The Impact of Training on Productivity: Evidence from a Large Panel of Firms*

Emilio Colombo† and Luca Stanca‡

January 2008

Abstract

This paper investigates the effects of training on labor productivity using a unique nationally representative panel of Italian firms for the period 2002 to 2005. We find that training has a positive and significant effect on productivity. Using a variety of panel estimation techniques, we show that failing to account for unobserved heterogeneity leads to overestimate the impact of training on productivity, while failing to account for endogeneity leads to substantially *underestimate* it. Training also has a positive and significant impact on wages, but this effect is about half the size of the effect on productivity. Within occupational groups, the effect of training on productivity is large and significant for blue-collars, but small and not significant for white collars.

JEL Classification: C23, D24, J31.
Keywords: On-the-Job-Training; Productivity; Wages; Panel Data.

*We thank Unioncamere for granting access to the Excelsior dataset. We also thank Pietro Aimetti, Cecilia Corrado and Andrea Gianni for advice on the project and assistance with the Excelsior data set. The views expressed are those of the authors and do not necessarily reflect the position of Unioncamere. The usual disclaimer applies.

†Economics Department, University of Milan Bicocca. Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy. E-mail: emilio.colombo@unimib.it

‡Corresponding author. Economics Department, University of Milan Bicocca. Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy. E-mail: luca.stanca@unimib.it

1 Introduction

Human capital is widely acknowledged as a key factor for economic performance at both the micro and macro level. Despite the fact that a large fraction of human capital accumulation takes place after the entry into the labor market, most of the existing literature that investigates the returns to investment in human capital has focused on education, due to measurement problems and data availability. Relatively little evidence is available, instead, on the accumulation of human capital through the lifelong training of workers and, more specifically, on the effects of training on productivity.

A number of studies have tried to fill this gap by analysing the impact of training on productivity using firm-level data. However, this literature does not provide a consistent picture, as the lack of longitudinal data has generally made it difficult to control for unobserved heterogeneity and endogeneity of training (e.g. Bartel, 1994, Bishop, 1994, Black and Lynch, 1996, Barrett and O'Connell, 2001). Some recent studies have tackled this problem by focusing on panel data at industry-level (e.g. Dearden et al., 2006, Conti, 2005). This approach, however, does not allow to estimate the private returns to training, as analyses based on industry-level data also capture spillover effects between firms. There exists a recent literature that investigates the effects of training on productivity using firm-level panel data, but it is generally hampered either by the specificity of the sample (e.g. Almeida and Carneiro, 2006), or by the limited number of observations in the sectional dimension (Ballot et al., 2006; Zwick, 2005, 2006) or in the time dimension (Black and Lynch, 2001).

This paper investigates the effects of training on productivity using a unique nationally representative sample of Italian firms in the period 2002 to 2005. Our paper contributes to the existing literature in several ways. Our study is the first in the literature to be based on a large and nationally representative panel data set at firm-level. The availability of longitudinal information on training and productivity allows us to deal with both unobserved heterogeneity and endogeneity of training, using a variety of panel estimation techniques. In addition, firm-level data on training and direct measures of productivity allow us to estimate the private returns to training for employers, while netting out the possible spillover effects that may lead to over-estimation when using industry-level data.

Second, we examine whether training has different returns for employers and employees, by comparing the effects of training on direct measures of labor productivity with the results obtained from the corresponding wage equations. We also check the robustness of the results in the baseline specification by allowing for different types of labor, and by focusing on sub-samples defined on the basis of firms characteristics (size, industry, region). In addition, we are able to account for the duration of training by constructing an alternative indicator of training intensity, the average number of days of training per worker, and comparing its effects on productivity with those of the standard indicator of training intensity generally used in the literature.

Third, Italy provides a particularly interesting case study for at least two reasons. On the one hand, it is one of the countries with the lowest incidence of on-the-job training in Europe (Bassanini et al., 2005). It is therefore interesting to assess to what extent this feature can affect the relationship between firms' training and productivity, while obtaining an indication of the efficiency costs implied by sub-optimal investment in training. On the other hand, the Italian labor market is known to be characterised by severe rigidities. Comparing the effect of training on productivity and wages allows to assess the effect of labor market rigidities on how the returns to training are shared between the firm and the workers.

