Is Training More Frequent When Wage Compression is Higher? Evidence from the European Community Household Panel

Andrea Bassanini
Giorgio Brunello

I Z A

Is Training More Frequent When Wage Compression Is Higher? Evidence from the European Community Household Panel

Andrea Bassanini
OECD

Giorgio Brunello
University of Padova and IZA Bonn

ABSTRACT

Is Training More Frequent When Wage Compression Is Higher? Evidence from the European Community Household Panel[*]

When labor 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 greater than the increase in pay. We propose a novel way of testing this proposition, which exploits the variation in training incidence and in the training wage premium within the European Union. Our results unambiguously show that (general) training incidence is higher in clusters – defined by country, sector, occupation and educational attainment – with a lower training wage premium, measured as the differential between the median wage growth of trained and untrained employees.

JEL Classification: J24, J31, J41

Keywords: training, wage compression, ECHP

Corresponding author:

Giorgio Brunello
Department of Economics
University of Padova
Via del Santo 33
35100 Padova
Italy
Email: brunello@decon.unipd.it

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, if not, which agent (employers or employees) has too little incentive to invest.

According to Becker (1964), when labor 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 for by the worker, who reaps all the benefits from the investment. Hence, only imperfections in other markets (e.g. the capital market) may prevent workers from choosing the optimal amount of human capital. The empirical evidence, however, is difficult to reconcile with Becker's model. 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 [2002]). 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.

Recent theories of imperfect competition in the labor market can explain why employers have an incentive to pay for general human capital. If the market for trained workers is less competitive 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. Under-investment in training may still occur, even in the absence of imperfections in other markets, because current employers typically cannot internalize the benefits from training that will accrue to the worker's future employers (see *e.g.* Katz and Zidermann [1990], Stevens [1994], and Acemoglu and Pischke [1999a]).

The purpose of this paper is to test the empirical relevance of these theories of employer-sponsored training based on labor market imperfections, by estimating the relationship between training incidence and the training wage premium, measured as the

differential between the median (log) wage growth of trained and untrained employees, and clarifying how the latter variable relates to wage compression.

A few recent papers have provided empirical evidence, which is consistent with models of firm-sponsored training based on imperfect competition in the labor 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 NLSY data, 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 [2002] 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).

However, the presence of labor market imperfections is not the only possible explanation of the finding that wages after a training spell grow faster if the worker changes jobs. A feature of the reviewed literature is that it does not control for the reasons why workers quit.[1] It seems rather implausible that a worker be laid-off shortly after having received training. As a consequence, quits occurring after training are likely to be triggered by better outside wage offers, while job changes in the absence of training may include lay-offs. Hence, it cannot be excluded that, when endogenous selection is properly taken into account, the estimated wage gains from switching jobs turn out not significantly different between trained and untrained workers. Additionally, some firms may use subsidized training contracts (such as apprenticeships), to pay employees less than their productivity and provide general

---

[1] The only exception is Blundell, Dearden, and Meghir [1999]. They partially control for mobility by using an Heckman-type correction to deal with the correlation between temporary shocks and training incidence.

4

training, with the implicit agreement that employees will quit at the end of the contract to find a better job.

An alternative way of testing whether wage compression induced by labor market imperfections provides an incentive for firms to pay for general training is to investigate the relationship between the minimum wage and training incidence. The reason is that the minimum wage compresses the lower tail of the wage distribution without necessarily affecting individual productivity. 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 qualify 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, it is not clear what time horizon is appropriate to analyze the effect of institutional changes such as the minimum wage. Last but not least, the degree of imperfection of the labor market might differ across countries.

In this paper we propose an alternative test of the hypothesis that wage compression encourages the provision of general 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 and investigate whether these measures are significantly correlated to general training. When labor markets are imperfectly competitive and training is employer provided, we expect to find a negative correlation between the training wage premium and training incidence. 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, since the (unobserved) productivity premium and the (observed) wage premium are positively correlated. This approach seems to us particularly suitable to testing theories of training based on imperfect competition using European data. Since the migration of labor between EU countries is still limited, we can meaningfully use the country dimension in the definition of clusters, which

---

[2] See Grossberg and Sicilian [1999], Neumark and Wascher [2001], and Acemoglu and Pischke [2003], for the United States, and Arulampalam, Booth and Bryan [2002], for the United Kingdom.

allows us to construct a sample with a large number of clusters and to obtain significant variation in wage compression measures.[3]

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

2 General Training and Wage Compression

In this section we briefly review a simplified version of the model by Acemoglu and Pischke [1999a] — which we take as an archetype of models of general training in imperfectly competitive labor markets — and explain in some detail our empirical approach to testing the relationship between training incidence and wage compression.

2.1 A simplified model of wage compression and training

One implication of removing the assumption of perfect competition is that, when a worker quits, she has to incur a cost — that is, her outside option is below her productivity $f(\tau)$ by the amount $D(\tau)>0$, with $\tau = 1$ indicating that the individual has received general training (while $\tau = 0$ indicates that no training takes place). As a consequence, the employer can retain the worker by offering her a wage below productivity. Labor market imperfections might also reduce the transferability of general training, thereby making the outside option, and thus the actual wage, increase less than productivity in the aftermath of training. When the gap between productivity and the wage increases with training, it can be said that there is "a *compressed wage structure*, since the return to skills for a worker is less than the one prevailing in a competitive labor market." (Acemoglu and Pischke [1999b], p.F120). The fundamental result of the theory of general training in imperfect competition is that the greater the wage compression, the greater the amount of training provided and paid for by the firm (see Acemoglu and Pischke [1999b] for a survey).

We can illustrate this idea more in detail using a simplified version of the model of Acemoglu and Pischke [1999a]. Assume that the current employer wants to retain the trained

---

[3] It must be noticed that a testing approach based on estimating the impact of minimum wages on training can be difficult to implement when continental European countries are added to the sample. In countries such as Austria, Germany and Italy, there are no statutory minimum wages, but wage floors are defined by collective agreements, often at a very disaggregate sectoral level, and it is not easy to collect precise information on them.

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, the expected profits from training and no training are respectively $\pi(1) = f(1) - w(1) - c(1) = (1 - \beta)(f(1) - v(1)) - c(1)$ and $\pi(0) = f(0) - w(0) = (1 - \beta)(f(0) - v(0))$, where c is the training cost and $c(0) = 0$. The employer decides to bear the training costs if and only if $\Delta f - \Delta w > c(1)$ or, equivalently, $\Delta f - \Delta v > c(1)/(1 - \beta)$, where $\Delta f = f(1) - f(0)$ and Δw and Δv are similarly defined.

If training costs vary across workers, the more compressed the wage structure (the greater $\Delta f - \Delta w$), the larger the number of employees for which employers are ready to pay for training. 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 labor market frictions, D is equal to zero, $f(\tau) = w(\tau) = v(\tau)$ and the employer never pays for general training.

With a similar argument it can be shown that, even when the 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.[4] In the special case of perfect competition, total investment in training is paid by the employee and is increasing with the wage premium.

To summarize, we can draw from the above discussion the following empirical implications: 1) wage compression in imperfect labor markets increases the provision of general training by firms; and 2) a negative relationship between the wage gain and general training is not consistent with competitive theories of the labor market.

2.2 The Empirical Strategy

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. For a given individual i, the probability that she receive employer – paid training is:

[4]In this respect, Acemoglu and Pischke's results are similar to those by Hashimoto in his model of firm – specific training, where the parties share the costs and benefits of the investment (see Hashimoto [1981]).

$$\text{Pr}\,ob\{\tau_i = 1\} = \text{Pr}\,ob\{\Delta f_i - \Delta w_i > c_i(1)\} \qquad [1]$$

Taking into account that the training that is reported to be employer-sponsored might be indirectly paid by the worker (at least in part) by accepting lower wages during training, equation [1] can be generalized as:

$$\text{Pr}\,ob\{\tau_i = 1\} = \text{Pr}\,ob\{\gamma\Delta f_i + \sigma\Delta w_i + \delta c_i(1) > 0\} \qquad [2]$$

with $\gamma > 0$, $\sigma \geq -\gamma$ and $\delta < 0$, in the case of the Acemoglu and Pischke model,[5] and $\gamma = 0$, $\sigma \geq 0$ and $\delta < 0$, in the case of the Becker model[6].

