The Role of Firm Size in Training Provision Decisions: evidence from Spain[•]

By Laia Castany

Universitat de Barcelona. AQR Research Group – IREA.
Avda Diagonal 690, 08034 Barcelona. Email: lcastany@ub.edu
Tel: +34 93 4021824. FAX: +34 93 402182

Abstract.: The level of training provided by small firms to their employees is below that provided by their larger counterparts. The provision of firm-related training is believed to be associated to certain characteristics of the firm. In this paper we argue that small firms provide fewer training opportunities as they are less likely to be associated with these characteristics than large firms. The suitability of estimating training decisions as a double-decision process is examined here: first, a firm has to decide whether to provide training or not and, second, having decided to do so, the amount of training to provide. The differences in training provision between small and large firms are decomposed in order to analyse the individual contribution of these characteristics to explaining the gap. The results show that small firms face greater obstacles in accessing training and that the main reasons for that are related to their technological activity and the geographical scope of the market in which they operate.

Keywords: Continuous training; Firm size; Innovative activity;

JEL codes: J21, L11, M53, L25

[•] I am grateful to Hessel Oosterbeek for his suggestions and discussion of the results reported here, to Enrique López-Bazo for his supervision and support and to Elisabet Motellón for her advice on Yun's methodology.

1. Introduction

Although the standards of education attained by the Spanish labour force have improved considerably over the last three decades, Spain still lags behind other advanced economies in this respect. However, employees' qualification does not depend solely on their schooling (formal education), but also on the quality of their life-long learning, including continuous and occupational training. The National Reform Programme for Spain (2005), developed within the framework of the Lisbon Agenda, stresses the importance of continuous training for already occupied workers as a means of acquiring additional knowledge and skills applicable to their present and future posts. Continuous training ensures that workers' skills can be adapted to the constantly evolving requirements of the workplace, which enhances their competitive position. As workers become more productive, their employers in turn improve their performance. Yet, in 2004, only 5.2% of the Spanish population received continuous training, while the EU-15 average stood at 10.7% and the EU-25 average at 9.9%.[1]

Workers in small and medium-sized enterprises (SMEs) are less likely to receive continuous training. Given that the Spanish economy is characterised by a smaller number of large firms and a smaller average firm size than other advanced economies, the difficulties faced by SMEs may constitute a limitation for the economy as a whole.

In this paper, the reasons as to why small firms may be facing greater difficulties in accessing continuous training are analysed. Based on the evidence that training is generally associated with certain types of firm and employees, the hypothesis propounded here is that large firms are more likely to be associated with these characteristics, or at least, to be associated with them more intensely, than their smaller counterparts. This, we argue, explains in part why small and large firms take different decisions regarding their training provision.

An empirical study conducted by Black *et al.* (1999) examines the relationship between different training measures and firm size for a sample of US firms and finds that large firms invest more heavily in training. They argue that such firms have scale economies in the provision of formal and informal training and provide more opportunities for undertaking co-worker training. Baldwin *et al*. (1995) contend that as large firms might enjoy a higher pay-off from their investment in training, they are encouraged to invest more. Holtmann and Idson (1991) maintain that large firms face

[1] The National Reform Programme seeks to raise the percentage in receipt of such training to 10% by 2008 and 12.5% by 2010.

lower investment risks as they can pool these risks. Barron *et al*. (1987) argue that there is a higher probability of shirking in large firms: as employees work cooperatively to produce a common output, it is more difficult to disentangle the participation of each individual worker. Thus, they claim, as large firms face higher monitoring costs, one way to reduce these, is providing training to their employees. According to Hashimoto (1979), large firms have access to cheaper capital for financing training. In the case of Spain, Rigby (2004) reports that small firms typically have access to training plans that "do not reflect the specific needs of employers and are promoted actively by social partners independently of employers".

Another strand of the literature is dedicated to examining the reasons why firms decide to train their workers and the amount of training they provide. Empirical studies of interest include Bartel (1989), Baldwin *et al*. (1995), Black and Lynch (1998) and Blundell *et al*. (1999). Evidence regarding the situation in Spain is reported in Alba-Ramírez (1994), Peraita (2005) and Albert *et al*. (2005a). These studies estimate the impact of certain characteristics of firms (determinants) that are believed to affect training decisions. In what follows, the role of these determinants and how they might vary with firm size are discussed.

To start with, training will tend to be provided to those who have previously shown an aptitude for formal education as they will benefit most from a firm's expenditure on training (Black and Lynch, 1998; Alba-Ramírez, 1994). Thus, firms that employ more qualified workers are more likely to provide training. On the role of firm size in this relationship, Evans and Leighton (1989) report evidence of some sorting on ability characteristics across firm sizes (with better educated workers being employed in larger firms) while Zábojník and Bernhardt (2001) propose a model where workers in larger firms and industries acquire more human capital.

Technological changes are introduced at high speed and require the continuous upgrading of the labour force. This use of advanced, specialized technology requires specific knowledge and skills that are not readily available in the labour market and so training is a way of acquiring such skills (Baldwin *et al*., 1995; Korpi and Mertens, 2004). The empirical evidence suggests that technological change leads to workers receiving more training (upskilling) (see Osterman, 1995). Additionally, firms that launch new products may need to train sales or technical staff, while those that implement process innovations may need to provide their production workers with technical training (Alba-Ramírez, 1994; Li *et al*., 2006). Similarly, the specific knowledge that the new process or product requires may not be easily found in the labour market. In line with Schumpeter (1942), various authors have argued that large firms enjoy an advantage over small companies in terms of innovation and the evidence from Spain seems to confirm this (Huergo and Jaumandreu, 2004). The same argument holds for a more intense use of advanced technologies in large firms.

Firms exposed to more competitive markets may invest more in training programmes as a strategy to enhance the competitiveness of their employees (see, for instance, Bartel, 1989). There are reasons to believe that both small and large firms are willing to provide training to increase their competitiveness: the former, because of their vulnerability to highly competitive markets; the latter, because they place themselves in such competitive markets, such as international markets.

Other studies argue that foreign-owned firms are more likely to provide training for their workers (Görg and Strobl, 2005; Hughes *et al*., 2004). Typically, such companies are multinational firms, characterised by managerial efficiency, the recruitment of more qualified workers and a more positive attitude regarding workers' skills than that of domestic firms.

Finally, firms with a high proportion of temporary workers are assumed to invest less in training. This state of affairs might be particularly important in the case of the Spanish labour market, which has a high degree of temporary employment (Alba-Ramírez, 1994; Albert *et al*., 2005b). Firms will not be interested in providing training for workers that may leave their employment in the short term as they will be unable to recover the returns from their investment. Furthermore, temporary workers lack incentives to acquire firm-specific human capital as they are unlikely to stay with the firm. Oi (1983) finds that large firms have less rotation because of internal labour markets.[2]

To analyse these questions, we use data drawn from the *Encuesta sobre Estrategias Empresariales* (ESEE). This survey gathers information from a sample of Spanish industrial firms with at least 10 employees, which has been widely used in empirical industrial organisation. This study aligns itself initially with that strand of literature that estimates the determinants of firm-related training. What is innovative, however, about this study is its consideration of the decision to provide training as a

[2] There are determinants of training for which we cannot control: the degree of unionization in the firm (Wagar, 1997), the personal characteristics of workers (Oosterbeek, 1996) and the workplace and personnel practices (Black and Lynch, 2004).

double-decision process: firms first decide whether they will provide training or not and, once an affirmative decision has been taken, they decide the amount they want to provide. To analyse the training differential between small and large firms, this paper decomposes this differential using the Oaxaca-Blinder methodology. This allows us to evaluate the individual contribution of each variable to this gap and to distinguish between differences due to firms' characteristics and differences due to the impact of these determinants on training.

In Section 2, the appropriateness of using a two-part model for estimating firms' decisions regarding the provision of training is discussed. Section 3 offers evidence that small firms spend less on training and that this is related to firms' characteristics. Section 4 provides the results of the estimation and Section 5 decomposes the training gap. Section 6 concludes.

2. Methodological questions and empirical model

It is common practice to estimate a probit model to analyse the factors that determine whether firms provide training to their employees or not. Likewise, in analysing the determinants of firms' expenditure on training, it is a fairly common approach to estimate a tobit model, which takes into account the fact that the dependent variable is censored at zero as, by nature, it can only take nonnegative values (Alba-Ramírez, 1994; Black and Lynch, 1998; Black *et al*., 1999). Estimating the specification by Ordinary Least Squares (OLS) instead, would provide inconsistent estimates, as it assumes that the dependent variable can take both positive and negative values. Moreover, as the logarithm of zero does not exist, a common solution is to add a small positive constant; but this constant is set arbitrarily.

