training is:

$$\text{Prob}\{\tau_i = 1\} = \text{Prob}\{\Delta f_i - \Delta w_i > c_i(1)\} \quad [2]$$

With a similar argument, it can be shown that, even when workers bear some of the training cost, the employer's investment in general training increases with wage compression. In this case, however, total investment in training may or may not increase with wage compression, since the incentives for the employee to invest in training are greater, the greater the training wage premium Δw . In the special case of perfect competition, the investment in training is paid by the employee and is increasing with the wage premium.

It stands out from this discussion that the case of perfect competition considered by Becker can be viewed as a special case of the Acemoglu and Pischke model: in the absence of labour market frictions, D is equal to zero, $f(\tau) = w(\tau) = v(\tau)$ and the employer never pays for general training.

Taking into account that the training that is reported to be employer-sponsored might be indirectly paid by the worker by accepting lower wages during training, equations [1] and [2] can be generalised as:

$$\text{Prob}\{\tau_i = 1\} = \text{Prob}\{\gamma \Delta f_i + \sigma \Delta w_i > \delta c_i(1)\} \quad [3]$$

In fact, from equation [1] we expect $\gamma = 0$, $\sigma = 1$ (or, more generally, $\sigma > 0$) and $\delta = 1$, in the case of the Becker model. Conversely, in the extreme case of the Acemoglu and Pischke model where only firms pay for training (equation [2]), we expect $\gamma = 1$, $\sigma = -1$ and $\delta = 1$. Finally, in the general case of the Acemoglu and Pischke model in which both firms and workers can pay for training, workers' investment will be represented by [1] and firms'

temporarily increase worker's cost of quitting (except for reasons unrelated to the content of training), thereby creating a captive market for the training firm.

investments by [2], therefore total training investments can be represented by equation [3] with $\gamma > 0$, $\sigma \geq -\gamma$ and $\delta > 0$.

Employers and employees decide on the investment in training by forming expectations about the training wage premium. We posit that agents form their expectations by looking at the current wage distribution for trained and untrained employees. The heterogeneity of workers and jobs suggests, however, that the relevant distribution should not be the entire wage distribution, but rather an appropriate portion of it. Therefore, we partition individuals into relatively homogeneous clusters and approximate the training wage premium with the difference between the median wage growth rates of those who reported to have received training in the period covered by the survey and of those who did not. Individual wages are affected both by the current training investment and by the accumulated training stock before the sample period. Since we can observe individual training history only for very few years, a substantial part of this stock is not observed. By using *growth* rates rather than *levels*, we are able to eliminate the influence on wages of the training stock accumulated before the reference period. We call this measure the wage growth differential.

Without cross-country comparable matched employer-employee datasets (see Bartelsman, Scarpetta and Schivardi [2005]), there are no cross-country data on individual productivity. Therefore, we treat the productivity gain from training Δf as an omitted variable and discriminate among competing theories on the basis of the estimate of the parameter σ . This implies that if we find $\hat{\sigma} < 0$ (with $\hat{\cdot}$ standing for estimate), then we will conclude that our evidence is in contrast with the empirical prediction of the Becker model. Conversely, we will not be able to discriminate among different theories if we find $\hat{\sigma} \geq 0$.

How does the omission of the productivity gain Δf bias our estimates? If the Becker model were true, omitting the productivity gain from the specification would induce no bias, since its expected coefficient is zero. Conversely, in the Acemoglu and Pischke model, to the extent that productivity and wage premia are not negatively correlated, the omission of the former would at most bias our estimates of σ against finding a negative relationship between the wage premium Δw and training incidence. In this case, our estimate of σ could be interpreted as a lower bound estimate of the impact of wage compression on training.

We do not observe training costs directly. Yet, we assume that worker's opportunity costs of training and liquidity constraints can be captured by a set of variables related to the employee's household, such as the household structure, family responsibilities and the financial situation of the household (we will call these variables "family variables" hereafter).

As training incidence typically vary with firm size, individual age, tenure in the job, educational attainment, type of labour contract, previous unemployment record, sector of activity, occupation and country, we also let training costs vary along these dimensions. Letting ΔW_c be the wage growth differential in cluster c and X the vector of controls potentially capturing training costs, our empirical specification then becomes

$$\Pr ob\{\tau_{ic} = 1\} = \Pr ob\{\alpha + \beta X_{ic} + \sigma \Delta W_c + \varepsilon_{ic} > 0\} \quad [4]$$

where the null hypothesis is $\sigma \geq 0$, as implied by the Becker model.

A potential objection to our empirical proxy of the wage premium is that it can be endogenous. For instance, if there are diminishing returns to human capital and trained and untrained workers are imperfect substitutes in production, we expect that the greater the *stock* of previously accumulated training the lower the productivity gain from training. Therefore, to the extent that the *stock* of previously accumulated training and the *flow* of training taken in the period covered by our data are correlated, we may find a negative relationship between the training wage premium and the probability of training even if the Becker's model is true.

To take this issue into account, we test the hypothesis of weak exogeneity of the wage growth differential by following the methodology suggested by Smith and Blundell [1986]. To implement this test, we need to select at least one exogenous variable which is correlated with ΔW_c but is independent of training incidence (or the probability of training), conditional on ΔW_c . Our selected instrument is the difference between the log median age of those who have received training and the log median age of those who have not (we call this difference the log age differential and denote it with ΔA_c).⁴

We expect ΔA_c and ΔW_c to be correlated because training wage premia have been found to decrease with age, at least in European countries (see e.g. OECD, 2004). This result is typically explained by the fact that, in the absence of training, the wedge between productivity and the wage tends to decrease with age (see e.g. Abowd and Kramarz, 2003 and Aubert and Crepon, 2004, for France), due to labour market institutions and sectoral characteristics that affect the degree of downward nominal wage rigidity as well as the extent

⁴ This measure roughly corresponds to the percentage difference between the two medians divided by 100.

of internal labour markets. As workers get older, thus, training simply allows matching non-decreasing wages with otherwise decreasing productivity, with no apparent wage premium.