We do not observe training costs directly but assume that they vary with firm size, individual age, tenure in the job, educational attainment, type of labor contract, previous unemployment record, sector of activity, occupation and country. We use these variables plus a standard error term to approximate training costs in equation [2].

Without cross-country comparable matched employer-employee datasets (see Bartelsman, Scarpetta and Schivardi [2003]), there are no 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, rejecting the Becker model if $\hat{\sigma} < 0$ (with ^ standing for estimate). Since productivity and wage premia are positively correlated, the omission of the former should bias our estimates of σ against finding a negative relationship between the wage premium Δw and training incidence. Hence, our estimate of σ can be interpreted as a lower bound estimate of the impact of wage compression on employer sponsored training.

It is reasonable to expect that employers infer the training wage premium Δw from the actual wage distribution. This distribution varies across countries. Given this variation, it is tempting to look at the correlation between country - specific measures of wage dispersion and training incidence (see *e.g.* OECD [1999] and Brunello [2001]). This approach, however,

---

[5] The parameter σ can be strictly higher than $-\gamma$ if employer-provided training is partly financed by the employee.

[6] Strictly speaking, the simple fact of observing some employer-paid general training would be inconsistent with competitive theories of the labor market, unless the employee fully compensates the employer by accepting lower wages during training spells. However, the evidence that this occurs is not conclusive. (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).

has two problems. First, there are other country - specific policy and institutional variables beside wage compression which can affect training. These effects can be adequately captured by using unrestricted country dummies. Second, the heterogeneity of individuals and jobs suggests that the relevant measure of the wage premium should not refer to the entire wage distribution in a country.

To illustrate the second point, consider the decision to invest in the training of a professional with a college degree who is employed in the service sector. Since earnings vary with education, occupation and sector, the portion of the wage distribution associated to employees with lower education and in completely different professions and sectors, is unlikely to be informative. Conversely, as a reasonably close proxy of the training wage premium for this professional, we can use the average training wage premium for employees in the same broad profession and sector, and with the same level of educational attainment.

A natural way to measure the training wage premium is therefore to partition individuals into relatively homogeneous clusters and to compute for each cluster the average training premium, which in its simplest form can be defined as the difference between the average wages of trained and untrained individuals. In order to better control for individual heterogeneity within clusters, however, we compute the premium as the difference between the average wage growth rates of those who reported to have received training in the period covered by the survey and of those who did not. We call this measure the wage growth differential. Defining ΔW_c as the wage growth differential in cluster c, we specify the empirical relationship linking it to the probability that individual i receive training as follows:

$$\Pr ob\{\tau_{ic} = 1\} = \Pr ob\{\alpha + \beta X_{ic} + \sigma\Delta W_c + \varepsilon_{ic} > 0\} \qquad [3]$$

where X is a vector of controls, including age, country, sector, firm size, occupation, tenure, education, previous unemployment and type of contract, and ε is a standard error term. If we find that the variable ΔW_c attracts a negative and significant coefficient in model [3], we can conclude that higher wage compression increases the probability of training, as predicted by the theory of training in imperfect labor markets.

A potential objection to our empirical proxy of the wage premium is that it can be endogenous. First, both the probability of training and individual wage growth are likely to be correlated with unobserved ability. Assume that a shock occurs and that an additional individual receives employer–provided training. We can expect that this additional worker

9

who receives training has lower ability than those who would have been trained even in the absence of the shock. If individual wage growth is correlated with ability, the mean wage growth of those who receive training and of those who do not will fall. However, since training incidence is generally low we can also expect that a marginal increase in the incidence of training has little impact on the mean wage growth of those who do not receive training. Hence, the wage growth differential will fall as a result of the greater training incidence induced by the shock and a negative bias will emerge. Such a bias could erroneously point to a positive correlation between training and wage compression even when labor markets are competitive. To solve this problem we use median wage growth rates instead of means. If the effect of ability on individual wage growth is positive and separable from that of training,[7] the relative movement of the median wage growth rates for the two groups does not depend on the size of each group (as in the case of the mean), but only on the concentration of each wage growth distribution (of those who receive training and those who do not) in the neighborhood of the medians. To the extent that the distribution of wage growth rates of workers who receive training is not less concentrated in the neighborhood of its median, the latter does not shift downwards more than the median wage growth of those who do not receive training and, hence, no downward bias emerges.[8] The absence of systematic differences in the degree of concentration of the two distributions around their medians can be easily tested (as we do) by looking for significant differences across the two groups of interquantile differences around the median.

An additional source of endogeneity is the possibility that diminishing returns to training induce reverse causality. If there are diminishing returns and trained and untrained workers are imperfect substitutes in production, we expect that the greater the stock of training the lower the wage gain from training. However, this possibility is less problematic than it might seem at first glance. To fix ideas, think of a standard augmented Solow growth model and, to simplify, assume that training is the only source of accumulation of human capital. In the absence of shocks and assuming that steady states are similar across clusters (or equivalently that cross-cluster differences in the steady states are captured by our control variables), the aggregate growth of human capital per worker is a decreasing function of the

---

[7] This assumption is standard in the literature on returns to training discussed above. However, in the sensitivity analysis, we will also consider alternative measures that allow relaxing this assumption.

[8] Another advantage of using medians instead of means is to reduce the weight of outliers.

average stock of human capital.[9] It follows that diminishing returns to training imply, at the aggregate level, not only a negative relationship between the stock of human capital and the training wage premium but also a positive relationship between the premium and the growth rate of human capital per worker. To the extent that obsolescence is relatively low, training incidence (the aggregate equivalent of our dependent variable) is closely related to the growth rate of average human capital in the neighborhood of the steady state and, therefore, is negatively related to the average stock of human capital. By transitivity, training incidence turns out to be positively correlated to the wage gain from training, which at worst would bias upwards our estimate of σ.[10]

Nevertheless, 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 the wage growth differential but is independent from training incidence conditional to the wage growth differential.

Our selected instrument is ΔA_c, 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). The presence of diminishing returns implies that changes in the stock of trained employees reduce the wage of trained workers relative to the wage of untrained workers but have no significant influence on characteristics such as relative age. Since younger workers are likely to be trained first, we expect that the greater the incidence of training, the greater the age of workers who receive training and of those who do not. The difference between the median log ages will not be affected, however, if the 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. If this is the case, as we show at the end of the next section, diminishing returns affect relative prices ΔW_c without affecting ΔA_c.

---

[9] This is true not only in the case of the augmented Solow model (see the Appendix for a formal proof), but also in the case of the Uzawa-Lucas model as well as for plausible parameters of its generalisation (see Barro and Sala-i-Martin [1995] for a numerical analysis)

[10] Note that the assumption of low obsolescence is standard in panel data studies of the returns to training. In fact, it is generally assumed that individual training stocks do no depreciate as time goes by so that training history before the sample can be eliminated through fixed effects.

3. Data

We use individual data from the December 2001 release of the *European Community Household Panel* (ECHP), which is a longitudinal survey modeled on the US Panel Study of Income Dynamics (PSID). 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. The data collection for the ECHP is made at the national level by National Data Collection Units (NDUs), with Eurostat providing centralized support and coordination.