Our results indicate that increasing the share of employees participating in training activities has a positive and significant effect on productivity at firm-level. When training intensity increases by 1 percentage point, productivity increases by about 0.07 per cent. Training intensity also has a significant effect on wages, but using wages as an indirect measure of productivity leads to substantially underestimate the impact of training. Within occupational groups, training has large and significant effects for blue-collar workers, while the effects for executives and clerks are negligible. We also show that using an indicator of training that does not account for training duration may lead to underestimate the effects of training on productivity.

More generally, the comparison of the results obtained with a variety of panel estimators indicate that, for both the productivity and the wage equation, failing to account for unobserved heterogeneity leads to overestimate the impact of training on productivity. However, failing to account for potential endogeneity of training leads to *underestimate* its impact on productivity: the estimated effect of training on productivity almost doubles when training is treated as a choice variable and the time dimension of the panel data set is exploited to obtain appropriate instruments.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 discusses the methodology used in our analysis. Section 4 describes the data set. The results of the econometric analysis are presented in section 5. Section 6 concludes with a discussion of the implications of the analysis. The data appendix provides more detailed information on the composition and representativeness of the data set.

2 Previous literature

A vast empirical literature has investigated the effects of training using wages as a proxy for productivity, generally finding that different types of training result in significantly higher earnings for workers.[1] Wages, however, only provide an indirect measure of productivity. The real wage rate is assumed to be equal to the marginal product of labor if the labor market is perfectly competitive and under restrictive assumptions about training. More generally, the benefits of training are shared between employers and employees depending on labor market imperfections, whether training is specific or general, and who pays for the costs of training (e.g. Acemoglu and Pischke, 1999, Booth et al. 2003), so that wage equations do not provide an appropriate indication of the effects of training on productivity. In recent years, as detailed firm-level data sets have become available, a number of studies have investigated the effect of training on direct measures of labor productivity.

Early studies are generally based on cross-sectional data (see Bartel, 2000, for a review of this literature). Bartel (1994) uses a survey dataset on personnel policies and economic characteristics of 495 manufacturing firms, finding that companies that implemented formal training programs experienced a 6 percent higher annual productivity. Black and Lynch (1996) use establishment-level data for 1993 to estimate a Cobb-Douglas production function that includes indicators of training, in addition to several controls for firms' characteristics. The results indicate that off-the-job training and computer training have a positive effect on productivity in manufacturing and non-manufacturing establishments, respectively. Barrett and O'Connell (2001) analyze a sample of about 700 Irish firms, finding that general training has a positive impact on productivity, whereas specific training does not have an impact on productivity.

The main problem with studies based on cross-section analysis is that they cannot control for the possible endogeneity of training. Training decisions can be endogenous for two reasons: on the one hand, there can be unobserved heterogeneity, if there are time-invariant company characteristics unaccounted for that determine both training and firms' economic performance (e.g. managers' quality, technological level). On the other hand, training may be a choice variable, so that idiosyncratic shocks at firm level affect both training decisions and productivity. The availability of panel data at the micro level allows to deal with both these problems.[2]

[1]See e.g. Bartel (1994, 2000), Blanchflower and Lynch (1994), Blundell et al. (1996), Booth (1991), Lynch (1992), Lillard and Tan (1992), Tan et al. (1992).

[2]See e.g. Ichniowki et al. (1997), Carriou and Jeger (1997) and Delame and Kramarz

Black and Lynch (2001) use panel data for 627 US establishments, with information on training derived from a survey administered for two years. They implement a two-step procedure to account for the possible endogeneity of training, finding a positive effect of training on productivity in the cross-section analysis, but no significant effect when controlling for unobserved fixed effects.[3] A similar approach is followed by Zwick (2005, 2006), who finds a positive effect of training intensity on productivity in a larger sample of German firms. Ballot et al. (2006) analyse a sample of French and Swedish firms over the period 1987-1993, finding that training has a positive effect on productivity in France but a non-significant effect in Sweden. Almeida and Carneiro (2006), using a panel of about 1,500 large Portugues manufacturing firms between 1995 and 1999, find that an increase of 10 hours per year in training per worker leads to an increase in productivity of about 0.6 per cent.