The main limitation of the tobit model is that it does not consider decisions regarding the provision of training as a double-decision process: first, firms decide whether they will invest in training or not (participation decision), and once the decision has been taken, they decide the amount they wish to invest (quantity decision). This is particularly true when the two decisions are motivated by different determinants. For instance, when the decision as to whether to provide training involves incurring fixed costs, such as designing a training plan. Then, these fixed costs determine the decision as to whether to spend a sum of money or not, but they do not necessarily affect the decision regarding the quantity. Even in the case that the two decisions are dependent on the same factors, the dependent variable may include observations that take a zero value with high frequency and this mass of zeros might respond in a different way to covariates than observations with positive values. When this occurs, there are reasons to model the decision of training as two separate mechanisms, which can be seen as a generalisation of the tobit model. These models, henceforth referred to as two-part models, add flexibility in the sense that they allow zeros and non-zeros to be generated from different processes (Cameron and Trivedi, 2005).

Two approaches to these flexible models can be adopted: the sample selection model and the two-part model itself. The main difference between them is that the former takes the sample selection effect into account, which, when omitted, can cause biased estimations. The most popular sample selection model is the bivariate sample selection model studied by Heckman (1979). The so-called Heckit model comprises a participation equation that determines the sample selection:

$$dTR_i^* = X_{1i}'\beta_1 + \varepsilon_{1i} \tag{1}$$

where dTR_i^* is a censoring latent variable that reflects whether each *i*-firm is willing to provide training, X_{1i} is a vector of variables that determines this decision and ε_{1i} is the error term. The willingness of firms to provide training cannot be observed, but we can observe whether the firm spends money on it. Define dTR_i as the censoring observed variable, which is a binary indicator that takes value 1 if we observe that the firm dedicates some amount of money to training. Thus, $dTR_i = 1$ if $dTR_i^* > 0$ and $dTR_i = 0$ otherwise.

Define TR_i as the firms' expenditure on training and $lnTR_i$ as its logarithm, which is determined by a vector of variables X_{2i}. The quantity equation can be expressed as:

$$\ln TR_i = X_{2i}'\beta_2 + \varepsilon_{2i} \tag{2}$$

where ε_{2i} is the error term. Assuming that the error terms ε_{1i} and ε_{2i} follow a bivariate normal distribution with zero means, standard deviation σ_1 and σ_2, covariance σ_{12} and correlation ρ:

$$E(\ln TR_i \mid dTR_i = 1) = X'_{2i}\,\beta_2 + \sigma_{12}\lambda_i\left(X'_{1i}\,\beta_1\right) \tag{3}$$

where $\lambda_i(X_{1i}'\beta_1) = \phi(X_{1i}'\beta_1)/\Phi(X_{1i}'\beta_1)$ is defined as the inverse Mills' ratio, ϕ is the standard normal density function and Φ is the standard normal cumulative distribution function.

The coefficients β_1 are obtained by a first-step probit regression of *dTR* on X_1: $P(dTR=1)=\Phi(X'_{1i}\beta_1)$. The Heckit model augments the OLS regression on the quantity of training by the inverse Mills' ratio and then uses the positive values of *TR* to estimate the model by OLS. The estimate of β_2 is consistent, as it takes the sample selection bias into account.[3] By introducing the inverse Mills' ratio, this model corrects for possible sample selection effects. Sample selection appears when the error terms of the two equations are not independent, and thus the covariance of the error terms, σ_{12}, is different from zero.

When σ_{12} equals zero, the Heckit model simplifies to the two-part model, which simply uses the positive values of *TR* to estimate the model by OLS, obtaining consistent estimates of the β_2 parameters. The two-part model was first proposed by Cragg (1971) and was specifically designed for data on expenditure that contains a large number of zeros and a right-skewed distribution. The two-part model also starts from a participation and a quantity equation and equation (3) simplifies to:

$$E(\ln TR_i \mid dTR_i = 1) = X'_{2i}\,\beta_2 \tag{4}$$

In this paper, the following quantity equations, corresponding to the Heckit and the two-part models respectively, are estimated. For *dTR=1*:

$$\ln TR_i = X'_{2i}\,\beta_2 + \sigma_{12}\lambda_i\left(X'_{1i}\,\beta_1\right) + v_i \tag{5}$$

$$\ln TR_i = X'_{2i}\,\beta_2 + \varepsilon_{2i} \tag{6}$$

3. The dataset and descriptive analysis

For the empirical analysis, I use a sample of Spanish industrial firms drawn from the *Encuesta sobre Estrategias Empresariales* (ESEE), carried out by the *Fundación Empresa Pública* (FUNEP). This survey is an unbalanced panel that covers the period 1990-2002 and gathers information on the strategic decisions and behaviour of firms. The reference population of the ESEE is firms with 10 or more employees dedicated to one of the activities corresponding to divisions 15 to 37 from the CNAE-93, excluding division 23 (activities related to refinement of oil and fuel treatment). In the base time period, all firms with more than 200 employees were required to participate (of which 70% actually did). The firms with 10 to 200 employees were sampled randomly by

[3] The bivariate sample selection model can also be estimated by ML although this would impose stronger assumptions on the distribution of the error terms.

industry and four size strata, retaining about 5%, so that representativity for every industry and firm size was guaranteed.[4] In the present paper, we use information corresponding to years 2001 and 2002,[5] in which data is available for 1,515 and 1,505 firms respectively, of which 30% are large.

The variables used in this analysis are defined as follows:

- Training is measured as a discrete variable (according to whether the firm provides continuous training or not) and as a continuous variable (the log of the real expenditure on continuous training per worker). Continuous training is measured as the external expenses on training per worker, including five different types of training: computation and information technologies, foreign languages, sales and marketing, engineering and technical training and other issues.
- Firm size is defined as the total number of employees, taking into consideration full time and part time employees as well as temporary employees.
- The percentage of white collars in the firm is the proportion of engineers, graduates, middle level engineers, experts and qualified assistants on the total number of employees.[6]
- The intensity of use of advanced technologies is measured by a set of three dummy variables labelled as low, medium and high, when firms use 0-1, 2-3 or 4-5 advanced technologies respectively. The ESEE considers these technologies as being: Computer Numerically Controlled (CNC) machines and tools, Robots, Computer-Aided Design (CAD), a combination of the previous systems by central computer (CAM, flexible manufacturing systems, etc) and Local Area Network (LAN) for factory use.
- Innovation is defined as a dummy variable that takes value 1 if the firm has introduced a product or a process innovation.
- The geographical scope of the market where the firm operates is defined by a dummy that takes values 1 when the firm operates in an international market and zero otherwise.
- The participation of foreign capital is defined as the percentage of foreign-owned capital in the firm.

[4] For further details on the dataset, see Fariñas and Jaumandreu (1999).
[5] The information on the firms' provision of continuous training in the ESEE is only available for 2001 and 2002.
[6] This variable enters the equations with one lag. Data on white collar workers are not available for 2000 and 2001 as they are not assumed to change substantially from one year to the next. We therefore interpolate the percentage of white collars, assuming that they change linearly.

- The percentage of temporary workers over the total employees is measured at the end of the year. When the firm reports that the number of temporary employees has changed considerably, it is calculated as a mean of every quarter.
- Finally, the control variables are: the use of the productive capacity (percentage), a dummy reflecting whether the firm is part of a group, a set of 20 sector dummies, a set of 17 regional dummies and year dummies.

Table 1 presents a descriptive analysis of training, both for the discrete variable (*dTR*) and for the expenditure per worker (*TR*) according to the above-mentioned characteristics. First, the data show that around 40% of the firms in the sample provide training. In terms of the firms' characteristics, we see that firms with a percentage of white collar above the median (labelled "high % of white collars"), innovative firms, firms that make an intense use of advanced technologies,[7] firms that operate in international markets, those that have a higher participation of foreign capital and those with a percentage of temporary workers below the median provide more training. Moreover, the differences between firms with these characteristics and without them are significant at 1%. This evidence supports the belief that training is associated with certain characteristics of the firm.

[Insert Table 1 about here]

As this study focuses on differences by firm-size, we divide the sample into two subsamples of small and large firms. Around 25% of small and 75% of large firms provide some training, while the average real expenditure per worker is 42 euros and 141 euros respectively. Thus, large firms provide more training and the differences are significant at 1%.

As Table 2 shows, among the firms with a percentage of white collars above the median, large firms provide significantly more training than their smaller counterparts. Similarly, among the firms with a percentage of white collars below the median, large firms also provide significantly more training. Analogue results are obtained for all the other firm characteristics considered here (significance at 1%).