We argue that ΔA_c also satisfies the orthogonality condition for instrument validity. First, there is no compelling reason for a direct causal impact of ΔA_c (which is an aggregate age variable) on the probability of training, once we have controlled for observable individual characteristics, including individual age.⁵ Second, ΔA_c is not affected by diminishing returns by construction: with diminishing returns, an increase in the number of trained employees is likely to affect the wage of all workers, but has no influence on demographic characteristics such as age. True, since younger workers are likely to be trained first, we expect that the greater the incidence of training, the greater the median (log) age of workers who receive training as well as of those who do not. But the log age differential ΔA_c will not be affected if the two distributions — of the log age of workers who receive training and of those who do not — have the same degree of concentration around their respective medians.⁶ If this is the case — as we show it is at the end of the next section — diminishing returns affect relative prices ΔW_c without affecting relative ages ΔA_c .

⁵ A direct effect might emerge in the presence of large human capital externalities, if the intensity of these externalities varies with age. We exclude this possibility since we are aware of no evidence supporting this hypothesis in the literature.

⁶ To see this, assume first that, in a given cluster, an individual is trained if and only if her age is below a given threshold, and that increasing training incidence implies shifting this threshold upwards, thereby making the youngest individuals in the group of the untrained switch to the group of the trained, of which they become the oldest. The crucial point is that both the median log ages of trained and untrained workers are lifted up, but the size of their relative increase depends on the concentration of the log age distribution in the neighbourhood of the two medians. If the distribution has the same degree of concentration in the neighbourhood of the two medians, the difference between them is independent from either the size of the groups and/or the exact age of switching individuals, and therefore the aggregate training rate (this statement is derived formally in the working paper version of this article - Bassanini and Brunello [2006] - which the reader is referred to also for the extension of the argument to the general case).

3. Data

We use individual data from the December 2001 release of the *European Community Household Panel* (ECHP), which is a longitudinal survey modelled on the British Household Panel Survey (BHPS). This survey provides a wealth of information on individual income and socio-economic characteristics for all EU countries and aims to be representative both in cross-sections and longitudinally. Due to the common questionnaire, the information contained in the ECHP is, in principle, comparable across countries, which is its main strength.

We use training data from the 1996 wave of the ECHP for 7 countries⁷ and restrict our attention to male employees (excluding apprentices), aged from 30 to 60 years⁸ and working full-time in the non-agricultural private sector, excluding sectors where non-profit organizations account for a non-negligible fraction of employment.

The main question on vocational training in the ECHP is as follows "Have you at any time since January (year before the survey year) been in any vocational education or training, including part-time and short-courses?". From this question, we construct a dichotomous variable "training participation", which takes value 1 if the individual responded "yes" and 0 if she responded "no". Conditional on a positive answer, the individual is asked to report additional information on the last course only (including the type, whether it is still ongoing at the date of the survey and whether it is paid for or provided by the employer). If more than one concurrent course are involved, only the information concerning the course considered by the respondent as the most important is reported. Table 1 shows training events and training incidence — trained individuals as a percentage of the relevant population — by country and selected characteristics. About 17 percent of the individuals in our sample have experienced training, but there is a large cross-country variation. The country ranking that is shown in the table is similar to what emerges from other cross-country European surveys, such as Eurostat

⁷ The choice of the survey year and the country sample is dictated by data availability (see Bassanini and Brunello [2006]). Included countries are Austria, Belgium, France, Germany, Italy, Spain, and the United Kingdom.

⁸ We exclude individuals under 30 to reduce the risk that our results be altered by different national institutions affecting the school-to-work transition (such as different apprenticeship systems, with different degrees of government support).

Continuing Vocational Training Survey (CVTS) (see Nestler and Kailis [2002] and OECD [2003]).

Table 1: Training events, by country and selected characteristics.

Country	Individuals receiving training (in % of the total)	Other Characteristics	Individual receiving training (in % of the total)
Austria	159 (21.1)	Less than upper sec. education	201 (7.7)
Belgium	110 (19.9)	Upper secondary education	550 (18.4)
France	201 (15.5)	More than upper sec. education	443 (33.7)
Germany	219 (19.9)	Mining, manuf. and utilities	653 (15.0)
Italy	81 (6.7)	Services	541 (21.4)
Spain	150 (11.8)	High-skilled occupations	622 (31.0)
United Kingdom	274 (39.1)	Medium-skilled occupations	538 (12.4)
Total	1194 (17.3)	Low-skilled occupations	34 (6.3)

Note: male employees, aged from 30 to 60 years and working full-time in the non-agricultural private sector, excluding sectors where non-profit organizations have a non-negligible share of employment. The table shows the number of individuals who reported to have received training in the 1996 survey, by country and selected characteristics. Training incidence (trained individuals as a percentage of the relevant population) is reported in parentheses.

Respondents who have been in vocational education or training are asked to select the type of training received among the following categories: a) third level qualification, such as technical college; b) specific vocational training at a vocational school or college; c) specific vocational training within a system providing both work experience and complementary instruction elsewhere; d) specific vocational training in a working environment, without complementary instruction elsewhere; e) other. The distribution of training events by type is 5.1 percent for type a, 16.3 percent for type b, 11.1 percent for type c, 64.6 percent for type d, and 2.8 percent for type e. The questionnaire also asks individuals whether the last vocational training course was paid for or organized by the employer. As expected, about 86.9% of the courses on which the information is available are paid for or organized by the employer. Interestingly, the more formal the training the lower the employer support: 57% for third level qualification, such as technical college; 76% for specific vocational training at a vocational school or college; 92% for specific vocational training within a system providing both work experience and complementary instruction elsewhere; and 95% for vocational training in a working environment.