Most of this recent literature using firm-level panel data suffers from limitations related to the specificity of the sample, the limited number of observations in the time or sectional dimension, or the limited availability of information on the type of training activity. Some recent studies have tackled these problems by constructing panel data sets at industry-level, matching information on training at employee-level with information on productivity at firm-level. Dearden et al. (2006) combine British data on productivity from firms' company accounts with information on training from labor force surveys. By aggregating information into clusters at regional and sectoral level, they obtain a panel of industries with a significant time dimension (1983-1996). By estimating a Cobb-Douglas production function, they find that increasing the proportion of trained workers by 1 percentage point leads to a 0.6 per cent increase in value added per hour, and a 0.3 per cent increase in hourly wages. Conti (2005) applies a similar methodology to construct a panel of Italian industries for the years 1996-1999, obtaining similar results. Despite the advantage of fully exploiting the time dimension and the detailed information about training, this approach has important shortcomings. By aggregating data at industry-level, there is a loss of micro-level information that may result in aggregation biases. More importantly, analyses based on industry-level data are likely to capture spillover effects between firms that may lead to over-estimate the private returns to training.

(1997) for early studies based on firm-level panel data.

[3] In the first step, they estimate a Cobb-Douglas production function with fixed effects, omitting variables with little time variability (including training intensity). In the second step, the quasi time-invariant variables are regressed over the residuals of the first step.

3 Methodology

The econometric analysis in this paper follows the literature in assuming that technology at firm-level can be characterized by a Cobb-Douglas production function (e.g. Dearden et al., 2006):

$$Y = AL^{\alpha}K^{\beta} \tag{1}$$

where Y, L, and K, are value added, labor and capital, respectively, A represents technological progress, α and β denote the elasticity of value added with respect to capital and labor.

Under the assumption that trained and untrained workers have different productivities, effective labor can be written as:

$$L = N^{U} + \gamma N^{T} \tag{2}$$

where N^T and N^U represent trained and untrained workers, respectively, L is effective labor, and γ is a parameter that characterizes trained workers' relative productivity. This parameter will be greater than 1 if trained workers are more productive than untrained workers.

Substituting equation (2) in equation (1) we obtain

$$\begin{aligned} Y &= A\left[N^{U} + \gamma N^{T}\right]^{\alpha} K^{\beta} = \\ &= A\left[1 + (\gamma - 1)\frac{N^{T}}{N}\right]^{\alpha} N^{\alpha}K^{\beta} \end{aligned} \tag{3}$$

where N is the total number of workers and $\frac{N^T}{N}$ is the fraction of trained workers over the total. Under the assumption of constant returns to scale ($\alpha + \beta = 1$) we can write the production function in intensive form and express labor productivity as follows:

$$\frac{Y}{N} = A\left[1 + (\gamma - 1)\frac{N^{T}}{N}\right]^{\alpha}\left(\frac{K}{N}\right)^{\beta} \tag{4}$$

Applying a log-transformation and approximating around 1, we obtain:

$$\log\left(\frac{Y}{N}\right) = \log(A) + \alpha(\gamma - 1)\frac{N^{T}}{N} + \beta\log\left(\frac{K}{N}\right) \tag{5}$$

If trained workers are as productive as untrained workers ($\gamma = 1$), the coefficient of training intensity will be zero. If the labor market is perfectly competitive, the real wage is equal to the marginal product, and a wage equation can be defined analogously to equation (5).

Following a similar approach, we can obtain an expression for labor productivity with different types of workers (e.g. by occupation, gender, etc.). Assuming two labor inputs:

$$L = N^U + \gamma_1 N_1^T + \gamma_2 N_2^T \tag{6}$$

so that the production function can be written as follows:

$$\begin{aligned} Y &= A\left[N^U + \gamma_1 N_1^T + \gamma_2 N_2^T\right]^\alpha K^\beta = \\ &= A\left[1 + (\gamma_1 - 1)\frac{N_1^T}{N} + (\gamma_2 - 1)\frac{N_2^T}{N}\right]^\alpha N^\alpha K^\beta \end{aligned} \tag{7}$$