We use training data from the 1996 wave of the ECHP for 7 countries[11] and restrict our attention to male employees (excluding apprentices), aged from 30 to 60 years[12] and working full-time in the non-agricultural private sector, excluding sectors where non-profit organizations account for a non-negligible fraction of employment.[13]

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 and whether it was paid for or provided by the employer). If more than one concurrent course are involved, only 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-

---

[11] The choice of the survey year and the country sample is dictated by data availability (see appendix). Included countries are Austria, Belgium, France, Germany, Italy, Spain, and the United Kingdom. In the sensitivity analysis we explore the robustness of results to changes in the survey year.

[12] 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). We also perform a sensitivity analysis by varying the age of the sample.

[13] This choice of sectors is due to the fact that we cannot distinguish between business enterprises and non-profit organizations. The model described in the previous section may not apply to the latter. Included sectors are mining, manufacturing, utilities, construction, wholesale and retail trade, hotels and restaurants, transport, storage and communication, financial services, real estates, renting and business activities (corresponding to ISIC Rev. 3 codes C to K).

country variation. The country ranking that is shown in the table is similar, however, to what emerges from other cross-country European surveys, such as Eurostat's 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;[14] 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. Categories a and b refer unambiguously to off-site training courses; category c refers to courses partly taken off-site and partly received in the workplace; and category d refers to workplace training only.[15] Following Loewenstein and Spletzer (1998, 1999) and OECD (2003) we use the distinction between off-site and workplace training to proxy the distinction between general and firm-specific training, which is key for our analysis.

[14] To illustrate the difference between these two categories, we can take the example of the UK, where higher BTEC, BEC, TEC, HNC, HND and teaching and nursing qualifications without degree are in the first category, while ordinary or general BTEC, RSA and City and Guilds or equivalent are in the second. Admittedly, the distinction between these two categories is somewhat arbitrary.

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

Assuming that all off-site training is at least partly general we treat categories a, b and c as "general training" and category d as "firm-specific training". Nevertheless, since training spells falling under category c are likely to generate both general and specific skills, 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".

The questionnaire also asks individuals whether the vocational training course was paid for or organized by the employer. As expected, 95% of the courses that we classify as firm-specific training are paid for or organized by the employer. In line with the existing literature, the data also show that 79% of the courses which we classify as general are paid for or organized by the employer.[16] Unfortunately, this question must be used with caution for three reasons. First, we cannot establish how courses which are *partly* paid by the employer are classified. Second, it might be difficult for the worker to assess the employer's contribution, because the latter might not be explicit (see for example Booth and Bryan, 2002). Last but not least, workers might indirectly pay for employer-sponsored training by accepting lower wages. For these reasons, in the analysis that follows we focus both on total training, independently of its financing, and on employer-provided training.

As said in the previous section, we compute our empirical proxy of the training wage premium (the wage growth differential) by cluster. We define clusters by four dimensions: the country, the educational attainment (*less than upper secondary*, *upper secondary*, *more than upper secondary*)[17], the broad group of sectors (*mining, manufacturing, utilities and construction*, and *services*)[18], and the broad occupational group (*high-skilled occupations* and *medium- and low-skilled occupations*).[19] By so doing we obtain 12 clusters per country, but some of them are empty.[20]

---

[16] 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; and 92% for specific vocational training within a system providing both work experience and complementary instruction elsewhere.

[17] Corresponding to ISCED categories, 0-2, 3, and 5-7, respectively.

[18] Corresponding to ISIC Rev. 3 industries C to F, and G and above, respectively.

[19] 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).

[20] 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.

The wage growth differential is obtained by computing for each cluster the difference between the median growth rates of gross hourly wages of employees who reported to have received training in the period covered by the survey and 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 ECHP data start in 1994, 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.

Individual wage growth is computed as the logarithmic difference between the wage reported in the current wave and the wage reported two waves earlier by the same individual.[21] Since our measure of the wage growth differential should capture the cluster-specific premium to completed training spells, we exclude from this calculation all individuals who reported to be still in training at the time of the survey. In the case of Austria, for which no data are available in 1994, we replace the 1996 wage growth differential with that of 1997. 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.[22] Finally, in order to reduce the weight of outliers, we drop clusters with 30 observations or less and with less than 5 reported training events. This leads us to 48 valid clusters.[23]

Table 2 shows how the wage growth differential varies by country, educational attainment, occupation, and sector of activity. On average, the wage growth differential is close to zero or even negative in Austria, Germany and France. This should not surprise the reader: indeed, the literature fails to find statistically significant training premia in at least two of these countries once time-invariant heterogeneity is controlled for (see Goux and Maurin [2000] and Pischke [2001] for France and Germany, respectively). Conversely, it is relatively

---

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 the sensitivity analysis, we explore the effect of different partitions by varying the grouping of occupations.

[21] Remember that training reported in a given year might have been taken in January of the year before.

[22] We check, however, the robustness of our results by repeating the analysis using wage growth differentials computed only for those sectors that we use in the final regressions.

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

high in the other four countries, but in line with available country estimates.[24] The wage growth differential is also relatively high among medium and low skilled workers as well as in manufacturing. Furthermore, no monotonic relationship between educational attainment and the wage growth differential emerges from the data. Finally, no clear pattern of bivariate association between training incidence and wage growth differential can be noticed at first glance (the correlation coefficient between these variables at the cluster level is 0.01).

Table 2: Average wage growth differentials by country and other characteristics

Country	Wage growth differential	Other Characteristics	Wage growth differential
Austria	0.006	Less than upper sec. education	0.040
Belgium	0.048	Upper secondary education	0.001
France	-0.004	More than upper sec. education	0.021
Germany	0.009	Mining, manuf. and utilities	0.033
Italy	0.029	Services	-0.006
Spain	0.033	High-skilled occupations	-0.005
United Kingdom	0.026	Medium/low-skilled occupations	0.027
Average	0.019		

Note: The wage growth differential refers to 1996 for all countries except for Austria, for which it refers to 1997, since no data for that country are available for 1994. Averages are weighted by cluster size.

In the previous section we argued that the use of medians rather than means in the computation of the wage growth differential could solve the endogeneity problem induced by the correlation of training and wage growth with unobserved ability. We have also 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. These statements are correct only if there is no systematic difference in the concentration of the wage growth and age distributions in the neighborhood of the median wage growth and median age of those who receive 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 wage growth (resp. age) distributions and by checking whether their cross-cluster averages are systematically different. We performed this test for

---

[24] For instance, Booth and Bryan (2002), use three waves of the BHPS and control for time-invariant heterogeneity by using fixed effects. They find that the wage gain from training participation in the United Kingdom is about 2.4%, except in the case of training received at previous employers, for which the estimate is much higher, but which, nonetheless, accounts for less than 10% of training events.

16

the difference 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 our choice of medians rather than means and the validity of ΔA as an instrument.

Table 3: Tests of differences in the concentration around the median

Interquantile difference	ΔW_c	ΔA_c
P55-P45	0.003 (*0.13*)	-0.012 (*0.12*)
P60-P40	-0.006 (*0.15*)	-0.008 (*0.10*)
P70-P30	-0.022 (*0.20*)	-0.013 (*0.16*)

Note: The table reports the cross-cluster average differences between a given interquantile difference for the wage growth (*resp.* age) distribution of those who received training and the same interquantile difference for the wage growth (*resp.* age) distribution of those who did not receive training. ΔW_c indicates the wage growth differential, ΔA_c indicates the log age differential. Averages are weighted by cluster size. Standard errors in parentheses.