As said in the previous section, our empirical measure of the training premium is based on comparing the wage growth rates of individuals who reported to have received

training in the period covered by the survey and of those who did not — hereafter we will refer to them as “trained” and “untrained”, respectively. Individual wage growth is computed as the difference between the log gross hourly wage reported in the current wave – year 1996 – and the log wage reported in the previous wave – year 1995 – by the same individual. In the case of Austria, for which 1995 wage data are not comparable with those from other years (see the appendix), we replace the 1995-1996 wage growth with that of 1996-1997. Additional information on wages is reported in the appendix.

Table 2 shows the difference between the median wage growth rates of trained and untrained individuals by country, educational attainment, occupation, and sector of activity. The median wage growth of trained individuals is lower than the median wage growth of the untrained in Germany and Italy, although the difference in these countries is close to zero. The table shows a wage return in the range 0-4 percent. Although these figures might seem low, they are consistent with panel data estimates based on European data (see Pischke [2001], Booth and Bryan [2005], Gerfin [2004], Schøne [2004] and OECD [2004]). The simple comparison of Table 2 and Table 1 shows no clear correlation pattern between training and wage premia at this aggregate level.

Table 2: Wage growth gaps between trained and untrained workers.

Country	Median wage growth, difference trained / untrained (%)	Other Characteristics	Median wage growth, difference trained / untrained (%)
Austria	1.05 [1.32]	Less than upper sec. education	-0.08 [1.50]
Belgium	0.33 [1.85]	Upper secondary education	0.34 [0.35]
France	1.38 [0.72]	More than upper sec. education	1.17 [0.93]
Germany	-0.17 [0.97]	Mining, manuf. and utilities	1.15 [0.63]
Italy	-0.21 [2.68]	Services	0.00 [0.63]
Spain	4.52 [2.32]	High-skilled occupations	-0.24 [0.90]
United Kingdom	0.05 [1.48]	Medium-skilled occupations	0.61 [0.52]
Average	0.29 [0.23]	Low-skilled occupations	5.19 [5.63]

Note: Percentage point difference between the median wage growth rates of those who reported to have received training in the period covered by the 1996 survey (1997 for Austria) and of those who did not. The sample is limited to full-time male employees aged from 30 to 60 years and working in the non-agricultural business sector, excluding individuals who reported to be still in training at the time of the survey. Wage data refer to 1995-1996 for all countries except for Austria, for which they refer to 1996-1997. Medians are weighted by cross-sectional weights. Bootstrapped standard errors, obtained with 100 replications, in brackets.

We compute our empirical proxy of the training wage premium (that we called "wage growth differential" in the previous section and denoted with ΔW_c in equation [4]) by cluster.

We define clusters by four dimensions: the country, the educational attainment (*less than upper secondary, upper secondary, more than upper secondary*), the broad group of sectors (*mining, manufacturing, utilities and construction, and services*), and the broad occupational group (*high-skilled occupations and medium- and low-skilled occupations*).⁹ By so doing we obtain 12 clusters per country, but some of them are empty.¹⁰

The wage growth differential ΔW_c is obtained by computing for each cluster the difference between the median wage growth rates of the employees who reported to have received training in the period covered by the survey¹¹ and of those who did not, excluding workers who either changed cluster between the two wage observations or have missing cluster affiliation in one of the relevant interviews (about 6.5% of the sample). Since our measure of the wage growth differential should capture the cluster-specific premium to completed training spells, we exclude from this calculation also all individuals who reported to be still in training at the time of the survey or have missing information on this question (about 19.4% of individuals who received training in the period covered by the survey).¹² Although we limit our regression analysis to workers employed in sectors where non-profit organizations are not important, we consider all employees in the non-public service sector in the computation of the wage growth differential. This is done because we believe that the whole private service sector should be considered as the relevant market for service sector workers. Finally, in order to reduce the weight of outliers, we drop clusters with 30

⁹ corresponding to managers, professional technicians and associate professional — ISCO-88 codes 1 to 3 — and to clerks, service and sales workers, craft and related trade workers, plant and machine operators and assemblers, and elementary occupations — ISCO-88 codes 4 to 9 —, respectively. Low-skilled occupations are aggregated to medium-skilled occupations since the size and training incidence of this occupational group is too small to be used separately in the definition of clusters (see Table 1).

¹⁰ Given that in our data sectors and, especially, occupations are defined at a lower level of aggregation (see appendix), it might be argued that a finer partition of the data should be used to define clusters. However, since our empirical measure of the training premium controls for time-invariant heterogeneity within clusters (see below), the error induced by heterogeneity in large clusters is likely to be smaller than the error induced by small cluster size. Nevertheless, in a sensitivity analysis, we explore the effect of different partitions by varying the grouping of occupations (see Bassanini and Brunello [2006]).

¹¹ That is, for growth rates between 1995 and 1996, those who reported to have received training in the 1996 survey and, for growth rates between 1996 and 1997, those who reported to have received training in the 1997 survey.

¹² Ideally, we would like to control for the number of spells of training in the period covered by the survey. Yet, this information is not available in the ECHP. The lack of this control constitutes an important limit to the conclusiveness of our results.

observations or less and with less than 5 reported training events, and remain with 47 valid clusters.¹³ Overall, we observe no clear pattern of cross-cluster bivariate correlation between training incidence and the wage growth differential (the correlation coefficient between these variables at the cluster level being only -0.13).