In addition, the statistics of the tests of equality of proportions and means that compare the provision of training in small and large firms are smaller for the group of firms with a high level of qualified workers than for the group with a low level. This indicates that differences between small and large firms reduce for firms with more

[7] Firms that make a high use of these technologies provide more training than firms with medium use, and firms that make medium use of advanced technologies provide more training than firms with low use.

human capital. Thus, having a high percentage of white collars seems to slightly mitigate the differences in training provision decisions between small and large firms. This result is obtained for all the other characteristics (except for temporary workers). Therefore, the results suggest that the presence of certain characteristics makes small firms behave as large ones with respect to the provision of training; nevertheless, the differences between the two groups remain notable.

[Insert Table 2 about here]

The objective here is to determine whether small and large firms present different patterns in their training decisions in relation to the characteristics of the firm. The hypothesis is that the difficulties faced by small firms in accessing training are related to the fact that they are not associated with the above-mentioned characteristics (or not associated with the same intensity as large firms). Table 3 confirms this reasoning: large firms employ more high-skilled workers, innovate more and use advanced technology with an intermediate and high intensity more than small firms do; they also operate more in international markets and they have a greater participation of foreign capital. As for small firms, they use advanced technology with low intensity more than large firms do and they have more temporary workers than large firms. The differences in these characteristics between the two groups are significant at 1%. These results suggest that large firms may provide more training because they are more associated with these characteristics. In the next section, we perform a regression analysis to study whether such characteristics are driving the training decisions and whether they have different impact on small and large firms.

[Insert Table 3 about here]

4. Estimation

4.1. The Heckit and the two-part model

As commented above, around 60% of the observations of the dependent variable *TR* take value zero. This percentage indicates the existence of a high degree of censoring and, thus, the need to consider the possibility that the zeros and positive observations are generated by different processes.[8] In this Section, we discuss whether it is more

[8] The distribution of expenditure on training per worker is clearly right skewed. The median is 90€ per worker in 2001 and 109€ per worker in 2002, while the average is 171 and 186 respectively. The skewness coefficient is 7 in 2001 and 5.3 in 2002.

appropriate to model firms' training decisions as a two-part model with sample selection or not.

The estimation of expressions (5) and (6) is shown in Table 4. The first and second columns show, respectively, the marginal effects and coefficients of the participation equation (which is the same in the Heckit and the two-part model). The third and fourth columns show the coefficients of the quantity equation in the Heckit model in the two-part model.[9] In the participation equation, all the variables of interest are significant (except for the percentage of temporary workers) and have the expected sign. In the quantity equation, the percentage of white collars, the innovative activity, the participation of foreign capital and the percentage of temporary workers are clearly significant. In the Heckit model, the intensive use of advanced technologies and operating in an international market also increase firms' expenditure on training significantly.

[Insert Table 4 about here]

The choice between the two models is a controversial question and has led to intense debate in recent years (Dow and Norton, 2003). First, the type of dependent variable to be modelled must be given careful consideration. To put it simply: when analysing training expenditure with a large proportion of zeros, do we observe *potential* training-providers that for some reason decided not to provide training? Or do we observe firms that do not wish to provide training (*observed* outcome)? Dow and Norton (2003) argue that when the observed zeros do not represent zero values for the potential outcome then a sample selection bias might appear.

Lynch (1993) argues that, in small firms, fixed costs of training are distributed across a smaller number of employees: for instance, the production losses associated with a worker being away from his or her workplace or the design of a firm's training plan can be more costly in a small than in a large firm. We argue that, in the presence of fixed costs, some firms cannot afford to provide training and we observe a zero in the variable measuring the expenditure on training. If these fixed costs were smaller, these firms would decide to provide training and we would observe a positive value. Seen in this light, fixed costs might be hiding a latent potential training provision. Here, we are

[9] Notice that the two sets of coefficients of the quantity equation cannot be directly compared: while in the two-part model, the coefficients are equal to the conditional marginal effects, in the Heckit, they are only part of the conditional marginal effect. For further details, see Cameron and Trivedi (2005).

specifically interested in this potential outcome and so the Heckit model appears to be more appropriate.

Secondly, the Heckit model may have problems of identification when the same regressors are included in the two equations, while in the case of the two-part model this is not a limitation. Although the Heckit model with normal errors is theoretically identified without exclusion restrictions on any regressor, when the same regressors are included in the two equations, this model is close to unidentified. Cameron and Trivedi (2005) explain that sometimes it can be very difficult to make defensible exclusion restrictions. In our case, it seems difficult to find at least one regressor that determines the decision as to whether or not to provide training, but which does not determine the quantity of training provided.

A t-test on the coefficient of the inverse Mills' ratio can be used to test the null hypothesis that the two-part model is correct against the alternative hypothesis that the Heckit is correct.[10] However, under collinearity between the covariates and the inverse Mills' ratio, the power of the test is limited and this test cannot be used as a criterion to select between the two models; with low collinearity, the test is reliable. According to Leung and Yu (1996), imposing no exclusion restrictions is a main source of multicollinearity. These authors recommend using the condition number to check for multicollinearity between the inverse Mills' ratio and the covariates in the quantity equation. Based on Monte Carlo experiments, Belsley *et al*. (1980) suggest that a condition number beyond 30 is indicative of collinearity problems. For the total sample, the condition number for the covariates is 26.9, and after including the inverse Mills' ratio it takes a value of 36.9. As suggested in Cameron and Trivedi (2005), although the condition number including the inverse Mills' ratio takes a value above 30, the increase when including this regressor is very small, for which multicollinearity problems can not be considered as severe. Then, the test on the inverse Mills' ratio can be considered a useful tool for selecting between the two models. Table 4 shows that the coefficient of the inverse Mills' ratio takes a value of 0.57 and is not statistically significant; thus, the null that the two-part model is correct cannot be rejected.

Finally, using statistical criteria to select between the two models, Dow and Norton (2003) recommend the test proposed by Toro-Vizcarrondo and Wallace (1968), which

---

[10] Dow and Norton (2003) stress that if the coefficient of the inverse Mills' ratio is zero, the Heckit reduces exactly to the two-part model, but the two-part model does not require the coefficient to be equal to zero. The two models simply make different implicit distributional assumptions and they are only partially nested.

they name an empirical mean squared error (EMSE) test. The original test statistic was derived for OLS models, but the intuition can be extended to the Heckit and two-part models. This test involves calculating the EMSE of both estimators, under the assumption that one model is consistent and correct. Then, the estimator with the lower EMSE is chosen. For most of the variables in the empirical specification (with the sole exception of the variable on temporary workers), the EMSE for the two-part model is smaller than the EMSE for the Heckit model, indicating that the former seems more appropriate. As for the control variables, the same result is obtained, with the exception of some regional dummies.[11] Under the two assumptions, the results are similar, indicating the robustness of the results.

Therefore, although from a theoretical point of view it can be argued that sample selection might exist, a significance test on the inverse Mills' ratio and the EMSE test suggest that in practice it seems more appropriate to estimate a two-part model to model firms' training provision. In the case of the subsamples of small and large firms similar results are obtained.

4.2. The two-part model with random effects

Empirical studies that use firm-level datasets reveal a high degree of heterogeneity among firms with similar observed characteristics. This particularity of the data requires estimating a model that takes firm-specific effects into account. If there are significant unobserved time-invariant, firm-specific effects that are correlated with the explanatory variables, the simple pooled regression may produce biased and inconsistent estimates. In the case of micro-databases, where firms in the sample are selected randomly from a larger population, it is quite common to estimate a random effects model, rather than a fixed effects model.[12]

The participation equation is estimated by means of a random effects probit model, which assumes a normal distribution for the random effects. The model is estimated by maximum likelihood (Guilkey and Murphy, 1993).[13] As for the quantity

[11] Due to limitations of space, the results for the EMSE test have not been included here. However, they are available from the author upon request.

[12] See for instance Groot and Maassen van den Brink (2003), Barrios *et al*. (2003), Máñez *et al*. (2004) or Licandro *et al*. (2004).

[13] The integral in the likelihood function is approximated with the non-adaptive Gauss-Hermite quadrature. The quadrature formula requires that the integrated formula is well approximated by a polynomial. As the panel size increases, the quadrature approximation becomes less accurate. If the results of the estimation change when the number of quadrature points changes, the results should be dismissed. We verified the magnitude of these changes and found that, for most variables, the relative

equation, a standard regression model including random effects is estimated by generalized least squares.

Table 5 offers the results of the two-part model including firm-specific effects, for the total sample and for the subsamples of small and large firms. In the case of the total sample (first set of columns in Table 5), the results are similar to those in Table 4: the same variables are significant and they show the same sign. Although the results are similar to the model without the inclusion of random effects, the tests reject the null hypothesis that firm-specific effects are zero. In the case of the participation equation, when the panel-level variance component is unimportant, the panel estimator is not significantly different from the pooled estimator. The test rejects the null that the panel-level variance component is equal to zero at 1%. As for the quantity equation, the Breusch and Pagan Lagrange-multiplier test rejects the null hypothesis at 1%. Similar conclusions from the tests are obtained for the subsamples of small and large firms. Therefore, the two-part model with random effects was chosen to carry out the rest of the analysis.