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

4 The Empirical Results

4.1 Main Results

Our empirical analysis is in three steps, following the organization of the questions on training in the ECHP questionnaire. First, as explained in the previous section, individuals are asked whether they have been in vocational education or training at any time since January of the year before the survey. As a preliminary step, we define the dummy variable *T* as equal to 1 in the event of training and to 0 otherwise, and estimate a probit model of training incidence, with the cluster specific wage growth differential ΔW_c as the indicator of the training wage premium. By so doing, we pool together general and firm specific training and investigate the correlation between training incidence and the training wage premium. Yet, a negative relationship between these two variables can be consistent with Becker if it is driven by firm specific training. However, in the survey, individuals who positively answer the

question on vocational training are asked to indicate the type of the last course they attended and who financed it. We group therefore the alternative options concerning course type into three categories, and define the dependent variable *TYPE* as equal to 0 in the case of no training, to 1 in the case of general training, and to 2 in the case of firm - specific training. Since these categories are not naturally ordered, but are mutually exclusive by design, we estimate a multinomial logit model, using no training as the baseline category, and the cluster-specific wage differential ΔW_c as the indicator of the training wage premium. Finally, we use the information on financing to define another dependent variable (*TYPEM*) which takes value 0 in the case of no training, 1 in the case of employer-sponsored general training, 2 in the case of employer-sponsored firm - specific training, and 3 in the case of non-sponsored training.[25] Again, we estimate a multinomial logit model for this variable, using no training as the baseline category, and the cluster-specific wage differential ΔW_c as the indicator of wage dispersion.

In the empirical specification we include the following variables: individual age, country, sector and occupation dummies, two education dummies, one for attained primary and lower secondary education and the other for attained upper secondary education, four tenure dummies, three firm size dummies, a dummy for permanent labor contracts and a previous unemployment dummy (*U*=1: presence of at least one unemployment spell since 1989, 0 otherwise). We take explicitly into account the fact that the training variables *T*, *TYPE* and *TYPEM* 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]).

We start the presentation of our results with the estimates of the probit model in Table 4. We present results from three different specifications. The first includes only basic controls (age, education, occupation, sector and country; Column 1), the second adds firm size to basic controls (Column 2), and the third includes an extended set of controls (tenure, permanent job and previous unemployment; Column 3). All the coefficients are rather stable across specifications. Our estimates show that the coefficient associated to the wage growth differential ΔW_c is negative and statistically significant.[26] The estimated impact, however, is

---

[25] General and firm-specific non-sponsored training are grouped together due to the small number of non-sponsored training events in our sample.

[26] Table 3 reports results obtained using ECHP personal weights. Un-weighted estimates are similar. We obtain also similar results when using means instead of medians.

rather small: conditional on the extended set of controls, an increase in the wage growth differential from its sample mean (0.019) to the average value for the UK (0.026), equivalent to a 36.8 percent increase, would reduce the probability of training by 0.478 percent.[27]

We test the weak exogeneity of the wage growth differential ΔW_c by following the procedure suggested by Smith and Blundell [1986]. In the first step, we regress the wage differential on the cluster – specific instrument ΔA_c and on country, occupation, education and sector dummies.[28] We compute the residual from the first step regression, add this variable to the probit specifications and test whether the associated coefficients are statistically different from zero (Columns 4 to 6 in Table 4). Consistent with our expectations (see Section 2), these coefficients are always insignificant, and therefore we cannot reject the null hypothesis of weak exogeneity of the wage growth differential ΔW_c in the training regression.[29]

The estimates in Table 4 do not explicitly distinguish between general and firm – specific training. We do so in Table 5, which reports the relevant coefficients from the estimation of the multinomial logit models for both total and employer-sponsored training. Each model is estimated subject to two different definition of general training (that we called "extensive" and "restrictive" in the previous section). The table shows that the estimated coefficient attracted by our measure of ΔW_c is always negative and significant in the case of general training and close to zero and not statistically significant in the case of firm – specific training. These estimates imply that the elasticity of the odds ratio of general training with respect to no training, evaluated at the sample mean value of ΔW_c, ranges between -0.064 and -0.090, depending on whether we consider total or employer-provided training and the extensive or restrictive definition of general training.[30]

We ask whether these elasticities are larger in absolute value than the elasticities obtained by pooling together general and firm specific training. To answer this question, we estimate a logit model of total training T and compute the elasticity of the odds ratio of

---

[27] The elasticity of the probability of training with respect to ΔW_c, evaluated at the sample means, is equal to -0.013.

[28] The F test on the significance of the instrument is 4.36 (statistically significant at the 5% level).

[29] Once the residual is added to the specification, the coefficient of the wage growth differential is only significant at the 10% level. However, the inclusion of the residual leads to inefficient estimates under the hypothesis of weak exogeneity.

[30] The coefficients of the multinomial logit can be interpreted as changes in the log odds ratio induced by a one unit change in the explanatory variable.

19

training with respect to no training. We find that this elasticity is equal to -0.022, less than one third of the estimates based on the multinomial logit. This result is consistent with our estimates in Table 5, showing the small and statistically insignificant effect of ΔW_c on firm specific training.

We interpret these findings as clearly inconsistent with competitive theories of the labor market, which predict a positive correlation between the training wage premium and the incidence of general training, and as consistent with the view that higher wage compression in imperfect labor markets helps firms to organize and pay for general training.

Table 4. *Probits for total training*. Dependent variable: training participation (*T*) in 1996.

	(1)	(2)	(3)	(4)	(5)	(6)
Age	-.017***	-.020***	-.022***	-.017***	-.020***	-.022***
	(.003)	(.003)	(.003)	(.003)	(.003)	(.004)
Educational attainment						
(ref: More than upper secondary)						
Upper secondary education	-.278***	-.289***	-.320***	-.295***	-.313***	-.340***
	(.095)	(.105)	(.103)	(.098)	(.107)	(.104)
Less than upper secondary education	-.521***	-.483***	-.533***	-.533***	-.500***	-.546***
	(.105)	(.121)	(.121)	(.105)	(.126)	(.120)
Previous unemployment spell			-.144**			-.144**
			(.061)			(.061)
Tenure						
(ref: 0-1 years)						
Tenure: 2-5 years			-.204*			-.202*
			(.122)			(.122)
Tenure: 6-9 years			-.213**			-.213**
			(.105)			(.105)
Tenure: 10-14 years			-.108			-.108
			(.099)			(.099)
Tenure: 15 years or more			-.174			-.173
			(.109)			(.118)
Permanent job			.253***			.253***
			(.097)			(.097)
Firm size						
(ref: 1-49 employees)						
Firm size: 50-99		.216***	.206**		.217***	.207**
		(.082)	(.082)		(.082)	(.082)
Firm size: 100-499		.411***	.362***		.414***	.365***
		(.072)	(.068)		(.071)	(.068)
Firm size: 500 or more		.529***	.503***		.531***	.505***
		(.073)	(.080)		(.073)	(.080)
Wage growth differential	-.543**	-.538**	-.603**	-1.023	-1.213*	-1.178*
	(.229)	(.262)	(.246)	(.660)	(.682)	(.616)
Residual from first stage regression				.517	.727	.623
				(.733)	(.763)	(.682)
Number of observations	6146	5968	5741	6146	5968	5741
Pseudo R-squared	.150	.169	.177	.150	.169	.177

Note: Cluster adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Additional controls not in the table are country, occupation and sector dummies. In the first stage the wage growth differential is regressed on the median log age differential, country, occupation, education and sector dummies. ***, **, *: significant at the 1%, 5% and 10% level of confidence, respectively.

21

Table 5. *General and specific training.* Trinomial and quadrinomial logits. Dependent variable: Total and employer provided training in 1996.

	(1)	(2)	(3)	(4)
	Extensive definition		*Restrictive definition*	
Total training (TYPE)	*General*	*Firm-specific*	*General*	*Firm-specific*
Basic controls	-3.388*** (.629)	.741 (.631)	-4.251*** (.797)	-.802 (.602)
Extended controls	-3.488*** (.598)	-.775 (.720)	-4.314*** (.799)	-.859 (.686)
Employer provided training (TYPEM)				
Basic controls	-3.630*** (.917)	-.640 (.684)	-4.551*** (1.151)	-.766 (.686)
Extended controls	-3.853*** (.888)	-.776 (.765)	-4.749*** (1.092)	-.938 (.762)

Note: In the first two rows the dependent variable is the type of training (TYPE), while in the other two rows the dependent variable is the type of employer-sponsored training (TYPEM). Cluster adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Basic controls are age, country, education, occupation and sector dummies. Extended controls add tenure, firm size, permanent job status and previous unemployment dummies. The extensive definition of general training is used in Columns (1) and (2). The restrictive definition of general training is used in Columns (3) and (4). ***: significant at the 1% level of confidence.