One problem with this measure of the training wage premium is that, given the formulation of the questions on training in the ECHP questionnaire (see above), when we compute the wage growth differential we misclassify as trained certain individuals who in fact received training only between January of the year before the current survey and the date of the previous survey (that is, except for Austria, between January 1995 and the date of the 1995 survey). As shown by Frazis and Loewenstein [1999], this implies that the computed wage growth differential will tend to be smaller than the value we would obtain if we focused only on training taken between the two survey dates (that is the dates of the two wage observations). Notice, however, that, to the extent that there is no reason why the proportion of misclassified individuals in a cluster should vary with training incidence, we can expect measurement error not to be systematically related to training incidence at the cluster level. In other words, this type of measurement error will, at most, bias our results towards zero, thereby making a rejection of the Becker model more difficult. Moreover, the absence of a systematic relationship between training incidence and measurement error can be tested by following the same instrumental variable procedure suggested by Smith and Blundell [1986], that we discussed in the previous section.

In the previous section we argued that the difference in the log median age of trained and untrained employees (the log age differential ΔA_c) is a valid instrument for the wage growth differential ΔW_c . We noted, however, that this statement is correct only if there is no systematic difference in the concentration of the log age distributions in the neighbourhood of the median age of those who received training and of those who did not. These hypotheses can be verified by computing for each cluster, and for two quantiles that are not far from the median, the interquantile difference for both age distributions and by checking whether their cross-cluster averages are systematically different. We performed this test for the difference

¹³ These threshold limits reduce the number of clusters by more than one third. Not surprisingly, however, retained clusters accounts for a much larger share of observations with non-missing wages (see appendix). Nevertheless, we checked the robustness of our results by lifting these size thresholds.

between the 55th and the 45th percentiles, the 60th and the 40th percentile and the 70th and the 30th percentiles. As shown in Table 3, no significant difference emerges for any of the three, which supports the validity of ΔA_c as an instrument.

Further details on the construction of the variables used in the empirical analysis as well as descriptive statistics are reported in the Appendix.

Table 3: Tests of differences in the concentration of the log age distribution around the medians of the trained and the untrained.

Interquantile difference	cross-cluster average difference
P55-P45	-0.007 (0.010)
P60-P40	-0.011 (0.013)
P70-P30	-0.012 (0.016)

Note: The table reports the cross-cluster average differences between any given interquantile difference for the log age distribution of those who received training and the same interquantile difference for the log age distribution of those who did not receive training. Averages are weighted by cluster size. Standard errors in parentheses

4. The Empirical Results

Our empirical analysis proceeds in two steps. First, we estimate a probit model of the probability of taking any type of training, using the cluster - specific wage growth differential ΔW_c as our measure of the training wage premium, and testing its weak exogeneity with respect to training. Second, we re-estimate our model by focusing exclusively on training that is likely to be general. A series of robustness checks on the latter specification are presented in the working paper version of this article (Bassanini and Brunello, [2006]).

Table 4 shows the results of the first model, in which the dependent variable, training participation T , is equal to one in the event of training and to zero in the event of no training. The table is divided into two panels: Panel A refers to all training with no distinction in terms of financing; in Panel B we also use the available information on who financed the last training course and repeat the analysis by focusing on the probability of receiving employer-

sponsored training.¹⁴ In this case, we redefine the dependent variable T and set it equal to one in the event of employer – provided training and to zero in the event of no training.

We take explicitly into account the fact that the training variable T and the wage growth differential ΔW_c are measured at different levels of aggregation and adjust the standard errors by allowing errors to be independent between clusters and correlated within clusters (as suggested by Moulton [1986]). The table reports only the relevant coefficients from each equation. Full results are available from authors.

We consider two specifications: the first specification includes only basic controls (family variables, age, education, occupation, sector and country; Columns 1, 3, 5 and 7), the second includes an extended set of controls (adding firm size, tenure, permanent job and previous unemployment; Columns 2, 4, 6 and 8). The first two columns of each panel report standard estimates, while the other columns report IV estimates obtained with the procedure suggested by Smith and Blundell [1986]: we compute the residual from the first stage regression in which the wage differential is regressed on the cluster-specific instrument ΔA_c and on country, occupation, education and sector dummies; then we add this variable to the probit specification and re-estimate it. The estimated coefficient of ΔW_c that we obtain with this procedure is a consistent IV estimate of the impact of the wage growth differential. Furthermore, we can also test the weak exogeneity of the wage growth differential ΔW_c by checking whether the estimated coefficient associated to the residual from the first stage regression is statistically different from zero.

¹⁴ While we do not know whether workers indirectly paid for employer-sponsored training by accepting lower wages, we can be relatively confident that employers did not pay for courses that are reported to be non-sponsored. It might therefore be desirable to eliminate these events from our dependent variable, since alternative theories have identical predictions for them under all circumstances. When the last training course is not reported to be employer-sponsored, we set therefore T to missing since we do not know whether the individual has taken employer-sponsored courses before the last one in the period covered by the survey.

Table 4. *Probits for total training*. Dependent variable: training participation T in 1996.
Panel A: all training

	(1)	(2)	(3)	(4)
Wage growth differential	-0.039 [0.083]	-0.067 [0.080]	-0.381 [0.259]	-0.460 [0.233]**
Residual from first stage regression			0.358 [0.278]	0.358 [0.278]
Basic controls	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
Extended controls	<i>no</i>	<i>yes</i>	<i>no</i>	<i>yes</i>
Number of observations	5898	5515	5898	5515
F-test on instrument significance	5.00	5.00	5.00	5.00
Wald-test of homoskedasticity ($\chi^2(2)$)	0.14	0.15	0.14	0.27
Pseudo R-squared	0.154	0.179	0.154	0.179

Panel B: Employer-sponsored training

	(5)	(6)	(7)	(8)
Wage growth differential	0.002 [0.091]	-0.036 [0.084]	-0.347 [0.245]	-0.417 [0.211]**
Residual from first stage regression			0.364 [0.268]	0.400 [0.238]*
Basic controls	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>yes</i>
Extended controls	<i>no</i>	<i>yes</i>	<i>no</i>	<i>yes</i>
Number of observations	5791	5412	5791	5412
F-test on instrument significance	5.00	5.00	5.00	5.00
Wald-test of homoskedasticity ($\chi^2(2)$)	0.59	1.20	0.21	0.92
Pseudo R-squared	0.167	0.201	0.168	0.202