[Insert Table 5 about here]

In general, the results obtained here confirm the findings of previous empirical studies. See for instance, Bartel (1989), Alba-Ramírez (1994), Baldwin *et al*. (1995), Black and Lynch (1998) and Hughes *et al*. (2004). More specifically, in the case of the total sample, the effect of firm size is positive and significant in the participation equation indicating the presence of effects associated with large firms even after controlling for the set of possible training determinants. In particular, increasing the firm size by one point increases the probability of firms providing training by 0.2. The fact that firm size is significantly positive in the participation equation, even after controlling for other variables and firm-specific effects, suggests the existence of scale economies in the provision of training as well as other effects associated with firm size.

Apart from this direct effect of firm size, the other covariates may have different effects in the subsamples of small and large firms, as suggested by the descriptive analysis in Section 3. For example, does an increase in the participation of skilled workers lead to a higher probability of training (or higher expenditure) in both small and large firms? Is this effect larger in magnitude in either group? To further analyse this question, the same equations are estimated for the subsamples of small and large

---

difference between the coefficients using different quadrature points was smaller than 0.01%. So, the results of the probit random effects model estimated here are reliable.

firms. Given that small firms are acknowledged to have more difficulties in accessing training, we are interested in analysing the impact of these variables on training decisions and whether they play different roles in firms of different size classes.

The second and third sets of Columns in Table 5 show the results for the estimation of the empirical specification for the subsamples of small and large firms respectively. Results suggest the existence of certain differences between small and large firms in their training provision decisions. Specifically, firm size has a negative effect on training expenditure per worker in small firms, while this is not the case with large firms. This could be explained by the high fixed costs of training, especially for the smallest firms.

In the case of the degree of qualification of the labour force, this factor does not determine whether large firms will decide to provide training, although it does have an impact on the amount of training provided. In small firms, however, the degree of qualification of the labour force is a determinant of both decisions. This result can be explained by the fact that large firms employ a wide range of employees, and so, *ceteris paribus*, there is a higher probability that they will provide training to at least one employee.

The variables related to technology appear to be important determinants of a firm's decision to provide training, both for small and large firms. However, in the case of the latter, the effects seem to be slightly smaller in magnitude than in the case of small firms. Moreover, in shifting from being a non-innovative to an innovative large firms increase their expenditure on training per worker by almost 22%, whereas in the case of small firms, this variable does not have a significant effect. These results suggest a relationship between size, technological activities and the training per worker. As we discuss below, technological activities appear to explain in part why large firms provide more training per worker.

In the case of small firms, competing in an international market and having foreign capital participation affect the two training decisions. This can be explained by the fact that small firms with these characteristics may decide to provide training as a way of guaranteeing success in their competitive environment. However, the impact of the geographical scope of the market on the decision as to whether to provide training is much larger in large firms than in their smaller counterparts.

Finally, the coefficient of the variable for temporary employment is only significant (presenting a negative sign) in the decision regarding the quantity of training in large firms. As before, given that large firms employ a wide range of workers, it does not affect their probability of providing training but rather the quantity of it.

With regards to the control variables on the group and the use of productive capacity, small and large firms do not differ in their respective behaviour. However, the sets of dummy variables on region and sector show differences between the two groups.

Overall, the technological activities and the geographical scope of the market appear to be important determinants of firms' training decisions. In addition, there are certain differences between small and large firms that may explain why small firms provide less training per employee than their larger counterparts. In the next Section, the training provision gap between small and large firms is decomposed to investigate further the contribution of these variables.

5. Decomposition of the training gap between small and large firms

5.1. The Oaxaca-Blinder decomposition in the two-part model

The Oaxaca-Blinder methodology is applied here to decompose the training provision differential between large and small firms.[14] It allows us to decompose the differences in the participation decision and in the amount of training into two components: differences in the *levels* of the determinants of training (firm characteristics) and differences in the *impact* of these determinants. The former reflects the fact that small and large firms have different characteristics, which are associated with different training levels. The latter reflects the differences by firm size in the impact of such characteristics on the training provision. For example, supposing that small and large firms had the same percentage of qualified workers, would they show a similar propensity to invest in training?

Starting from two auxiliary regressions for the sub-samples of small and large firms:

$$\begin{aligned} \hat{T}_S &= F(X'_S\,\hat{\beta}_S) \\ \hat{T}_L &= F(X'_L\,\hat{\beta}_L) \end{aligned} \tag{7}$$

[14] The Oaxaca-Blinder decomposition methodology has been widely used to study wage gaps associated with differences in workers' characteristics and discrimination by gender or race (Oaxaca, 1973 and Blinder, 1973). To the best of our knowledge, Smith *et al.* (2004) and Castany *et al*. (2007) are the only papers that have applied this method to analyse differences between firms.

where T denotes training, both as a discrete (*dTR*) or continuous variable (*lnTR*), X is the matrix of the regressors, $\hat{\beta}$ is the conforming vector of estimated coefficients and subscripts L and S refer to large firms and small firms respectively.[15]

Notice that $F(\cdot)$ can be both a linear function —quantity equation— or a non-linear function —participation equation. The traditional detailed Oaxaca-Blinder decomposition can be applied in linear models, but it is not suitable for non-linear specifications. Instead, for the latter we apply a recent proposal (Yun, 2004) to compute detailed decompositions for non-linear models that are linear in their arguments, such as the participation equation.[16]

According to the standard Oaxaca-Blinder decomposition, the differences in the quantity of training between small and large firms can be decomposed as:

$$\overline{\ln \hat{T}R_L} - \overline{\ln \hat{T}R_S} = \left(\overline{X}_L' - \overline{X}_S'\right)\hat{\beta}_L \quad + \quad \overline{X}_S'\left(\hat{\beta}_L - \hat{\beta}_S\right) \tag{8}$$

where the first term on the right-hand side is the part of the training gap due to differences in characteristics between the representative small and large firms and the second term on the right-hand side is the contribution of differences in the impact between both types of firm.

The standard version of the Oaxaca-Blinder decomposition builds on the assumption that one of the two equations is the "natural" model (for instance, in the case of the wage gap decomposition by gender, it may appear quite natural to assume that women are the "discriminated" group and, thus, to analyse what their wages would have been if they had had the returns of men). Nevertheless, in the present case there is no compelling reason to calculate the differences in firms' endowments assuming that all the firms had the coefficients of either large or small firms. It is difficult at times to establish which is the natural model and the results may often differ considerably. One strand of the literature suggests a variation on the standard decomposition that avoids having to make this assumption. According to this approach, there exists a "non-discriminatory structure of coefficients" $\hat{\beta}^*$ in relation to which one group is

[15] The Gardeazábal and Ugidos (2004) transformation has been applied in the estimation of in order to distinguish the effects due to the different sets of dummies.

[16] As far as we know Yun's detailed decomposition was applied for the first time by Motellón and López-Bazo (2005) and Hernanz and Toharia (2006).

"discriminated" while the other is "favoured" (Oaxaca and Ransom, 1994).[17] Then, the training differential can be expressed as:

$$\overline{\ln \hat{T}R_L} - \overline{\ln \hat{T}R_S} = \left(\overline{X}_L' - \overline{X}_S'\right)\hat{\beta}^* + \overline{X}_L'\left(\hat{\beta}_L - \hat{\beta}^*\right) + \overline{X}_S'\left(\hat{\beta}^* - \hat{\beta}_S\right) \quad (9)$$

where $\hat{\beta}^* = \Omega\hat{\beta}_L + (I - \Omega)\hat{\beta}_S$ and $\Omega = (X'X)^{-1}(X'_L X'_L)$. The first term on the right-hand side of (9) reflects training differences due to differences in firms' characteristics. The second and third terms are estimates of the large firms' advantage and small firms' disadvantage in relation to the non-discriminatory coefficients structure. The two terms together are considered as differences in the expenditure on training by firm size associated with differences in coefficients without imposing a discriminated group.