As above, assuming constant returns to scale, applying the log-tranformation and approximating around 1, we obtain:

$$\log\left(\frac{Y}{N}\right) = \log(A) + \alpha(\gamma_1 - 1)\frac{N_1^T}{N} + \alpha(\gamma_2 - 1)\frac{N_2^T}{N} + \beta\log\left(\frac{K}{N}\right) \tag{8}$$

More generally, with M labor inputs:

$$\log\left(\frac{Y}{N}\right) = \log(A) + \alpha\sum_k^M\left[(\gamma_k - 1)\frac{N_k}{N}\right] + \beta\log\left(\frac{K}{N}\right) \tag{9}$$

Turning to the empirical specification, we estimate the baseline equation in (5) and the multi-factor specification in (9) allowing for differences in labor quality (executives, clerks, workers), while controlling for a number of other factors affecting productivity, captured in A, such as innovation (proxied by research and development and patents expenditures), export activity, and a number of other firm characteristics (size, industry, region, age, part of a group and listed status). The resulting equation to be estimated can be represented as follows:

$$y_{it} = \beta x_{it} + \gamma z_i + \varepsilon_{it} \tag{10}$$

where y is the log of labor productivity, x is a vector of (potentially endogenous) time varying regressors that include training intensity, z is a vector of time invariant firms' characteristics, ε is the error term, i is the individual (firm) index and t the time (year) index.

We assume that the error term includes a time-invariant individual component (α_i), an individual-invariant time effect (τ_t) and an idiosynchratic component (η_{it}):

$$\varepsilon_{it} = \alpha_i + \tau_t + \eta_{it} \tag{11}$$

The appropriate estimation method for equation (10) depends on the assumptions regarding the relationship between training and the error term. If training is strictly exogenous with respect to the idiosynchratic component, the problem is how to deal with the presence of the fixed effects. If the individual effects are not correlated with the regressors, both OLS and GLS are unbiased and consistent, but only the GLS estimator is efficient. If they are correlated with the regressors, the OLS and GLS (random effects) estimators are biased and inconsistent, while the within (fixed effects) estimator is unbiased and consistent.

However, if training is predetermined (training decisions respond to past productivity shocks) or endogenous (training decisions respond to past and current productivity shocks), the fixed effect estimator is inconsistent. In the absence of any obvious instruments, a possible solution is to exploit appropriate moment conditions to construct a GMM estimator. One approach is to remove the fixed effects by taking first differences of equation (10):

$$\Delta y_{it} = \beta \Delta x_{it} + \Delta \tau_t + \Delta \eta_{it} \tag{12}$$

Under the assumption that η_{it} is serially uncorrelated, if x_{it} is predetermined, x_{it-1} and earlier lags provide valid instruments that can be used to construct a GMM estimator for equation (10) in first differences (Arellano and Bond, 1991). Similarly, if x_{it} is endogenous, x_{it-2} and earlier lags provide valid instruments. The available moment conditions can be written as

$$E\left(x_{it-s} \Delta \eta_{it}\right) = 0 \tag{13}$$

with $s \geq 1$ if x_{it} is predetermined and $s \geq 2$ if x_{it} is endogenous.

A well known problem with this estimator is that for variables with a strong persistence over time (such as capital) the correlation between their first difference and the lagged levels used as instruments will be low, resulting in a substantial bias in finite samples. Under the assumption that the initial change in productivity and in the endogenous variables are uncorrelated with the fixed effect, lags of the first differences of the endogenous variables can be used to instrument equation (10) in levels:

$$E\left(\Delta x_{it-s}\left(\alpha_i + \eta_{it}\right)\right) = 0 \tag{14}$$

with $s \geq 0$ if x_{it} is predetermined and $s \geq 1$ if x_{it} is endogenous.

Under these assumptions, the equations in differences (12) and in levels (10) can be combined to obtain a more efficient GMM system estimator, while also providing a way of controlling for transitory measurement error (Blundell and Bond, 1998).