Note: The table reports marginal percentage-point effects. Cluster adjusted robust standard errors within brackets. Observations are weighted by ECHP personal weights. Individuals reporting non-sponsored training are excluded from Panel B. Each column of each panel refers to a different specification. Basic controls are country, education, occupation and sector dummies plus age and family variables. Specifications with extended controls add firm size, tenure, permanent job status and previous unemployment dummies to basic controls. In the first stage the wage growth differential is regressed on the median log age differential, country, occupation, education and sector dummies. The Wald test statistic, which is used to test the null hypothesis of homoskedasticity against the alternative of multiplicative heteroskedasticity modelled as a function of the wage growth differential and its square, is distributed as a $\chi^2(2)$ under the null. **, *: significant at the 5% and 10% level of confidence, respectively.

The results that emerge from Table 4 are inconclusive. On the one hand, standard estimates point to no correlation between the wage growth differential ΔW_c and the probability of training. On the other hand, IV estimates tend to suggest a much more negative and sometimes significant relationship.¹⁵ Exogeneity tests, however, are not clear-cut as regards which estimation procedure should be preferred. In fact, even though in Panel A

¹⁵ One possible explanation of this difference is that standard estimates are strongly affected by measurement error, while diminishing returns are a second order problem.

residuals are insignificant, thereby not rejecting the exogeneity hypothesis, they are often large in absolute terms. Furthermore, according to the results reported in Panel B, the exogeneity hypothesis is rejected at the 10% level in one specification.¹⁶

In Table 4 we do not distinguish between general and firm-specific training. This is unsatisfactory for two reasons. First, alternative theories have different predictions on the relationship between the training wage premium and training only with regard to general training. Second, the inconclusive results reported in the table might be driven by the heterogeneity of training.¹⁷

We do not have in our dataset information on the generality of skills imparted through training. However, as noted in the previous section, respondents who have been in vocational education or training are asked to classify their last course in one (and only one) of the following mutually exclusive categories: a) third level qualification, such as technical college; b) specific vocational training at a vocational school or college; c) specific vocational training within a system providing both work experience and complementary instruction elsewhere; d) specific vocational training in a working environment, without complementary instruction elsewhere; e) other. Following Loewenstein and Spletzer [1998] and OECD [2003] we can use the distinction between off-site and workplace training to proxy the distinction between general and firm-specific training. Assuming that off-site training tend to be more general than workplace training – assumption that appears to be supported by empirical evidence on US data (see Loewenstein and Spletzer [1999]) – we treat categories a, b and c as “general training” and category d as “firm-specific training”.¹⁸ Nevertheless, since training falling under category c is partly taken off-site and partly received in the workplace, we also experiment with a different classification by assigning category c to “firm-specific training”. Throughout the remainder of the paper we will refer to the former classification as “extensive definition of general training” and to the latter as “restrictive definition of general training”.

¹⁶ The ambiguity of these results is not thoroughly surprising given that our instrument is relatively weak: the value of the F test on the significance of the instrument is 5.0, which is relatively low (although statistically significant at the 5% level).

¹⁷ For instance, if the true wage premium to firm-specific training is close to zero or at least much smaller than the true wage premium to general training, our wage growth differential ΔW_c will tend to be more correlated with the latter than with the former; therefore, the greater the share of firm-specific training in total training in our sample, the greater the importance of measurement error.

¹⁸ The option "other" cannot be classified and we choose to drop it from the sample.

We must be clear, however, that both classifications are just proxies of the distinction we are trying to capture. These proxies are based on the common sense assumption that the generality of skills imparted through training is decreasing from a to d. Yet, despite the terms we adopt in the remainder of the paper to simplify our presentation, the reader should not view training falling into each category as entirely general or firm-specific.

Table 5 reports results obtained by re-estimating the models of Table 4 (Panel A) using general training as the dependent variable. In this case, we exclude the observations where the last training course is reported to be firm – specific, since we do not know whether these individuals received also general training in the period covered by the survey (as discussed above, only one answer is allowed in the questionnaire). Three results stand out. First, under both definitions, standard estimates suggest a statistically significant negative correlation between the wage growth differential ΔW_c and the probability of receiving general training.¹⁹ Second, since the coefficient of the residual from the first stage is relatively small and statistically insignificant, we cannot reject the null hypothesis of weak exogeneity of the wage growth differential ΔW_c . Third, once the residual is added to the specification, the estimated coefficient of the wage growth differential becomes less statistically significant. Notice, however, that the inclusion of the residual leads to inefficient estimates under the hypothesis of weak exogeneity.²⁰ Overall, these results suggest that there is a negative relationship between the training wage premium and the individual probability of attending a general training course. These findings appear inconsistent with the Becker model, which predicts a positive relationship between the training wage premium and the incidence of general training. The estimated impact, however, is rather small: taking, for example, the estimates in the first panel, we obtain that doubling the wage growth differential from its

¹⁹ One might argue that we should use as wage growth differential the difference between the wage growth of those who received general training and of those who received no training. However, due to the relatively low incidence of general training, we would be left with only 15 valid clusters after applying our size thresholds. For this reason, we do so only in the sensitivity analysis reported Bassanini and Brunello [2006], with no qualitative change of results.