As for the decomposition of the participation equation, the Yun's methodology consists of finding the contribution of every *n*-variable to the total difference. Using the variation suggested by Oaxaca and Ransom (1994) for the Yun-Oaxaca-Blinder decomposition of the participation equation, we have:

$$\begin{aligned}\overline{d\hat{T}R_L} - \overline{d\hat{T}R_S} = {} & \sum_{n=1}^{N} W_{\Delta X}^{n}\left[\overline{\Phi(X_L\hat{\beta}^*)} - \overline{\Phi(X_S\hat{\beta}^*)}\right] + \sum_{n=1}^{N} W_{\Delta\beta fav}^{n}\left[\overline{\Phi(X_L\hat{\beta}_L)} - \overline{\Phi(X_L\hat{\beta}^*)}\right] + \\ & \sum_{n=1}^{N} W_{\Delta\beta disc}^{n}\left[\overline{\Phi(X_S\hat{\beta}^*)} - \overline{\Phi(X_S\hat{\beta}_S)}\right]\end{aligned} \quad (10)$$

where, Φ is a standard normal cumulative distribution function, $W_{\Delta X}^{n}$ and $W_{\Delta\beta}^{n}$ are the weights for each *n*-variable and subscripts *Δβfav* and *Δβdisc* indicate that the weights correspond to the effect of large firms' advantage and small firms' disadvantage in relation to the non-discriminatory coefficients structure. The key question is finding proper weights for the variables. Yun (2004) suggests evaluating the value of the function using mean characteristics and then using a first order Taylor expansion to linearize Φ around $\overline{X}_L\hat{\beta}_L$, $\overline{X}_S\hat{\beta}_S$ and $\overline{X}_S\hat{\beta}^*$. In this way, the weights can be expressed as:

$$W_{\Delta X}^{n} = \frac{(\overline{X}_L^n - \overline{X}_S^n)\hat{\beta}^{*n}}{(\overline{X}_L - \overline{X}_S)\hat{\beta}^*}\ ;\ W_{\Delta\beta fav}^{n} = \frac{(\hat{\beta}_L^n - \hat{\beta}^{*n})\overline{X}_L^n}{(\hat{\beta}_L - \hat{\beta}^*)\overline{X}_L}\ ;\ W_{\Delta\beta disc}^{n} = \frac{(\hat{\beta}^{*n} - \hat{\beta}_S^n)\overline{X}_S^n}{(\hat{\beta}^* - \hat{\beta}_S)\overline{X}_S} \quad (11)$$

The first term on the right-hand side of equation (10) reflects training differences due to differences in characteristics. This term is an estimate of the differential in the probability of providing training between small and large firms in the absence of differences in the impact of these characteristics. The second and third terms are

[17] It can be easily proved that a consistent estimate of β* can be obtained by OLS in the whole sample of firms.

estimates of the differential in probability of providing training due to differences in the *impact* of firms' characteristics. Together, they collect the effect of large firms' advantage and small firms' disadvantage in relation to the non-discriminatory coefficients structure.

5.2. Results of the decomposition of the training gaps

In this Section, we assess the individual contribution of firms' characteristics in explaining the training gap between small and large firms in two ways: differences in the *level* of the determinants of the training provision and differences in their *impact* on the training provision decisions. To perform this analysis, the detailed decomposition described in expressions (9) and (10) is applied.

Table 6 shows the results of the Oaxaca-Blinder decomposition based on the two-part model with and without firm-specific effects.[18,19] The magnitudes of the effects of each variable are similar under the two models. The differential in the probability of providing training between small and large firms is 2.7. The decomposition for all the variables together shows that most of the gap is due to differences in characteristics, while differences in the impact of characteristics explain only 5% of the gap (and even less in the model including firm-specific effects). However, we are especially interested in the individual decomposition for analysing the contribution of each variable.

[Insert Table 6 about here]

The fact that large firms employ more white-collar workers explains a very small part of the differential of the probability of providing training both as differences in characteristics and as differences in the impact of characteristics.

In the case of the variables related to technological activities, the differences in the intensity of their use explain around 20% of the gap, while the global impact of this variable has a very small effect. The differences in innovative activity between small and large firms explain about 10% of the differential in the probability of providing training, while the global impact of this variable is also very small in magnitude.

The differences in the variable related to the geographical scope of the firms' market explain about 8% of the gap in the probability of training, while differences in

[18] In the RE model the transformed residuals have zero mean, but not the residuals from the original specification. This prevents obtaining an exact decomposition of the training gap based on the RE estimates of the coefficients.

[19] Table 6 shows the most relevant results of the decomposition. For more complete results, see Table A1 at the Appendix.

the global impact of this variable are quite small. The differences in the participation of foreign capital and the percentage of temporary workers show a small contribution to explain the differences in the probability of providing training.

The differential in the logarithm of the expenditure on training per worker between small and large firms is 0.4. The decomposition for all the variables together shows that differences in firms' characteristics explain around 65% of the differential, while differences in the impact of characteristics explain 35%.

Again, the percentage of white-collar workers has an almost negligible contribution in explaining the gap in the quantity of training.

The use of advanced technologies explains more than 15% of the differential in the quantity of training (25% in the case of the RE model). Of this figure, around 5% of the differential is due to differences in the impact of using advanced technologies, while the rest is due to differences in characteristics (both in favour of large firms). The innovative activity also explains about 15% of the gap —more than 10% is due to differences in characteristics and the remaining portion is due to differences in the impact. Differences in the geographical scope of firms' markets explain more than 16% of the training gap and both differences in characteristics and differences in the impact of these characteristics have a similar contribution, both in favour of large firms.

The participation of foreign capital explains a fairly sizeable part of the differential: around 14% is due to the fact that large firms enjoy a greater participation of foreign capital. However, the impact of this variable is also quite large and favours small firms, with values around 24%. In other words, under equal impact of characteristics (i.e. coefficients), the gap in the probability of providing training would be larger favouring large firms.

Finally, the percentage of temporary workers makes an important contribution in explaining the differential in the quantity of training and this is mainly due to differences in the impact of characteristics in favour of small firms, taking values of almost 40%. Thus, if small and large firms recorded the same impact from the variable of temporary workers, *ceteris paribus*, the gap in the probability of providing training between small and large firms would be even wider.

All in all, the variables that play the most relevant role in explaining the gap between small and large firms in terms of the probability of providing training are: the use of advanced technology, the innovative activity and the international scope of the market in which firms operate. Together they explain about 40% of this gap and their effect is primarily due to differences in characteristics. As for the quantity equation, the variables that make the main contribution in explaining the gap are the same as in the case above, explaining more than half of the gap. However, here this is due to differences in characteristics as well as differences in the impact of these characteristics. Additionally, the participation of foreign capital in the firms and the percentage of temporary workers seem to explain a large part of the effect, which is due in particular to differences in the impact of these characteristics on the quantity of training in favour of small firms.

6. Conclusions

This paper has sought to understand why small firms provide their employees with less training than their larger counterparts. The initial hypothesis holds that large firms provide more training opportunities because they are endowed with certain characteristics that *permit* them to expend greater efforts in training their workers. These include employing more white-collar workers and engaging fewer temporary workers. Furthermore, the provision made by large firms is greater because they are endowed with certain characteristics that, in their turn, *require* more training. These include their adoption of more advanced technologies or innovative activity, operating in more competitive markets (e.g. internationally) and being partially owned by foreign capital.

The empirical evidence reviewed here seems to support the hypothesis that training provision is indeed closely associated with these characteristics. Using the ESEE, we have presented evidence that large Spanish industrial firms invest more training (per worker) and that they are more closely associated with these characteristics than their smaller counterparts.

The paper has discussed the suitability of adopting two-part models for analysing training decisions, both from a theoretical and applied perspective. Although the Heckit model seems to be more appropriate from a theoretical point of view, no evidence of strong sample selection was apparent in the case of Spain's industrial firms, suggesting that the two-part model might be more appropriate for modelling their training decisions. Based on previous evidence that small and large firms follow different patterns in their training decisions, we estimated the two subsamples separately.

The results of these estimations suggest that technological activities and the geographical scope of the market are important determinants of firms' training decisions for both small and large firms. Indeed, the effects of technological variables on the participation decision are larger in magnitude in the case of small firms. Likewise, a large firm that switches from being non-innovative to innovative will increase its expenditure on training considerably. In the case of the geographical scope of the market, this factor determines both decisions for small firms, while for large firms it has a greater effect on the probability of providing training.

The decomposition of the training gap between small and large firms allows us to assess the relative contribution to the gap made by these firms' characteristics. In the case of the decision as to whether to provide training, the most important contributions are related to firms' technological activity and the geographical scope of the market in which they operate; these effects are mainly due to differences in characteristics in favour of large firms. In the case of the decision regarding the quantity of training per worker, the variables related to technological activity and market scope explain a sizeable part of the gap. Here the effects are due both to differences in characteristics and differences in the impact of these characteristics in favour of large firms. In addition, the participation of foreign capital and temporary workers explain a large part of the gap, basically as differences in the impact of characteristics in favour of small firms.

Overall, this study confirms the finding that small firms face greater restrictions in gaining access to training. The results suggest that the differences in training provision between small and large firms are related to differences in the firms' *requirements* to update the skills of their employees to ensure that they acquire the specific knowledge to use new technologies and to make the firms more competitive in international markets. And these differences in necessity favour large firms. In other words, the differences between small and large firms do not, in general, seem to be related to those characteristics that allow firms to provide more training (i.e. the fact of employing more qualified workers or non-temporary workers). This can, perhaps, be seen as a factor that impedes small firms from becoming more competitive because of a more restricted access to training, a tool that enables employees to upgrade their skills and so become more competitive.