4.2. Robustness and Extensions

In this subsection we check the sensitivity of our results to a number of changes in the definition of the sample and key variables used in the main regressions. We start by experimenting with alternative measures of the wage differential. In subsection 2.2 we have argued that medians should be preferred to means in the computation of the wage growth differential in order to lessen the endogeneity bias induced by the correlation between ability, training and wage growth. In the previous section we have tested whether our measure of the training wage premium was weakly exogenous. In this section we offer two alternative measures of the premium which are less precise but at the same time less likely to be affected by the correlation between training and ability. For each cluster, we take the difference between the maximum wage growth of those who received training and the minimum wage growth of those who did not (Maxmin herafter) and the difference between the maximum wage growth of those who received training and the maximum wage growth of those who did not (Maxmax hereafter). The logic is as follows: if the correlation between training, ability and wage growth is strong, then we can assume that the most able worker in a cluster will be

22

the first to be trained and will have the greatest wage growth, while the least able worker will have the lowest wage growth and will never be trained, no matter how favorable the shock is. Under these assumptions, Maxmin will be unaffected by training incidence. Alternatively, since the more favorable the shock, the greater the size of the group of those who receive training and the smaller the size of the group of those who do not, by definition of maximum, a positive training shock can only increase (or leave unchanged) the maximum of the former group and reduce (or leave unchanged) the maximum of the latter. Therefore Maxmax will be at most positively affected by training incidence, inducing an upward bias in our estimates.

Table 6 shows the estimates obtained by substituting these measures of the wage growth differential in the trinomial and quadrinomial logit models of total and employer-sponsored training. In both cases the estimated coefficients of ΔW_c are smaller than those reported in Table 6, but remain with a negative sign and statistically significant.[31]

Table 6. *Maxmin and Maxmax.* Trinomial and quadrinomial logits. Dependent variable: Total and employer provided training in 1996.

|	(1)	(2)
	Maxmin	*Maxmax*
Total training (TYPE)		
Extensive definition	-.526*** (.090)	-.184*** (.062)
Restrictive definition	-.540*** (.140)	-.217** (.089)
Employer provided training (TYPEM)		
Extensive definition	-.526*** (.128)	-.213*** (.083)
Restrictive definition	-.466** (.209)	-.260* (.152)

Note: In the first two rows the dependent variable is the type of training (TYPE), while in the other two rows the dependent variable is the type of employer-sponsored training (TYPEM). Only estimated coefficients of the Maxmin and Maxmax measures of the wage growth differential for general training are reported (only in the case of employer-sponsored training for TYPEM). Cluster adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Other controls are age plus country, education, occupation, sector, tenure, firm size, permanent job status and previous unemployment dummies. The extensive definition of general training is used in the first and third row. The restrictive definition of general training is used in the second and fourth row. ***, **, *: significant at the 1%, 5%, 10% level of confidence, respectively.

The small number of countries in our sample (7) suggests the possibility that our results can be driven by a single country. We assess the robustness of our findings to variations in country coverage by eliminating one country at a time and re-estimating our

---

[31] The fact that the estimates using Maxmin are smaller should not surprise the reader because of the larger measurement error, which biases the estimated coefficient towards zero.

models. Chart 1 reports the results of this sensitivity analysis for the estimated coefficients of the median wage growth differential in the case of employer-sponsored general training, using the extensive definition of general training (with extended controls only). It turns out that coefficient estimates are substantially robust to variation of country coverage. We obtain similar results when we do not distinguish the source of financing.

Chart 1. **Sensitivity to variation of country coverage**

Employer-sponsored general training (extensive definition): estimated coefficient of the wage growth differential
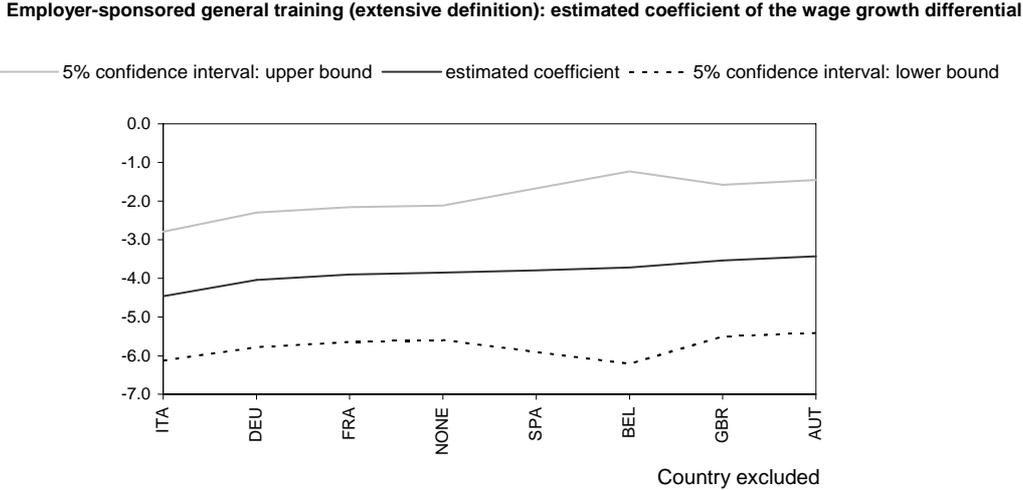
The figure shows central estimates and confidence intervals obtained by re-estimating the quadrinomial logit model (with extended controls) after excluding one country at a time from the sample. NONE identifies the specification with all 7 countries for the purpose of comparison.

In our preferred sample we have excluded workers aged 29 years or less to improve cross-country comparability. In Table 7, we use the specification with extended controls and the extensive definition of general training and change the age range of the sample (accordingly, we also vary the sample on which the wage growth differential is computed). Our results are confirmed at standard confidence levels. When young workers are added to the sample, estimated coefficients maintain their negative sign, albeit with a smaller magnitude.

Table 7. *Sensitivity to variation of the age of the sample.* Trinomial and quadrinomial logits. Dependent variable: Total and employer provided training in 1996.

	(1)	(2)
Age group	*Total training*	*Employer provided training*
Age 20-40	-2.586*** (.685)	-4.861*** (.867)
Age 20-60	-1.993*** (.547)	-2.715*** (.669)
Age 40-60	-3.941*** (.813)	-4.568*** (.591)

Note: In Column (1) the dependent variable is the type of training (TYPE), while in Column (2) the dependent variable is the type of employer-sponsored training (TYPEM). Only estimated coefficients of the median wage growth differential for general training are reported. General training is defined according to the extensive definition. Cluster adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Other controls are age plus country, education, occupation, sector, tenure, firm size, permanent job status and previous unemployment dummies. ***: significant at the 1% level of confidence.

We also verify whether our results can be attributed to the selected year (1996) by re-estimating the baseline specification on data from the 1997 wave (1995-1997 for the wage growth differential). The resulting estimates must be interpreted with caution, however, since the 1997 sample is much smaller and includes only four countries and 19 clusters (see the Appendix). As shown in Table 8, the estimated coefficient of the wage growth differential for total general training remains negative, but small and statistically insignificant. Nevertheless, and more importantly, the estimate for employer-sponsored general training is significant at the 10% level. Similar, but more significant, results are obtained if we replace contemporaneous wage growth differentials with those obtained from a different wave.

In Table 9 we consider a number of other robustness checks using the specification with extended controls and the extensive definition of general training. First, we try to better control for composition effects by adding to our regressors country by sector dummies and the employment rate in 1996 by country and educational attainment (which we matched to our sample using data from the European Labour Force Survey).[32] The inclusion of these variables, both separately and combined, does not affect our results in a significant way.