²⁰ This argument appears more compelling in the case of the extensive definition, where the residual from the first stage is close to zero. However, when the restrictive definition is used, IV estimates remain significant at least at the 10% level. In interpreting this figure, note that it might be argued that, given the structure of our null hypothesis, a one-tail test would be more appropriate than a two-tail test. P-values derived on the basis of two-tail tests can be seen as lower bounds to the actual P-value.

sample mean (1.95 percentage points) would reduce the probability of general training (using its “extensive” definition) by only 3.1 percent.²¹

Table 5. *Probits for general training*. Dependent variable: Participation in general training in 1996.
Panel A: Extensive definition

	(1)	(2)	(3)	(4)
Wage growth differential	-0.079 [0.025]***	-0.083 [0.024]***	-0.091 [0.073]	-0.101 [0.072]
Residual from first stage regression			0.013 [0.073]	0.019 [0.072]
Basic controls	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>Yes</i>
Extended controls	<i>no</i>	<i>yes</i>	<i>no</i>	<i>Yes</i>
Number of observations	5178	4857	5178	4857
F-test on instrument significance	5.00	5.00	5.00	5.00
Wald-test of homoskedasticity ($\chi^2(2)$)	1.02	0.20	0.99	0.23
Pseudo R-squared	0.226	0.243	0.226	0.243

Panel B: Restrictive definition

	(5)	(6)	(7)	(8)
Wage growth differential	-0.048 [0.019]***	-0.053 [0.019]***	-0.088 [0.045]**	-0.092 [0.047]*
Residual from first stage regression			0.043 [0.041]	0.041 [0.043]
Basic controls	<i>yes</i>	<i>yes</i>	<i>yes</i>	<i>Yes</i>
Extended controls	<i>no</i>	<i>yes</i>	<i>no</i>	<i>Yes</i>
Number of observations	5090	4771	5090	4771
F-test on instrument significance	5.00	5.00	5.00	5.00
Wald-test of homoskedasticity ($\chi^2(2)$)	0.77	0.53	1.01	0.69
Pseudo R-squared	0.244	0.250	0.244	0.250

Note: The table reports marginal percentage-point effects. Cluster adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Individuals reporting firm-specific training are excluded from the sample. Each column of each panel refers to a different specification. The extensive definition of general training is used in Panel A, the restrictive one in Panel B. Basic controls are country, education, occupation and sector dummies plus age and family variables. Specifications with extended controls add firm size, tenure, permanent job status and previous unemployment dummies to basic controls. In the first stage the wage growth differential is regressed on the median log age differential, country, occupation, education and sector dummies. The Wald test statistic, which is used to test the null hypothesis of homoskedasticity against the alternative of multiplicative heteroskedasticity modelled as a function of the wage growth differential and its square, is distributed as a $\chi^2(2)$ under the null. ***, **, *: significant at the 1%, 5% and 10% level of confidence, respectively.

²¹ More precisely, Table 5, Panel A shows that a 1 percentage point change in the wage growth differential would induce a change in the probability of training of about 0.08 percentage points. Taking into account that the sample incidence of general training is 5.19 percent, this translates into an elasticity of 0.031.

One potentially serious problem in probit models is the violation of the assumption of homoskedasticity. For this reason, in order to check for heteroskedastic error terms of the latent variable, we allow their standard deviation to be a multiplicative function of a subset of explanatory variables and their powers, as suggested by Harvey [1976]. In so doing, we follow the standard approach of choosing parsimonious specifications for the error variances (see e.g. Greene Knapp and Seaks [1992]). Tables 4 and 5 present Wald test statistics for error variances that are functions of the wage growth differential and its square and show no sign of heteroskedasticity. Similar results occur if error variances are also allowed to vary across any of the cluster dimensions.

A second potentially important problem that might affect our results is that we do not distinguish between short and long courses. To the extent that wage premia to longer courses might be greater, a negative relationship between wage premia and training might not be inconsistent with the Becker model. For instance, there might be clusters with many mini courses (that is, with high training incidence) that are likely to yield no training wage premium and other clusters with few long courses (that is, low training incidence) yielding quite naturally high wage premia. Information on course length in ECHP data is, however, difficult to use, since it is often missing in certain countries, and the distribution of training courses is heavily concentrated in the neighbourhood of zero (with half of the courses of a duration of less than 10 days). To check that our results are not due to this simple composition effect, in a sensitivity analysis, we eliminate progressively courses of duration inferior to 1 day, 2 days, etc. up to 7 days and re-estimate our models of Table 5.²² While eliminating one-day courses reduces the estimated elasticity of training to the wage premium by about 10-15% — which remains nonetheless significant at conventional statistical levels — this elasticity tends to increase as longer courses are eliminated and is even bigger than our baseline estimate if courses shorter or equal to 7 days are excluded.²³

²² We stop the exercise at 7 days since, after elimination of courses of length smaller or equal to 1 week, we are left with only 25 valid clusters, due to our size thresholds in their definition.

²³ Several other sensitivity exercises — as regards specification, estimation method, computation of the wage growth differential and variations in the sample — are reported in the working paper version of this article (Bassanini and Brunello [2006]).

5. Conclusions

According to Becker [1964], when labour markets are perfectly competitive, only the worker will invest in general training, since she is the only agent who can reap the benefits from the investment. To the extent that labour market imperfections are of limited empirical relevance as regards training, the Becker model represents a simple and powerful model to think about training in practical terms. We can test this statement by looking at the predictions of the model concerning the relationship between the training wage premium and training. In the Becker model, we expect that the greater the wage premium the greater the training investment. By contrast, as a large body of literature has pointed out, when labour markets are imperfectly competitive firms may be willing to finance general training if the wage structure is compressed, that is if the increase of productivity after training is faster than the increase in pay. In this case, a negative relationship between the training wage premium and training might emerge, to the extent that the training wage premium is not positively correlated with wage compression.

In this paper, we contribute to the literature that tries to shed light on the empirical relevance of alternative theories of training, by exploiting the cross-country variation in training incidence and training wage premia within the European Union. We find that the probability of receiving general training, proxied in the paper with off-site training, is higher in clusters — defined by country, sector, occupation and educational attainment — with a lower differential between the median wage growth of trained and untrained employees. Importantly, the negative and statistically significant correlation between the training wage premium and training incidence does not appear to reflect the potential endogeneity of the former variable.