4 Data

The data set was constructed by merging information from two different sources. Firm-level information on training was obtained from Excelsior (see Unioncamere, 2007), a joint project of the Italian ministry of Labor and Unioncamere (Italian Association of Chambers of Commerce). Excelsior is a survey conducted yearly on a sample of approximately 100,000 Italian firms, with the aims of assessing firms' occupational needs and providing detailed information on the qualifications of expected new hirings. The sample includes all firms with more than 50 employees and an unbalanced panel of smaller firms. The selection of small firms is obtained through a stratified sampling method that ensures representativeness of the population.[4]

The Excelsior data set contains a section on training activity that provides detailed information on the number of employees undertaking some form of training. This information is available for the firm as whole and disaggregating by occupation (managers, clerks and workers) and gender. The survey also provides information on the type of training activity (internal and external courses, on the job, self-learning) and, for a subset of firms, the average duration of training (average number of days of training per trained employee) and the cost of training activities. Although the survey has been conducted since 1996, internal data consistency allowed us to retain only the four years between 2002 and 2005. The sample contains all firms that compiled the section on training in the questionnaire for at least two non-consecutive years over the period 2002-2005.[5]

Company account data were obtained from AIDA, a database containing annual accounts for all Italian non-financial firms with a turnover greater than 500,000 Euros.[6] This database was the source for information on value added, capital, labor, R&D expenditure, in addition to size, industry, geographic region, age and other company characteristics (see the Data Appendix for details on the definition and construction of individual variables). Productivity is defined as value added per worker. Capital is measured by the book value of total fixed assets.[7] R&D intensity is expenditure for re-

[4]Firms with less than 50 employees are divided into two classes of 1-9 and 10-49 employees, respectively. For each class, the sample is drawn from about 8,700 cells defined over 43 sectors and 103 regional areas.

[5]Firms that did not undertake any form of training are included in the sample, as they could complete the section on training by answering negatively to the first question: "Did you undertake any form of training on personnel during the last year?".

[6]Aida is provided by Bureau van Dijk (http://www.bvdep.com/en/aida.html).

[7]We experimented with alternative definitions of capital, the paper's findings were unaffected.

search and development and avertising over capital. Patent Intensity is the capitalized costs for patents over capital. The wage rate is the total wage bill divided by the number of employees. All nominal variables were deflated with producer price indices at two digit industry level, obtained from ISTAT, the Italian National Institute of Statistics.

We merged the Excelsior and the Aida data sets by using company tax codes. The resulting data set was thoroughly checked for consistency, leading to eliminate approximately 100 observations that contained incorrect or implausible values. Matching and data validation left us with an unbalanced panel of 11,123 firms for a total of 33,815 observations. The sample coverage is wide and well representative of the population of Italian firms. A detailed description of the composition and structure of the sample is contained in the Data Appendix. Summary statistics for all the variables used in the analysis are reported in table 1.

5 Results

This section presents the results of the econometric analysis. We start by examining the effects of training on productivity and wages in the whole sample, using the baseline specification in (10). Next, we check the robustness of the results by estimating the effects of training separately for different types of labor and sub-samples of firms. Finally, we consider an alternative indicator of training activity, effective training intensity, that takes into account the duration of training activity per worker.

5.1 Baseline specification

Table 2 presents estimation results for equation (10), treating all explanatory variables as exogenous (columns 1-3) or using a fixed effect estimator to take into account unobserved heterogeneity (column 4). All regressions include firms' characteristics (age, export, group and listed status) and a full set of time, industry, region, and size dummy variables. The first two columns report OLS estimates obtained, respectively, without and with the inclusion of occupational shares and indicators of innovation in the set of regressors. Capital per worker is strongly and significantly related to productivity in both specifications, although the coefficient (0.23) is lower than capital's share of value added. The more general specification indicates that labor productivity is positively related to the quality of labor: the share of executives has a positive and strongly significant coefficient, while the coefficient for the share of workers is negative and strongly significant. The coefficient for patents intensity has the expected sign, whereas the coefficient for R&D intensity is negative.[8] The coefficient for training intensity is positive and strongly significant in both specifications estimated by OLS. The point estimate drops from 0.112 in the restricted specification to 0.045 in the general specification, reflecting the relevance of occupational proportions as a proxy for labor quality. The OLS estimates with the full set of controls indicate that raising the training variable by 1 percentage point is associated with an increase in productivity of about 0.05 per cent.