---

[32] The reason to include country by industry dummies is that the composition of each sector varies by country, which could bias our estimates if abler workers flow into sub-sectors where training opportunities and wage prospects are better. We include the employment rate, which varies by country and educational attainment, because greater statutory or contractual minimum wages may reduce employment among the untrained (particularly among the least able workers) without affecting the number of workers who are trained, thereby driving up training incidence among the employed and driving down the wage growth differential (we also experimented with education by country dummies with similar results).

Table 8. *Sensitivity to variation of the survey year.* Trinomial and quadrinomial logits. Dependent variable: Total and employer provided training.

	(1)		(2)	
	Total training		*Employer provided training*	
A. Wave 4 (1997)	-.141	(.575)	-1.205*	(.686)
B. Wave 4 (1997), ΔW from wave 3 (1996)	-3.990***	(1.246)	-4.896***	(1.391)
C. ΔW from wave 4 (1997)	-.397	(.552)	-2.173**	(1.004)

Note: In Column (1) the dependent variable is the type of training (TYPE), while in Column (2) the dependent variable is the type of employer-sponsored training (TYPEM). Only estimated coefficients of the median wage growth differential for general training are reported (only in the case of employer-sponsored training for TYPEM). General training is defined according to the extensive definition. Cluster adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Other controls are age plus country, education, occupation, sector, tenure, firm size, permanent job status and previous unemployment dummies. In Row A the baseline specification is estimated using data from 1997 (1995-1997 for the wage growth differential); in row B the baseline specification is estimated using data from 1997 but from 1994-1996 for the wage growth differential); in row C the baseline specification is estimated using data from 1996 but from 1995-1997 for the wage growth differential). ***, **, *: significant at the 1%, 5%, 10% level of confidence, respectively.

Next, we modify cluster definitions by dividing the group of medium/low skilled workers into low skilled and medium skilled workers. This is equivalent to excluding the former category, since training incidence is small in that group (see Table 1). The same occurs if we separate clerks from all other medium/low skilled occupations, since the size of the former group is small. We also consider a split between white collars and sales workers, on the one hand, and other blue collars, on the other hand, as well as another split where craft and related trades workers are switched to the former group. Although sample size vary because of the threshold limits of more than 30 workers per cluster (of which at least 5 receiving training), no important difference from the baseline specification emerges, particularly in the case of employer-sponsored general training.

We check whether our results are affected by the fact that sectors where non-profit organizations are not important are included in the computation of the wage growth differential but not in the final regressions. Reassuringly, as shown in Table 9, excluding these sectors from the computation yields no appreciable difference in the estimated parameters. We also lift the size threshold of 30 observations per cluster (of which at least 5 receiving training). Estimated coefficients remain significant, although their absolute value is smaller (due to the greater measurement error induced by small clusters and/or clusters with few training events).

Table 9. *Other robustness tests.* Trinomial and quadrinomial logits. Dependent variable: Total and employer provided training.

	(1)	(2)
	Total training	*Employer provided training*
A. Country X sector dummies	-4.265*** (.695)	-5.741*** (.969)
B. Employment rate	-3.710*** (.666)	-4.389*** (.952)
C. Country X sector dum. + empl. Rate	-4.543*** (.755)	-6.397*** (1.064)
D. No low skilled	-3.003*** (.686)	-3.174*** (.970)
E. No clerks	-2.445*** (.756)	-2.746** (1.107)
F. Three occupational groups (1)	-2.930*** (.581)	-2.978*** (.912)
G. Three occupational groups (2)	-2.018** (.792)	-2.648** (.947)
H. No non-market sectors to compute ΔW	-3.599*** (.640)	-4.298*** (1.003)
I. No size thresholds	-1.283*** (.459)	-1.301** (.516)
J. ΔW computed for general training only	-1.520*** (.466)	-1.714*** (.309)

Note: In Column (1) the dependent variable is the type of training (TYPE), while in Column (2) the dependent variable is the type of employer-sponsored training (TYPEM). Only estimated coefficients of the median wage growth differential for general training are reported (only in the case of employer-sponsored training for TYPEM). General training is defined according to the extensive definition. Cluster adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Other controls are age plus country, education, occupation, sector, tenure, firm size, permanent job status and previous unemployment dummies. In Row A country by sector dummies are added to the baseline specification; in Row B the employment rate (by country and education) is added to the baseline specification; in row C both country by sector dummies and the employment rate are added to the baseline specification; in row D elementary occupations are excluded; in row E clerks are excluded; in row F the group of medium/low-skilled occupation is divided into two groups (clerks, service workers and shop and market sales workers and other medium/low skilled workers) and clusters are defined accordingly; in row G craft and related trades workers are switched from other medium/low skilled workers to clerks, service workers and shop and market sales workers; in row H sectors with ISIC Rev.3 code greater than K are not used in the computation of the wage growth differential; in row I the size threshold (30 observations per cluster, of which at least 5 receiving training) is removed; in row J the wage growth differential is computed by excluding those who reported specific training (according to the extensive definition of general training). ***, **, *: significant at the 1%, 5%, 10% level of confidence, respectively.

It might be argued that the relevant training wage premium for our analysis should refer to general training only. With ECHP data, however, it is not advisable to compute training premia separately for general and firm-specific training because individuals report the type of training only for the last course. Since workers reporting specific training may have taken general training as well (or vice versa), type-specific training wage premia could be severely affected by measurement error. Nevertheless, we check the robustness of our results to the introduction of type-specific training wage premia, using the extensive

definition of general training. As shown in Table 9, the estimated impact of the premium on general training remains negative and statistically significant.[33]

Finally, we explore the effect of the wage growth differential on each separate category of training. In order to do so, we define two new dependent variables: *DTYPE*, which takes value 0 in the case of no training, 1 in the case of third level qualification, 2 in the case of specific vocational training at a vocational school or college, 3 in the case of dual systems with both workplace and school training and 4 in the case of workplace training only; and *DTYPEM* which is equal to *DTYPE* for employer-sponsored training and to 5 when training is not employer-provided. Since these categories are not naturally ordered, but are mutually exclusive by design, we estimate two multinomial logit models with 5 and 6 categories, respectively, using no training as the baseline category.

Needless to say, the results of this exercise can only be exploratory, since the small number of observations corresponding to certain training categories prevents us from undertaking a rigorous analysis at this level of aggregation. Table 10 shows the estimated coefficients associated to the wage growth differential. Reassuringly, point estimates are much higher for training which is thoroughly off-site - categories 1 and 2 - than for other types of training, although standard errors for third level qualification training are large, due probably to the small number of events in this category.

Overall, these sensitivity exercises clearly suggest that our key findings in the previous sub-section are robust to changes in the definition of the key variables and model specification.

---

[33]The fact that point estimates are smaller can be attributed to the greater measurement error. We also experiment with the simultaneous inclusion in the specification of the differential computed for general training only and the differential computed for specific training only. In the case of employer-sponsored general training, the resulting estimates are -1.704, with a standard error .259, and -.278, with a standard error of .264, for the general and specific training premia, respectively.

Table 10. *Disaggregating training type.* Dependent variable: Total and employer provided training. Multinomial logits with 5 and 6 characteristics, respectively.

	(1)		(2)	
	Total training		*Employer provided training*	
a. third level qualification	-2.857*	(1.743)	-5.665**	(2.823)
b. vocational training at school or college	-5.133***	(1.049)	-5.081***	(1.389)
c. workplace training and complementary instruction elsewhere	-1.270	(1.401)	-2.689*	(1.572)
d. workplace training without complementary instruction elsewhere	-.778	(.719)	-.784	(.763)

Note: In Column (1) the dependent variable is the disaggregated type of training (DTYPE), while in Column (2) the dependent variable is the disaggregated type of employer-sponsored training (DTYPEM). Only estimated coefficients of the median wage growth differential are reported. Cluster adjusted robust standard errors within parentheses. Observations are weighted by ECHP personal weights. Other controls are age plus country, education, occupation, sector, tenure, firm size, permanent job status and previous unemployment dummies. ***, **, *: significant at the 1%, 5%, 10% level of confidence, respectively.