While statistically significant, the estimated impact of changes in the training wage premium on the probability of general training is rather small: conditional on an extended set of controls, a one percent increase in the wage growth differential is expected to reduce the probability of general training by about 0.03 percent. However, to the extent that productivity gains from training and training wage premia are not negatively correlated, this number is likely to be only a lower bound estimate of the effect of wage compression on general training. Should matched employer-employee data on training become available on a cross-country comparable basis, the magnitude of the effect of wage compression on training

incidence could be precisely estimated using our empirical framework, by controlling for the productivity gains from training at the individual and firm level.

Overall, our findings suggest that competitive theories of the labour market, which imply a positive relationship between the training wage premium and the incidence of general training, provide insufficient guidance to interpret empirical training patterns. Conversely, training patterns are not inconsistent with the view that economic environments with higher wage compression can help firms organize and pay for general training, as predicted by the recent theories of training in imperfect labour markets.

Appendix:

A.1 Definition of co-variates and descriptive statistics

Wage growth differentials are defined in terms of gross hourly wages, computed from normal gross monthly earnings in the main job at the date of the interview, by dividing them by 52/12 and by the number of usual weekly hours of work. Overtime pay and hours are included, but individuals who are either still in training in the final year or work less than 30 hours per week (despite declaring to work full-time) or more than 70 hours per week are excluded. Similarly, we exclude workers who change cluster between the two interviews (or with missing cluster affiliation in the first interview). Wage growth differentials (as well as log age differentials) are computed using ECHP personal weights. Although in principle there are 84 clusters, non-missing wage growth information for both individuals receiving training and not receiving training is available in 1996 only in 79 clusters (with an average training incidence of 15.1%).²⁴ Furthermore, the thresholds of 30 observations and 5 training events imply that the sample is reduced to 47 clusters, although, not surprisingly, retained clusters are larger and account for about 90% of observations with non-missing wage growth information (with an average training incidence of 14.9%).

As described in the text, all specifications control for age, educational attainment, occupation, sector and country as well as for a set of family variables. Family variables (see Table A1) include household type (grouped into 8 categories), a dummy for presence of children aged less than 12 years, a dummy that equals 1 if the respondent thinks that looking after other persons prevented him from working as much as he wanted, and a series of variables proxying the financial situation of the family (including indicators of home ownership, the burden of accommodation costs including mortgage, the number of rooms divided by household size and whether the household received an heritage of at least 2000 Eur in the last 12 months). In the regressions, age is measured in years, while all other co-variates (with the exception of wage growth differentials and rooms per household member) are categorical, and coded as sets of dummy variables (omitting one dummy per set for identification). We

²⁴ In the case of Austria, wage growth is computed for 1997, due to data quality problems.

consider three educational attainment levels (less than upper secondary, upper secondary, more than upper secondary, corresponding to standard ISCED categories 0-2, 3, and 5-7), 8 occupational groups (corresponding to standard ISCO-88 codes 1 to 9 with the exclusion of skilled agricultural workers, or ISCO-88 code 6), 13 sectors (defined according the ISIC rev. 3 classification), 4 firm size classes, a dummy for permanent contract, a dummy for at least one spell of unemployment since 1989, and 5 tenure classes. Tenure is obtained as the difference between the survey year and the calendar year of start of the current job. We grouped the data into 5 classes rather than using a continuous variable for two reasons: *i*) the information is censored at 15 years; and *ii*) computed this way, tenure measures are highly imprecise. For example, an individual who is surveyed in December but was hired in January of the survey year would result having tenure shorter than another individual who started in December of the year before the survey year but was surveyed in January.²⁵ The means of the co-variates used in the regressions are reported in Table A1. The standard deviations of age and rooms per household member, the only continuous variables in the sample beside the wage growth differential, are 8.1 years and 0.74, respectively. Information on ECHP data quality and the choice of the sample is available in Bassanini and Brunello [2006].

²⁵ Information on the month of the interview is also available in the ECHP. However, this information is always missing in Germany (for confidentiality reasons) and often missing in a few other countries. For this reason, we opt for not using it.

Table A1: Means of co-variates (baseline sample: age 30 to 60 years, year: 1996).

<i>Variable</i>	<i>Mean</i>	<i>Variable</i>	<i>Mean</i>
Age (years)	42.2	% ISIC G	14.2
% tenure 1 year or less	12.6	% ISIC H	2.4
% tenure 2 to 5 years	16.4	% ISIC I	6.6
% tenure 6 to 9 years	15.1	% ISIC J	6.5
% tenure 10 to 14 years	11.5	% ISIC K	6.9
% tenure 15 years or more	44.4	% Austria	10.9
% firm size less than 50 employees	45.1	% Belgium	8.0
% firm size 50-99 employees	10.5	% France	18.8
% firm size 100-499 employees	18.5	% Germany	16.2
% firm size 500 employees or more	25.9	% Italy	17.6
% tertiary education	19.1	% Spain	18.4
% upper secondary education	43.3	% United Kingdom	10.1
% less than upper secondary education	37.6	Family variables:	
% Legislators, senior officials and managers	9.3	Number of rooms per household member	1.39
% Professionals	7.1	% received an heritage in the last 12 months	1.98
% Technicians and associate professionals	12.8	% housing costs are a heavy burden	19.55
% Clerks	10.6	% housing costs are somewhat a burden	46.45
% Service and shop and market sales workers	5.5	% housing costs are not a problem	33.99
% Craft and related trades workers	31.0	% house owners	71.43
% Plant and machine operators and assemblers	16.0	% tenants	24.59
% Elementary occupations	7.8	% with rent-free accommodation	3.98
% previous unemployment	29.3	% limited in work effort by looking after others	0.95
% permanent contract	89.7	% children aged less than 12 in the household	43.73
% ISIC C+E	2.8	% one person household	5.92
% ISIC DA	4.8	% single parent with one or more children	3.10
% ISIC DB+DC	2.7	% couple without children	14.96
% ISIC DD+DE	4.7	% couple with one child (aged less than 16)	13.06
% ISIC DF-DI	7.9	% couple with two children (all children aged less than 16)	17.51
% ISIC DJ+DK	14.2	% couple three or more children (all children aged less than 16)	5.43
% ISIC DL-DN	11.1	% couple with children (at least one child aged 16 years or more)	33.46
% ISIC F	15.2	% other households	6.56

References

Abowd, J.M. and Kramarz, F. 2003, The cost of hiring and separations, *Labour Economics*, 10, 499-530.