The third and fourth columns report estimates obtained with the random and fixed effects estimators, respectively. Training has a smaller but statistically significant effect on productivity in the efficient random effect estimation. More importantly, the effect of training is strongly significant also in the specification that controls for fixed effects, and the magnitude of the coefficient falls only slightly (0.028). These results indicate that failing to account for unobserved heterogeneity leads to overestimate the impact of training on productivity by about 30 per cent: the estimated increase in productivity associated to a 1 percentage point increase in training intensity falls from 0.045 per cent to 0.028 when we use a fixed effect estimator to take into account unobserved firm characteristics potentially correlated with both training and productivity.

The results reported in Table 2 do not allow for the possible endogeneity of training or other explanatory variables. To deal with this, we implemented the GMM approach described above. Table 3 summarizes the results. The same set of explanatory variables as in Table 2 is included in these specifications. Columns 1 to 3 present the results for the equations in differences (GMM-DIF). The first specification (column 1) assumes that all variables are exogenous with respect to the idiosyncratic component of the error term. Training has a positive and significant impact on productivity, and the point estimate (0.028) is very close to the fixed effect coefficient estimate. The changes in the other coefficients are also relatively small.

The next two columns report estimation results under the assumptions that training and capital are either pre-determined (column 2), or endogenous (column 3). The innovation indicators are assumed to be predetermined, while occupation shares are treated as exogenous. When training is treated as predetermined, using levels dated $t-1$ and $t-2$ as instruments for both capital and training, the estimated effect of training intensity rises to 0.044.

[8]This result could be explained by the fact that the balance sheet item used to construct the R&D intensity indicator also includes expenditures for advertising.

However, the coefficient is estimated less precisely, given that earlier lags are less informative about current differences, and is therefore only marginally significant using a one-tailed test. Although the Hansen test statistic does not lead to reject the null hypothesis of instrument validity for this specification, we also consider the results under the assumption that training is endogenous, using levels dated $t-2$ and $t-3$ as instruments.[9] The coefficient estimate rises further to 0.074, but it is even less precisely estimated, and is not statistically significant.

Columns 4 to 5 present the results for the system estimator (GMM-SYS) that combines equations in differences and levels, assuming that training is either pre-determined or endogenous. When training is treated as predetermined, the coefficient for training intensity almost doubles with respect to the specification in differences. The point estimate is 0.074, and it is strongly statistically significant. The Hansen test statistic does not lead to reject the null hypothesis of instrument validity both for the whole set of instruments (p-value=0.32) and for the equations in levels (p-value=0.29). When training is treated as endogenous, the point estimate rises further to 0.202 but is not statistically significant.

Overall, the results of GMM estimation indicate that failing to account for potential endogeneity of training leads to *underestimate* the impact of training on productivity: the estimated effect on productivity of a 1 per cent increase in training intensity rises from 0.028 to 0.074 per cent when we exploit the time dimension of the panel data set to obtain appropriate instruments for training. This may suggest that a firm increases its training activity when it faces negative market conditions and labor productivity is low.

Table 4 presents results for equation (5), using the log of the wage rate as the dependent variable. We restrict the attention to the OLS, random effects, fixed effects and system-GMM estimators. Similarly to the productivity equation, we find that failing to take into account the endogeneity of training leads to underestimate its effect on wages. The results from both fixed effect and system-GMM estimation indicate that training has a positive and significant effect on wages. However, irrespective of the estimation technique, the effect of training on wages is much smaller than the effect on labor productivity. Coefficient estimates for training intensity range from 0.17 in OLS and RE estimation, to 0.02 in FE and 0.044 in system-GMM

[9] In addition, we assessed the validity of each individual instrument with difference-in-Hansen tests. The AR tests (not reported in the tables) do not reject the null hypothesis that the error term is not serially correlated.

estimation, that is, about half the size of the corresponding estimates for the productivity equations.

This result, consistently with the findings at industry level in Dearden et al. (2006) for the UK and in Conti (2005) for Italy, indicates that using wages as an indirect measure of productivity leads to substantially underestimate the impact of training on productivity. The relative large size of the difference that we obtain between estimates in productivity and wage equations may reflect the institutional features of the Italian labor market, where the centralisation of wage bargaining and the strong role of unions weaken the link between wages and productivity at firm level.