Conclusions

When labor 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 paper, we have proposed a novel way of testing this proposition, which exploits the cross – country variation in training incidence and training wage premia within the European Union. We have found that the incidence of general training, which we have proxied 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 correlation between the latter variable and training incidence holds and is statistically significant also when general training is employer – provided. Conversely, we have found no evidence that firm – specific training (proxied by workplace training) is significantly correlated with our measure of the training wage premium.

While statistically significant, the economic impact of changes in the training wage premium on the probability of general training - relative to no 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 relative probability of general training by 0.06-0.09 percent.

Our findings are inconsistent with competitive theories of the labor market, which imply a positive correlation between the training wage premium and the incidence of general training. They suggest instead that economic environments with higher wage compression can help firms in organizing and paying for general training, as predicted by the recent theories of training in imperfect labor markets.

Appendix A: Data quality, sample selection, variable definition and descriptive statistics

A.1 Definition of co-variates and descriptive statistics

Wage growth differentials are defined in terms of gross hourly wages, computed from gross monthly earnings in the main job, by dividing them by 52/12 and by the number of usual weekly hours of work. Overtime pay and hours are included, but individuals that are either still in training in the final year, or working less than 30 hours per week (despite declaring to work full-time) or more than 70 hours per week are excluded.[34] 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 rate of 15.1%).[35] Furthermore, the thresholds of 30 or observations and 5 training events imply that the sample is reduced to 48 clusters, although, not surprisingly, retained clusters are larger and account for 90% of observations with non-missing wage growth information (with an average training rate of 14.9%).[36]

In the regressions, *age* is measured in years, while all other co-variates (with the exception of wage growth differentials) are categorical, and codified as sets of dummy variables (omitting one dummy per set for identification). We 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, corresponding to ISCO-88 code 6), 13 *sectors* (listed in Table A1), 4 *firm size* classes (1 to 49 employees, 50 to 99 employees, 100 to 499 employees, and 500 employees or

---

[34] Ideally, we would like to exclude also those individuals that were in training at the moment of the interview two years before. However, this information is not available in the 1994 wave.

[35] In the case of Austria wage growth is computed for 1997, since there are no data in 1994 (remember that wage growth differential are computed as logarithmic difference between the wage reported in the current wave and the wage reported two waves before).

[36] The training rate in the sample used to compute wage growth differentials is smaller than in the sample used for the final regressions because individuals still in training are excluded in the former but obviously not in the latter.

more), a dummy for *permanent contract*, a dummy for at least one spell of *unemployment* since 1989, and 5 *tenure* classes (0 to 1 year, 2 to 5 years, 6 to 9 years, 10 to 14 years, and more than 15 years). 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.[37] Means of co-variates are reported in Table A2.[38]

Table A1. List of sectors

ISIC Rev.3 Codes	Sector definition
C+E	Mining and quarrying + Electricity, gas and water supply
DA	Manufacture of food products, beverages and tobacco
DB+DC	Manufacture of textiles, clothing and leather products
DD+DE	Manufacture off wood and paper products; publishing and printing
DF-DI	Manufacture of coke, refined petroleum/chemicals/rubber & plastic products
DJ+DK	Manufacture of metal products, machinery and equipment n.e.c.
DL-DN	Other manufacturing
F	Construction
G	Wholesale and retail trade; repair of motor vehicles, motorcycles and personal/household goods
H	Hotels and restaurants
I	Transport, storage and communication
J	Financial intermediation
K	Real estate, renting and business activities

[37] 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.

[38] The standard deviation of age, the only continuous variable in the sample beside the wage growth differential, is 8.1 years.

Table A2: Means of co-variates (baseline sample: age 30 to 60 years, year: 1996).

Variable	Mean	Variable	Mean
Age (years)	42.2	% permanent contract	89.7
% tenure 1 year or less	12.6	% ISIC C+E	2.8
% tenure 2 to 5 years	16.4	% ISIC DA	4.8
% tenure 6 to 9 years	15.1	% ISIC DB+DC	2.7
% tenure 10 to 14 years	11.5	% ISIC DD+DE	4.7
% tenure 15 years or more	44.4	% ISIC DF-DI	7.9
% firm size less than 50 employees	45.1	% ISIC DJ+DK	14.2
% firm size 50-99 employees	10.5	% ISIC DL-DN	11.1
% firm size 100-499 employees	18.5	% ISIC F	15.2
% firm size 500 employees or more	25.9	% ISIC G	14.2
% tertiary education	19.1	% ISIC H	2.4
% upper secondary education	43.3	% ISIC I	6.6
% less than upper secondary education	37.6	% ISIC J	6.5
% Legislators, senior officials and managers	9.3	% ISIC K	6.9
% Professionals	7.1	% Austria	10.9
% Technicians and associate professionals	12.8	% Belgium	8.0
% Clerks	10.6	% France	18.8
% Service and shop and market sales workers	5.5	% Germany	16.2
% Craft and related trades workers	31.0	% Italy	17.6
% Plant and machine operators and assemblers	16.0	% Spain	18.4
% Elementary occupations	7.8	% United Kingdom	10.1
% previous unemployment	29.3		

A.2 Data quality and sample selection

Our ECHP release contains longitudinal data for 5 years (1994-1998), although not all countries are available in each wave. The country sample is chosen on the basis of data availability: Sweden has no wage data; Greece, Ireland, Luxembourg and Portugal, have either very small sample or very small training incidence, so that almost no cluster would be retained for these countries after applying the thresholds of 30 observations and 5 training events; and Denmark, the Netherlands and Finland has no information on training type (or that information is almost always missing). Thus, these eight countries are excluded from the sample, leaving us with 7 countries (Austria, Belgium, France, Germany, Italy, Spain, and the United Kingdom), although Germany and the UK are available only until 1996 included.

Unit non-responses and attrition rates in the ECHP are comparable to those of other longitudinal household surveys (Peracchi, 2002). Nevertheless, due to small entry rates, attrition results in a reduction of the sample size that is increasing with time. Focusing only on the countries that are present in two consecutive waves, the loss of observations due to attrition and insufficient replacement amounts to about 7% between the first (1994) and the

second (1995) wave, 4% between the second and the third (1996) wave, and 6% between the third and the fourth (1997) wave as well as between the fourth and the fifth (1998) wave.

Item non-response is a much more serious problem in the ECHP, except for the 1996 wave. In 1996 there are 7,196 male wage and salary employees, aged from 30 to 60 years and working full-time in the non-agricultural market-based sectors. Excluding observations with missing values either in the training variables or other covariates (with the exception of the wage growth differential), this sample drops to 6,447 individuals, corresponding to 90% of the original sample. If we also drop observations for which the wage growth differential is not defined (because of the thresholds of 30 observations and 5 trained individual per cluster), the sample drops further to 5,741 observations (80% of the original sample). In waves other than 1996 the problem is far more serious: male wage and salary employees, aged from 30 to 60 years and working full-time in the non-agricultural market-based sectors are 3,827 and 2,546 in 1997 and 1998, respectively. These figures drop to 2,413 (63% of the original sample) and 1,842 (72%), respectively, when observations with missing values are excluded.[39]

We choose to undertake a cross-sectional analysis of 1996 data only, because this year has much larger country coverage[40], a higher number of clusters with non-missing wage growth differentials[41] and a bigger sample size. For the same reason, we do not exploit further the longitudinal structure of the ECHP by pooling together more than one wave.