Acemoglu, D. and Pischke, J. 1999a, The structure of wages and investment in general training, *Journal of Political Economy*, 107, 539-572

Acemoglu, D. and Pischke, J. 1999b, Beyond Becker: Training in Imperfect Labour Markets, *The Economic Journal*, 109, F112-F142

Acemoglu, D. and Pischke, J., 2003, Minimum Wage and On-the-Job Training, *Research in Labor Economics*, 23, 159-202.

Arulampalam, W., Booth, A. and Bryan, M., 2004, "Training and the New Minimum Wage", *The Economic Journal*, 114, C87-C94.

Aubert P. and Crépon B., 2004, "La productivité des salariés âgés: une tentative d'estimation", *Economie et Statistique*, 368, 95-119.

Barron, J.M., Berger, M.C. and Black, D.A., 1999, Do Workers Pay for On-the-Job Training?, *Journal of Human Resources*, 34, 235-252.

Bartelsman, E., Scarpetta, S. and Schivardi, F., 2005, Comparative Analysis of Firm Demographics and Survival: Micro-level Evidence for the OECD Countries, *Industrial and Corporate Change*, forthcoming.

Bassanini, A and Brunello, G., 2006, Is training more frequent when the wage premium is smaller? Evidence from the European Community Household Panel, OECD Social Employment and Migration Working Paper, forthcoming, http://bassax.freeyellow.com/Babru_OECD.pdf.

Becker, G., 1964, *Human Capital*, NBER, New York.

Bishop, J.H., 1997, What We Know about Employer-Provided Training: A Review of the Literature, *Research in Labor Economics*, 16, 19-87.

Blundell, R., Dearden, L. and Meghir, C., 1999, Work-related training and earnings, IFS working paper, Institute for Fiscal Studies, London.

Booth, A and Bryan, M., 2005, Testing some predictions of human capital theory: New training evidence from Britain, *Review of Economics and Statistics*, 87, 391-394.

Booth, A. and Zoega, G., 2004, Is wage compression a necessary condition for firm-financed general training?, *Oxford Economic Papers*, 56, 88-97.

Frazis, H. and Loewenstein, M., 1999, Re-examining the Returns to Training: Functional Form, Magnitude, and Interpretation, BLS Working Paper 325.

Gerfin, M., 2004, Firm-Sponsored General Training in Frictional Labour Markets: An Empirical Analysis for Switzerland, IZA Discussion Paper 1077.

Greene Knapp, L., and Seaks, T.G., 1992, An Analysis of the Probability of Default on Federally Guaranteed Student Loans, *Review of Economics and Statistics*, 74, 404-411.

Grossberg, A.J. and Sicilian, P., 1999, Minimum Wages, On-the-Job Training, and Wage Growth, *Southern Economic Journal*, 65, 539-556.

Harvey, A.C., 1976, Estimating Regression Models with Multiplicative Heteroskedasticity, *Econometrica*, 44, 461-465.

Katz, E. and Ziderman, A, 1990, Investment in general training: the role of information and labor mobility, *The Economic Journal*, 100, 1147-58.

Lazear, E.P., 2003, Firm-Specific Human Capital: A Skill-Weights Approach, NBER Working Paper, No. 9679, May.

Lengermann, P., 1999, How Long Do the Benefits of Training Last? Evidence from Long Term Effects across Current and Previous Employers, *Research in Labor Economics*, 18, 439-461.

Loewenstein, M. and Spletzer, J., 1998, Dividing the costs and the returns to general training, *Journal of Labor Economics*, 16, 142-171.

Loewenstein, M and Spletzer, J., 1999, General and Specific Training: Evidence and Implications, *Journal of Human Resources*, 34, 710-733.

Moen, E. and A. Rosen, 2002, Does Poaching Distort Training?, CEPR Discussion Paper, No. 3468, July.

Moulton, B., 1986, Random Group Effects and the Precision of Regression Estimates, *Journal of Econometrics*, 32, 385-397.

Nestler, K. and Kailis, E., 2002, Disparities in Access to Continuing Vocational Training in Enterprises in Europe, Eurostat Statistics in Focus No. 22/2002.

Neumark, D. and Wascher, W., 2001, Minimum Wages and Training Revisited, *Journal of Labor Economics*, 19, 563-595

OECD, 2003, Employment Outlook.

OECD, 2004, Employment Outlook.

Pischke, J.-S. 2001, Continuous Training in Germany, *Journal of Population Economics*, 14, 523-48

Schøne, P., 2004, Why Is the Return to Training So High?, *Labour*, 18, 363-78

Sicilian, P., 2001, On-the-Job Training and Starting Wages, *Journal of Labor Research*, 22, 809-16.

Smith, R. and Blundell, R., 1986, An exogeneity test for a simultaneous equation tobit model with an application to labor supply, *Econometrica*, 54, 679-85

Stevens, M., 1994, A theoretical model of on-the-job training with imperfect competition, *Oxford Economic Papers*, 46, 537-62