5.2 Further Results

In this section we extend the baseline specification by estimating the effect of training separately for different types of labor. We also check the robustness of our results focusing on different subsamples of firms, restricting the attention to the system-GMM estimator. Table 5 reports estimation results for equation (9), where the effect of training on productivity is allowed to differ depending on the position of employees (executives, clerks, workers). We first consider, in column 1, the disaggregation by skill into two groups: blue collar (manual workers) and white collar (clerks and executives). Interestingly, the results indicate that the effect of training on productivity is large (0.13) and strongly significant for blue collar employees, while it is small and not significant for white collars. We then further disaggregate white collars into executives and clerks (column 2). The effect of training is small and positive for clerks, while negative and large for executives. However, the coefficients are not precisely estimated, and are not statistically significant in either case. The larger productivity effect for blue collar relative to white collar workers could be interpreted by considering that, for white collar jobs, productivity-enhancing skills are generally acquired through advanced education. In the presence of diminishing returns to human capital we should therefore expect a stronger effect of training on productivity for relatively unskilled workers.

In order to further check the robustness of the aggregate results presented above, table 6 presents the results obtained estimating equation (10) by system-GMM for different subsamples, defined according to firms size, industry and geographic region. The results for the sample-split by industry indicate that the effect of training on productivity in the service sectors is statistically significant and larger (0.09) relative to non-service sectors (0.05). Restricting the attention to the manufacturing sector, the coefficient estimate is 0.06, only marginally significant under a one-tailed hypothesis (p-value 1.68). Across regions, the effect of training is large and significant in North and Central regions (0.08 and 0.12, respectively), while small and not significant for firms located in the South. This may indicate that in the South training activity tends to be used to absorb excess labor in periods of negative firm-specific market conditions. Finally, the disaggregation by size shows that the effect of training on productivity is quite similar for small and large firms (0.08), while about half the size (0.04) for medium firms.

5.3 Accounting for training duration

One of the major drawbacks of the use of training intensity as an indicator of training activity, is that it does not take into account the duration of training, thus implicitly assuming that every worker is trained the same number of days per year in all firms. In fact, there is substantial variability in training duration across firms. The Excelsior survey also contains information on the average duration (number of days) of training per worker. This allowed us construct an alternative indicator of training activity to take into account the actual duration of training. Unfortunately, information on training duration is only available for a smaller sample of firms, and not availabe in the year 2002. It is nevertheless informative to check to what extent accounting for training duration would affect our results and, in particular, whether the impact of training on productivity may be underestimated by using a purely quantitative indicator such as training intensity. We therefore constructed a measure of "effective training intensity" by multiplying training intensity by the average number of training days per worker. In order to compare the effects of the standard and alternative measure of training intensity on firm's productivity we standardised the two indicators and restricted the sample to the years 2003-2005. The estimated coefficients should therefore be interpreted as the change in the dependent variable following a one standard deviation change in the corresponding training indicator.

Table 7 presents the results. First, comparing alternative estimators, the coefficient of effective training intensity displays the same pattern identified above: failing to account for the potential endogeneity of training leads to underestimate the effect of training on productivity (point estimates are 0.009, 0.005 and 0.022 for OLS, fixed effects and system GMM, respectively). Second, focusing on the system GMM as our preferred estimates, the coefficient for effective training is larger (0.022) than that of training intensity (0.019), although the difference is quite small. This indicates that using a measure of training intensity that does not account for training duration may lead to underestimate the effects of training on productivity.

6 Conclusions

This paper presented an empirical investigation of the effects of training on labor productivity. Our analysis is based on a large and representative panel data set of Italian firms in the years 2002–2005. The availability of longitudinal data allowed us to deal with the effects of both unobserved heterogeneity and potential endogeneity of training, using a variety of panel estimation techniques. The use of firm-level data allowed us to estimate the private returns to training, netting out possible spillover effects between firms that are captured in similar studies based on industry-level data. We also checked the robustness of the results by considering disaggregations by occupations and firm-characteristics, and by using an alternative indicator of training that takes into account the duration of training.

We find that training has a positive and significant effect on productivity. A one per cent increase in training is associated with an increase in value added per worker of about 0.07 per cent. This result is consistent with the evidence in Dearden et al. (2006), who obtain a coefficient estimate of about 0.6 per cent using a panel of British industries between 1983 and 1996, but find a much smaller effect using individual-level data, concluding that the larger magnitude of the training effects in their paper is largely due to the use of industry-level data.[10] We also find that training has a significant effect on wages. However, irrespective of the estimation technique, this effect is found to be about half the size than the effect on training productivity, reflecting the low flexibility of wages in the Italian labor market. This indicates that using wages as a proxy for productivity may lead to significantly underestimate the impact of training on labor productivity.