References

Alba-Ramirez, Alfonso, (1994). "Formal Training, Temporary Contracts, Productivity and Wages in Spain," *Oxford Bulletin of Economics and Statistics*, Blackwell Publishing, vol. 56(2), pages 151-70, May.

Albert, C., García-Serrano, C., and Herranz V., (2005a), "Los determinantes de la formación es España y sus rendimientos", *WP Universidad de Alcalá de Henares*.

Albert, C., García-Serrano, C., and Herranz V., (2005b), "Firm-provided training and temporary contracts," *Spanish Economic Review*, Springer, vol. 7(1), p. 67-88, January.

Baldwin, J. R., Gray, T., and Johnson, J. (1995), "Technology use, training and plant-specific knowledge in Manufacturing Establishments", *WP Statistics Canada,* No 86.

Barrios, S. Görg, H. and Strobl, E. (2003), "Explaining Firms' Export Behaviour: R&D, Spillovers and the Destination Market", *Oxford Bulletin of Economics and Statistics*, vol. 65, p. 475-496, September.

Barron, J.M., Black, D.A and Loewenstein, M.A. (1987), "Gender Differences in Training, Capital, and Wages", *Journal of Human Resources*, Vol. 28 (2), p. 343-364, Spring.

Bartel, A. (1989), "Formal employee training programs and their impact on labor productivity: evidence from a human resources survey", *NBER Working Paper Series*, No. 3026.

Belsley, D., Kuh, E. and Welsch R. (1980), *Regression Diagnostics, Identifying Influential Data and Sources of Collinearity.* New York: John Wiley and Sons.

Black, D., B. Noel, and Z. Wang (1999). "On-the-Job Training, Establishment Size, and Firm Size: Evidence for Economies of Scale in the Provision of Human Capital," *Southern Economic Journal*, July 66(1) 82-100

Black, S.E. and L.M. Lynch (1998), "Beyond the Incidence of Training: Evidence from a National Employers' Survey", *Industrial and Labor Relations Review*, Oct. 1998, pp. 64-81.

Black, S.E. and Lynch, L.M. (2004), "What's driving the new economy?: the benefits of workplace innovation", *The Economic Journal*, Vol. 114(493), p. F97-F116, February.

Blinder, A. S. (1973), "Wage Discrimination: Reduced Form and Structural Estimates," *The Journal of Human Resources,* vol. 8(4), p. 436–455.

Blundell, R., Dearden, L., Meghir, C. and Sianesi, B. (1999). Human Capital Investment: The Returns from Education and Training to the Individual, the Firm and the Economy, *Fiscal Studies*, vol. 20, no. 1, pp. 1–23

Cameron, A.C. and P.K. Trivedi (2005), Microeconometrics, Princeton University Press, Princeton.

Castany, L., López-Bazo, E. and Moreno, R. (2007), "Decomposing differences in total factor productivity across firm size", Working Paper XREAP2007-01.

Cragg, J. G. (1971). "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods." *Econometrica* 39: 829-44.

Dow, W., and Norton, E. (2003), "Choosing Between and Interpreting the Heckit and Two-Part Models for Corner Solutions", *Health Services and Outcomes Research Methodology*, vol. 4(1), p. 5-18.

Evans, D. S. and Leighton, L. S. (1989), "Why Do Smaller Firms Pay Less?," *Journal of Human Resources*, vol. 24(2), p. 299-318.

Fariñas, J. C. and Jaumandreu, J. (1999), "Diez años de Encuesta sobre Estrategias Empresariales (ESEE)", *Economía Industrial*, No. 329, p. 29–42.

Gardeazábal, J. and Ugidos, A. (2004), "More on Identification in Detailed Wage Decompositions", *The Review of Economics and Statistics*, vol. 86(4), p. 1034–1036.

Görg, H. and Strobl, E. (2005), "Do Government Subsidies Stimulate Training Expenditure? Microeconometric Evidence from Plant Level Data", *IZA DP* No. 1606.

Groot and Maassen van den Brink (2003), "Firm-related training tracks: a random effects ordered probit model," *Economics of Education Review*, vol. 22, p. 581-589.

Guilkey, D. K. and Murphy, J. L. (1993), "Estimation and testing in the random effects probit model", *Journal of Econometrics*, Vol. 59, p. 301–317.

Hashimoto, M. (1979), "Bonus Payments, on-the-Job Training, and Lifetime Employment in Japan", *Journal of Political Economy*, Vol. 87(5), pp. 1086-1104.

Heckman, James J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, Vol. 47, No. 1 (Jan., 1979) , pp. 153-162

Holtmann, Alphonse G. and Todd L. Idson (1991), "Employer Size and On-the-Job Training Decisions", *Southern Economic Journal*, Vol. 58, No. 2 (Oct., 1991) , pp. 339-355

Huergo, E. and Jaumandreu, J. (2004b), "How does probability of innovation change with firm age?," *Small Business Economics*, vol. 22, p.193–207.

Hughes, G. O´Connell, P.J. and Williams, J. (2004), "Company training and low-skill consumer-service jobs in Ireland," *International Journal of Manpower*, Vol. 25(1), p. 17-35.

Korpi, T. and Mertens, A. (2004), "Training and Industrial Restructuring: Structural Change and Labour Mobility in West Germany and Sweden", *International Journal of Manpower,* v. 25 (1), pp. 90-103.

Leung, S. F., and S. Yu, "On the choice between sample selection and two-part models", *Journal of Econometrics*, 72, 1996:197-229.

Li, Y., Zhao, Y., and Liu, Y. (2006), "The Relationship between HRM, Technology Innovation and Performance in China", *International Journal of Manpower,* v. 27 (7-8), pp. 679-97.

Licandro, O., Maroto, R. and Puch, L. (2004), "Innovation, Investment and Productivity: Evidence From Spanish Firms." *EUI Working Paper* Eco No. 2004/7.

Lisbon Strategy, "Convergencia y Empleo. Programa Nacional de Reformas para España", Octubre (2005), http://www.fundaciontripartita.org

Lynch, L.M., (1993), "The Economics of Youth Training in the United States," *Economic Journal*, Royal Economic Society, vol. 103(420), p. 1292-1302, September.

Máñez, J. A., Rochina, M. E., Sanchis, A. and Sanchis, J. A. (2004), "A Dynamic Approach to the Decision to Invest in R&D: the role of sunk costs", mimeo.

Motellón, E. and López-Bazo, E. (2005), "Discriminación por género en el acceso a la contratación indefinida", VIII Encuentro de Economía Aplicada, Murcia.

Oaxaca, R. (1973), "Male-Female Wage Differentials in Urban Labor Markets", *International Economic Review,* vol.14, p. 693-709.

Oaxaca, R., and M. R. Ransom (1994), "On discrimination and the decomposition of wage differentials" *Journal of Econometrics,* vol. 61, p.5-21.

Oi, W.Y., (1983), "Heterogeneous firms and the organization of production", *Economic Inquiry,* vol. 21 (2), p.147-171.

Oosterbeek, H. (1996). "A Decomposition of Training Probabilities," *Applied Economics*, Taylor and Francis Journals, vol. 28(7), p. 799-805, July.

Osterman, P. (1995), "Skill, Training, and Work Organization in American Establishments", *Industrial Relations*, Vol. 34(2), p. 125-46, April.

Peraita, C. (2005): «Firm sponsored training in regulated labor markets: evidence from Spain», *Applied Economics*, 37, 1885-1898.

Rigby, M (2004), "Training in Spain: an evaluation of the continuous training agreements (1993-2001) with particular reference to SMEs", Human Resource Development International, Vol. 7(1), p. 23-37.

Schumpeter, J.A. (1942), Capitalism, Socialism and Democracy, New York, Harper and Brothers.

Smith, V., Dilling-Hansen, M., Eriksson, T. and Strojer-Madsen, E. (2004), "R&D and productivity in Danish firms: some empirical evidence", Applied Economics, vol.36, p.1797–1806.

Toro-Vizcarrondo, C., and Wallace, T. (1968), "A Test of the Mean Square Error Criterion for Restrictions in Linear Regression", Journal of the American Statistical Association, Vol. 63(322), p.558-572.

Wagar, T.H. (1997), "Determinants of formal training in large Canadian organizations," *International Journal of Training and Development,* Vol. 1(2), p. 82-90.

Yun, M. (2004), "Decomposing differences in the first moment," *Economics Letters*, Elsevier, vol. 82(2), p. 275-280, February.

Zábojník, J. and Bernhardt, D. (2001), "Corporate Tournaments, Human Capital Acquisition, and the Firm Size–Wage Relation", *The Review of Economic Studies,* vol. 68 (3), p. 693-716.