Eurostat's guidelines specify that National Data Collection Units (NDUs) are requested to use country-specific categories of education and training according to the classification used in the national Labour Force Surveys (LFS). In order to facilitate the coding of training questions by the NDUs, Eurostat provides a correspondence between LFS and ECHP categories. In the case of France and the United Kingdom, no national category corresponds to category d (specific vocational training in a working environment, without complementary instruction elsewhere). However, some training is reported to fall under this category in the British and French data. We choose to keep these countries in our sample, but

---

[39] These figures are based on only four and three countries, respectively, taking into account that there are no data for Germany and the United Kingdom in these years and that, due to missing values for certain co-variates, France and Italy must be excluded from the 1997 and 1998 and the 1998 samples, respectively.

[40] In 1996 we can cover seven countries in our data. Conversely, taking into account attrition and item non-response we can include only four countries (Austria, Belgium, Italy and Spain) in 1997 and only three (Austria, Belgium and Spain) in 1998.

[41] After applying our size thresholds, there are 48 clusters in 1996 (44 without Austria), 19 clusters in 1997 and only 14 clusters in 1998.

we also check whether our findings are driven by misclassification by replicating our regression analysis on a sub-sample of countries which exclude France and the UK. All results that are presented in the paper are robust to the simultaneous elimination of these two countries (detailed results are available from authors).

Appendix B: Relationship between the growth rate of average human capital and its stock in the augmented Solow model.

In this appendix we show that, conditional on the level of the steady state, the transitional dynamics of the augmented Solow model implies that the growth rate of average human capital is a decreasing function of the current average stock of human capital.

Let us start by defining the production function at time t:

$$Y(t) = K(t)^{\alpha} H(t)^{\beta} \left(A(t)L(t) \right)^{1-\alpha-\beta}$$

where Y, K and H are output, physical and human capital respectively and α and β are the partial elasticities of output to physical and human capital, with $\alpha + \beta < 1$ (*i.e.* decreasing returns to reproducible factors). The time paths of the right-hand side variables can be described by the following system of equations:

$$
\begin{aligned}
\dot{k}(t) &= s_k(t)A(t)^{1-\alpha-\beta} k(t)^{\alpha} h(t)^{\beta} - (n(t)+d_k(t))k(t) \\
\dot{h}(t) &= s_h(t)A(t)^{1-\alpha-\beta} k(t)^{\alpha} h(t)^{\beta} - (n(t)+d_h(t))h(t) \\
\dot{A}(t) &= g(t)A(t) \\
\dot{L}(t) &= n(t)L(t)
\end{aligned}
\qquad [A1]
$$

where $k = K/L$ stands for physical capital in intensive terms, $h = H/L$ for average human capital, s_k and s_h for the investment rate in physical and human capital, n is the growth rate of labour, g is the rate of technological change and d_k and d_h are depreciation rates for physical

and human capital, respectively. To simplify, let us assume that $d_k = d_h = d$, and that g and n are equal to zero.[42] Expressing the first two equations [A1] in growth rates yields:

$$\frac{d\ln k}{dt} = s_k A^{1-\alpha-\beta} k^{\alpha-1} h^{\beta} - d$$

$$\frac{d\ln h}{dt} = s_h A^{1-\alpha-\beta} k^{\alpha} h^{\beta-1} - d$$

Expressing saving rates as a function of steady state variables (taking into account that growth rates are zero at the steady state) and re-arranging we obtain:

$$\frac{d\ln k}{dt} = d e^{-(1-\alpha)\ln\frac{k}{k^*}} e^{\beta\ln\frac{h}{h^*}} - d$$

$$\frac{d\ln h}{dt} = d e^{\alpha\ln\frac{k}{k^*}} e^{-(1-\beta)\ln\frac{h}{h^*}} - d$$

[A2]

where k^* and h^* are the steady state values of k and h, respectively. Log-linearizing and solving the system [A2] yields the following solution for $\ln h$:

$$\ln h(t) = \psi \ln h^*(t) + (1-\psi) \ln h(t-1)$$

[A3]

where ψ is a function of α, β and d, with $0 < \psi < 1$ se $\alpha+\beta < 1$. Re-arranging equation [A3] yields the following expression relating $\Delta\ln h$ to $\ln h$, where Δ stands for the difference between time t and time t-1:[43]

$$\Delta\ln h(t) = \frac{\psi}{1-\psi} \ln h^*(t) - \frac{\psi}{1-\psi} \ln h(t)$$

Thus, conditional to the steady state level, the growth rate of average human capital is a decreasing function of its current stock level.

---

[42] The argument we develop below can be easily generalized.

[43] $\Delta\ln h$ is an approximation of the growth rate of average human capital for a discrete time difference.

36

References

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, *Economic Journal*, 109, F112-F142

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

Arulampalam, W., Booth, A. and Bryan, M., 2002, Work Related Training and the New National Minimum Wage in Britain, IZA Discussion Paper No. 595.

Barro, R. and Sala-i-Martin, X., 1995, *Economic Growth*, McGraw Hill, New York.

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., 2003, Comparative Analysis of Firm Demographics and Survival: Micro-level Evidence for the OECD Countries, OECD Economics Department Working Paper No. 348.

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, Istitute for Fiscal Studies, London.

Booth, A and Bryan, M., 2002, Who pays for general training? New evidence for British men and women, IZA Discussion Paper 486.

37

Brunello, G., 2001, Labor market institutions and the complementarity between education and training in Europe, forthcoming in Checchi, D. and Lucifora, C., (eds.), Education, training and labor market outcomes in Europe, Palgrave.

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

Goux, D. and Maurin, E., 2000, Returns to Firm-Provided Training: Evidence from French Worker-Firm Matched Data, *Labour Economics*, 7, 1-19.

Hashimoto, M., 1981, Firm-Specific Human Capital as a Shared Investment, *American Economic Review*, 71, 475-82.

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

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.

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, 1999, Employment Outlook.

OECD, 2003, Employment Outlook, forthcoming.

Peracchi, F., 2002, The European Community Household Panel: A Review, *Empirical Economics*, 27, 63-90.

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

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

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

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

IZA Discussion Papers

No.	Author(s)	Title	Area	Date
825	A. Filippin A. Ichino	Gender Wage Gap in Expectations and Realizations	5	07/03
826	K. T. Hansen J. J. Heckman K. J. Mullen	The Effect of Schooling and Ability on Achievement Test Scores	6	07/03
827	H. Buddelmeyer E. Skoufias	An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA	6	07/03
828	D. Checchi T. Jappelli	School Choice and Quality	3	07/03
829	J. J. Heckman X. Li	Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000	6	07/03
830	T. J. Hatton	Emigration from the UK, 1870-1913 and 1950-1998	1	07/03
831	J. H. Abbring G. J. van den Berg	Analyzing the Effect of Dynamically Assigned Treatments Using Duration Models, Binary Treatment Models, and Panel Data Models	6	07/03
832	P.-C. Michaud	Joint Labour Supply Dynamics of Older Couples	3	07/03
833	H. Gersbach	Structural Reforms and the Macroeconomy: The Role of General Equilibrium Effects	2	07/03
834	T. Boeri J. I. Conde-Ruiz V. Galasso	Protecting Against Labour Market Risk: Employment Protection or Unemployment Benefits?	3	07/03
835	G. Joseph O. Pierrard H. R. Sneessens	Job Turnover, Unemployment and Labor Market Institutions	3	07/03
836	M.-S. Yun	A Simple Solution to the Identification Problem in Detailed Wage Decompositions	3	07/03
837	W. Schnedler	What You Always Wanted to Know About Censoring But Never Dared to Ask - Parameter Estimation for Censored Random Vectors	6	07/03
838	S. Girma H. Görg E. Strobl	Government Grants, Plant Survival and Employment Growth: A Micro-Econometric Analysis	6	07/03
839	A. Bassanini G. Brunello	Is Training More Frequent When Wage Compression Is Higher? Evidence from the European Community Household Panel	6	08/03

An updated list of IZA Discussion Papers is available on the center's homepage www.iza.org.