Appendix. Decomposition for the two-part model. Detailed results

[Insert Table A1 about here]

Table 1. Expenditure on training by firms' characteristics.

	dTR	Eq prop test	Training/worker (euros)	Eq mean test	Num of obs
Total sample	40.5%		72.4		3,020
% of white collars – low	24.4%	18.0***	28.8	11.9***	1,511
% of white collars – high	56.6%		116.2		1,509
Advanced technology – low	24.8%	14.9***	43.8	5.1***	1,683
Advanced technology – medium	54.2%		90.5		906
Advanced technology – high	72.6%	6.4***	146.3	3.5***	431
No innovation	27.4%	17.7***	44.5	8.5***	1,790
Innovation	59.5%		113.1		1,230
National market	29.8%	17.4***	53.6	7.3***	2,047
International market	63.0%		112.0		973
% of foreign capital – low	30.8%	20.7***	50.4	11.4***	2,368
% of foreign capital – high	75.6%		152.5		652
% of temporary workers – high	37.9%	2.8***	54.6	4.8***	1,510
% of temporary workers – low	43.0%		90.3		1,510

Note: (***) denotes significant at 1%.

Table 2. Expenditure on training by firms' characteristics and size.

	dTR	Eq prop test	Training/worker (euros)	Eq mean test	Num of obs
Total sample	40.5%		72.4		3,020
Small	24.9%	26.0***	41.9	10.8***	2,086
Large	75.2%		140.7		934
% of white collars – low - small	13.8%	18.3***	16.5	6.8***	1,188
% of white collars – low - large	63.2%		73.8		323
% of white collars – high - small	39.6%	16.1***	75.4	7.2***	898
% of white collars – high - large	81.5%		176.1		611
Advanced technology – low - small	17.1%	17.3***	28.2	7.9***	1,424
Advanced technology – low - large	67.6%		129.6		259
Advanced technology – medium - small	40.4%	9.9***	64.8	3.6***	532
Advanced technology – medium - large	73.8%		127.0		374
Advanced technology – high - small	47.7%	7.6***	97.7	2.4***	130
Advanced technology – high - large	83.4%		167.4		301
No innovation - small	17.5%	18.6***	26.2	5.5***	1,425
No innovation - large	66.0%		115.9		365
Innovation - small	41.0%	14.3***	75.6	6.0***	661
Innovation - large	81.0%		156.7		569
National market - small	20.0%	18.9***	31.9	6.7***	1,625
National market - large	67.3%		137.3		422
International market - small	42.3%	12.7***	77.0	5.0***	461
International market - large	81.6%		143.5		512
% of foreign capital – low - small	20.7%	20.6***	32.4	6.3***	1,875
% of foreign capital – low - large	69.0%		119.0		493
% of foreign capital - high - small	62.1%	5.6***	126.3	2.3***	211
% of foreign capital - high - large	82.1%		165.0		441
% of temporary workers - high – small	23.4%	18.1***	34.9	7.3***	1,068
% of temporary workers - high - large	73.1%		102.1		442
% of temporary workers – low - small	26.5%	18.6***	49.2	8.4***	1,018
% of temporary workers – low - large	77.0%		175.5		492

Note: (***) denotes significant at 1%.

Table 3. Firms' characteristics by firm size

	Total sample		Small firms		Large firms		
	Mean	Std dev	Mean	Std dev	Mean	Std dev	Eq mean test (•)
Size	242.5	698.7	47.2	46.8	678.6	1,139.9	16.9***
% of white collars	11.3	12.8	9.7	12.6	14.7	12.8	10.0***
Advanced technology – low	0.6	0.5	0.7	0.5	0.3	0.4	22.7***
Advanced technology - medium	0.3	0.5	0.3	0.4	0.4	0.5	7.8***
Advanced technology - high	0.1	0.3	0.1	0.2	0.3	0.5	16.1***
Innovation	0.4	0.5	0.3	0.5	0.6	0.5	15.4***
International market	0.3	0.5	0.2	0.4	0.5	0.5	17.5***
% of foreign capital	19.2	38.2	8.5	26.7	43.1	48.0	20.7***
% of temporary workers	18.2	21.6	19.8	23.8	14.7	15.3	7.1***
Num of obs	3,020		2,086		934		

Note: (***) denotes significant at 1%.

Table 4. Estimation of the determinants of training. The Heckit and two-part models.

	Participation equation		Quantity equation	
	Marginal effects	Coefficients	Heckit model Coefficients	Two-part model Coefficients
Size	0.1318***	0.3507***	0.0812	-0.0246
	(0.0107)	(0.0285)	(0.0752)	(0.0395)
White collars	0.0053***	0.0142***	0.0236***	0.0198***
	(0.0009)	(0.0024)	(0.004)	(0.0033)
Advanced technology - medium	0.1324***	0.3463***	0.1167	-0.0043
	(0.025)	(0.065)	(0.1158)	(0.0898)
Advanced technology - high	0.1511***	0.3893***	0.2274*	0.1164
	(0.0366)	(0.0927)	(0.1271)	(0.1094)
Innovation	0.1668***	0.4411***	0.3209***	0. 1935***
	(0.022)	(0.0585)	(0.1104)	(0.0806)
International market	0.1109***	0.2912***	0.1719*	0.0792
	(0.0244)	(0.0634)	(0.098)	(0.0819)
Foreign capital	0.001***	0.0027***	0.0029***	0.0023***
	(0.0003)	(0.0009)	(0.001)	(0.0009)
Temporary workers	-0.0002	-0.0005	-0.0065***	-0.006**
	(0.0006)	(0.0015)	(0.0025)	(0.0031)
Controls				
Productive capacity	-0.0012	-0.0032	-0.0009	0
	(0.0008)	(0.0021)	(0.003)	(0.0029)
Group	0.0361	0.0958	0.1624	0.1218
	(0.0299)	(0.079)	(0.1032)	(0.0991)
Year	-0.0307	-0.0816	-0.2731***	-0.2525***
	(0.0208)	(0.0555)	(0.0735)	(0.0719)
Sector dummies	yes	yes	yes	yes
Region dummies	yes	yes	yes	yes
Constant		-2.0382***	2.8695***	3.9481***
		(0.4101)	(0.8945)	(0.605)
Num of obs	3,020		1,222	1,222
Pseudo R	0.34		-	0.17
Pseudo lnL	-1,338.83		-	-
Correlation (ρ)	-		0.44	-
σ2	-		1.29	-
σ12	-		0.57	-
			(0.35)	
H0: Sector=0	47.80***		56.89***	3.11***
H0: Region=0	67.08***		30.48***	1.83**

Note: standard deviation in parentheses; (***) (**) and (*) denote significant at 1%, 5% and 10%.

Table 5. Estimation of the determinants of training. The two-part model with random effects.

	Total sample			**Small firms**			**Large firms**		
	Participation eq		Quantity eq	Participation eq		Quantity eq	Participation eq		Quantity eq
	Mg Eff	Coeff	Coeff	Mg Eff	Coeff	Coeff	Mg Eff	Coeff	Coeff
Size	0.2024***	0.6273***	-0.0341	0.0662***	0.6955***	-0.2347**	0.0681	0.379***	0.0299
	(0.0212)	(0.0683)	(0.0484)	(0.0149)	(0.1161)	(0.1075)	(0.1897)	(0.154)	(0.082)
White collars	0.0089***	0.0276***	0.0195***	0.0036***	0.0375***	0.0197***	0.0011	0.0062	0.0177***
	(0.0017)	(0.0053)	(0.0038)	(0.0009)	(0.0069)	(0.0048)	(0.0034)	(0.0082)	(0.0058)
Advanced technology	0.2420***	0.7019***	0.0436	0.109***	0.8085***	-0.0142	0.0784	0.4604**	0.1931
-medium	(0.0518)	(0.1471)	(0.1123)	(0.0352)	(0.1924)	(0.1536)	(0.2235)	(0.2367)	(0.1727)
Advanced technology	0.2801***	0.7696***	0.1576	0.136*	0.8152***	0.1408	0.0979	0.6219**	0.2403
-high	(0.0794)	(0.2073)	(0.1395)	(0.0805)	(0.3249)	(0.2488)	(0.2868)	(0.2709)	(0.183)
Innovation	0.1928***	0.5832***	0.1414*	0.0738***	0.6272***	0.0567	0.1026	0.529***	0.2244**
	(0.0365)	(0.1078)	(0.0795)	(0.0239)	(0.1406)	(0.1285)	(0.2662)	(0.1701)	(0.0968)
International market	0.1835***	0.5413***	0.096	0.0458*	0.397**	0.2674*	0.1205	0.6401***	0.051
	(0.0449)	(0.1282)	(0.0904)	(0.0254)	(0.1773)	(0.1543)	(0.3155)	(0.1889)	(0.1121)
Foreign capital	0.0014**	0.0043***	0.0022**	0.0007**	0.0075**	0.0047**	0.0005	0.0028	0.0013
	(0.0006)	(0.0018)	(0.0011)	(0.0003)	(0.0032)	(0.0022)	(0.0014)	(0.0021)	(0.0013)
Temporary workers	-0.0001	-0.0002	-0.0066*	0.0000	0.0005	-0.002	-0.0005	-0.0029	-0.0156***
	(0.0010)	(0.003)	(0.0034)	(0.0003)	(0.0036)	(0.0042)	(0.0018)	(0.0062)	(0.0053)
Controls									
Productive capacity	-0.0013	-0.004	-0.002	-0.0002	-0.0024	-0.0016	-0.0010	-0.0056	-0.0036
	(0.0013)	(0.0041)	(0.0031)	(0.0005)	(0.0051)	(0.0043)	(0.0031)	(0.0075)	(0.0042)
Group	0.0682	0.2078	0.113	0.0137	0.1339	0.0297	0.0114	0.0621	0.1983
	(0.0572)	(0.1717)	(0.1234)	(0.0268)	(0.2435)	(0.193)	(0.0555)	(0.2455)	(0.1674)
Year	-0.0462*	-0.1431*	-0.2154***	-0.0062	-0.0656	-0.2004**	-0.0541	-0.3019***	-0.2147***
	(0.0245)	(0.0761)	(0.0507)	(0.0093)	(0.0973)	(0.0937)	(0.1502)	(0.1267)	(0.0587)
Sector dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Region dummies	yes	yes	yes	yes	yes	yes	yes	yes	yes
Random effects	yes	yes	yes	yes	yes	yes	yes	yes	yes
Constant		-4.0818	4.2892***		-5.0014***	4.3705***		5.7329	4.7874***
		(0.9042)	(0.709)		(1.1196)	(0.9879)		(5.09)	(0.9618)
Num of obs	3,020		1,222	2,086		520	934		702
Num of firms	1,538		734	1,068		335	493		409
Pseudo lnL	-1,223.59			-777.5287			-421.57		
H0:Sector=0	30.81**		39.15**	25.84		40.77***	15.08		47.09***
H0:Region=0	41.52***		20.31	36.91***		16.77	13.46		30.35***
H0:RE=0	230.48***		68.17***	164.98***		15.72***	48.31***		42.65***

Note: standard deviation in parentheses; (***) (**) and (*) denote significant at 1%, 5% and 10%.

Table 6. Decomposition of the training gap between small and large firms

	OLS estimation				**RE estimation**			
	Participation eq		Quantity eq		Participation eq		Quantity eq	
Training differential	**2.7414**		**0.4393**		**2.7414**		**0.4393**	
	Charact	Impact	Charact	Impact	Charact	Impact	Charact	Impact
Total	0.464	0.026	0.294	0.145	0.523	0.005	0.261	0.144
	94.69%	5.31%	66.98%	33.02%	99.05%	0.95%	64.43%	35.56%
White collars	0.015	-0.009	0.005	-0.004	0.018	-0.003	0.004	-0.030
	3.07%	-1.80%	1.03%	-0.84%	3.43%	-0.55%	1.10%	-7.42%
Advanced Technology	0.095	-0.004	0.052	0.018	0.118	-0.006	0.081	0.022
	19.41%	-0.73%	11.91%	4.19%	22.32%	-1.10%	19.89%	5.32%
Innovation	0.055	-0.002	0.052	0.021	0.045	-0.001	0.038	0.030
	11.13%	-0.47%	11.94%	4.67%	8.47%	-0.26%	9.47%	7.31%
International Market	0.040	0.005	0.035	0.035	0.046	0.005	0.042	0.034
	8.22%	0.92%	7.95%	7.99%	8.80%	1.03%	10.45%	8.46%
Foreign capital	0.02	-0.001	0.059	-0.105	0.019	0.000	0.058	-0.096
	4.04%	-0.19%	13.50%	-23.89%	3.69%	-0.03%	14.24%	-23.77%
Temporary workers	0.001	-0.001	0.015	-0.175	0.000	0.000	0.017	-0.198
	0.12%	-0.23%	3.48%	-39.77%	0.03%	-0.04%	4.17%	-48.99%

Note: given that the decomposition is not exact in the case of using the RE estimates, the sum of the shares of the components does not equal 100%.

Table A.1. Decomposition of the training gap between small and large firms. Complete results

	OLS estimation						**RE estimation**					
	Participation eq			Quantity eq			Participation eq			Quantity eq		
Training differential	**2.7414**			**0.4393**			**2.7414**			**0.4393**		
	Charact	Advant	Disadv	Charact	Advant	Disadv	Charact	Advant	Disadv	Charact	Advant	Disadv
Total	0.464	0.033	-0.007	0.294	-0.009	0.154	0.523	0.017	-0.012	0.261	-0.018	0.162
	94.69%	6.71%	-1.40%	66.98%	-2.13%	35.16%	99.05%	3.16%	-2.21%	64.44%	-4.49%	40.05%
White collars	0.015	-0.013	0.005	0.005	-0.002	-0.002	0.018	-0.009	0.006	0.004	-0.028	-0.002
	3.07%	-2.73%	0.93%	1.03%	-0.38%	-0.46%	3.43%	-1.72%	1.17%	1.10%	-6.98%	-0.44%
Advanced Technology	0.095	-0.001	-0.003	0.052	0.03	-0.011	0.118	-0.001	-0.005	0.081	0.03	-0.009
	19.41%	-0.19%	-0.54%	11.91%	6.78%	-2.59%	22.32%	-0.20%	-0.90%	19.89%	7.53%	-2.21%
Innovation	0.055	-0.001	-0.002	0.052	0.017	0.003	0.045	0	-0.001	0.038	0.026	0.004
	11.13%	-0.15%	-0.33%	11.94%	3.91%	0.76%	8.47%	-0.06%	-0.20%	9.47%	6.42%	0.89%
International Market	0.001	-0.003	0.002	0.015	-0.114	-0.061	0	-0.001	0.001	0.017	-0.124	-0.074
	0.12%	-0.57%	0.34%	3.48%	-25.9%	-13.9%	0.03%	-0.21%	0.17%	4.17%	-30.6%	-18.4%
Foreign capital	0.04	0	0.004	0.035	-0.009	0.044	0.046	0	0.005	0.042	-0.009	0.043
	8.22%	0.10%	0.83%	7.95%	-2.11%	10.11%	8.80%	0.05%	0.98%	10.46%	-2.12%	10.58%
Temporary workers	-0.001	-0.004	0.003	0	-0.084	0.06	-0.001	-0.004	0.008	-0.003	-0.134	-0.03
	-0.25%	-0.75%	0.54%	-0.01%	-19.0%	13.71%	-0.17%	-0.74%	1.57%	-0.84%	-32%	-7.44%
Size	0.02	-0.002	0.002	0.059	-0.045	-0.06	0.019	-0.002	0.002	0.058	-0.044	-0.053
	4.04%	-0.51%	0.32%	13.50%	-10.3%	-13.7%	3.69%	-0.36%	0.33%	14.24%	-10.8%	-13.0%
Productive capacity	0.024	-0.002	0.003	0.111	0.059	-0.037	0.033	-0.002	0.003	0.103	0.052	-0.025
	4.96%	-0.48%	0.55%	25.28%	13.36%	-8.37%	6.20%	-0.44%	0.58%	25.45%	12.97%	-6.09%
Group	0.201	-0.045	0.005	-0.055	0.209	0.743	0.223	-0.044	0.015	-0.076	0.397	0.799
	41.10%	-9.22%	1.03%	-12.5%	47.55%	169.1%	42.32%	-8.29%	2.85%	-18.8%	98.13%	197.3%
Sector	0.015	-0.02	0.002	0.057	0.201	-0.108	0.021	-0.01	0.001	0.058	0.205	-0.081
	3.16%	-4.08%	0.48%	13.02%	45.70%	-24.7%	3.98%	-1.84%	0.14%	14.42%	50.66%	-20.1%
Region	-0.002	-0.325	0.008	-0.039	-0.918	-0.337	-0.001	-0.192	0.012	-0.063	-0.89	-0.328
	-0.38%	-66.3%	1.73%	-8.93%	-209%	-76.7%	-0.11%	-36.4%	2.30%	-15.4%	-219%	-80.9%
Year	0.001	0	0	0.001	0	-0.001	0.001	0	0	0.001	0	0
	0.10%	0.04%	0.01%	0.29%	-0.03%	-0.18%	0.10%	0.02%	0.01%	0.27%	0.00%	-0.07%

Note: given that the decomposition is not exact in the case of using the RE estimates, the sum of the shares of the components does not equal 100%.