More generally, our results indicate that failing to account for unobserved heterogeneity leads to overestimate the impact of training on productivity: the estimated increase in productivity associated to a 1 percentage point increase in training intensity falls from 0.045 per cent to 0.028 when we take into account unobserved firm characteristics potentially correlated with both training and productivity using a fixed effect estimator. However, the results also indicate that failing to account for potential endogeneity of training leads to *underestimate* the impact of training on productivity: the estimated effect on productivity of a 1 per cent increase in training intensity rises to 0.074 per cent when we exploit the time dimension of the panel data set to obtain

[10]The magnitude of our coefficient estimates are not directly comparable with those of Dearden et al (2006), as their indicator of training intensity is constructe as the proportion of workers in an industry who received training over a given 4-week period in the first quarter of the LFS.

appropriate instruments for firms training.

This is an important finding, consistent with the hypothesis that firms engage in training activities in periods of negative demand conditions, when the opportunity cost of training is lower (Bartel, 1991, Black and Lynch, 2001). It indicates that the returns to on-the-job training are likely to be severely underestimated if training is not treated a choice variable. Further research will have to explicity address the simultaneous and dynamic nature of the relationship between training decisions and economic performance.

Data Appendix

This appendix provides further details on the composition and structure of the data set analyzed in this paper. Table 8 describes the composition of the sample by year, size, industry and geographic region. The manufacturing sector accounts for 48.6 per cent of the sample, followed by business services (14.2 per cent) and trade (11.1 per cent). The bulk of economic activity is located in the northern regions, that together account for almost 70 per cent of the firms in the sample. These distributions are largely stable over time. Small firms account for 42.4 per cent of the total in our sample, against 40 and 17.6 per cent of medium and large firms, respectively. Both the turnover threshold in AIDA and the fact that the unbalanced part of the Excelsior survey refers mainly to small firms imply that medium and large firms are relatively over-represented in our sample with respect to the population of Italian firms. The under-representation of small firms should not be a cause of major concern when considering firms' training decisions, given that a large fraction of small firms in Italy have no employees.[11]

Table 9 examines the representativeness of the sample, by comparing the sectoral distribution for selected variables (value added, output, and employment) in the sample and in the population of Italian firms. Overall, the sample provides a good approximation of the sectoral distribution at the national level, with the main exceptions represented by the over-representation of the manufacturing sector, and the under-representation of the financial sector (not included in AIDA).

Table 10 provides a description of the time structure of the data set. Observations are available in all four years 2002-2005 for about 31 per cent of the firms in the sample. Three consecutive observations are available for about 32 per cent of the sample. Note that, due to the highly unbalanced nature of the panel data set, in the econometric analysis the sample size may change substantially depending on the specification adopted.

Table 11 reports the percentage of firms undertaking any training activity (training propensity), disaggregating by year, sector and geographical region. On average, 71.5 per cent of the firms in the sample undertake some form of training. This relatively large figure is in line with the figures reported in Zwick (2006) for Germany. The propensity to train is higher in large firms. The sectoral decomposition reveals a higher propensity to train in the service sector and, in particular, in the advanced service sectors (business

[11] The Italian Statistical Office reports that, in 2004, 67 per cent of all registered firms did not have employees.

services, finance and banking, education, health). Hotels and restaurants and Constructions are the sectors with the lowest training propensity. The propensity to train is clearly higher in the North as opposed to the Centre-South. Table 12 reports the distribution of firms' training intensity. The differences between small and large firms and between North and South are less pronounced, reflecting a composition effect as large firms are located mainly in the North. Overall, the sectoral distribution of training intensity displays a similar pattern to the one for the propensity to train.[12] The average value of training intensity is in line with the industry-level figures, based on individual data, reported for Italy by Conti (2005).[13]

[12]The high training intensity in the mining sector is due to a composition effect since the number of firms belonging to this sector in our sample is extremely limited.

[13]Note that these figures are not directly comparable as Conti (2005) uses labor supply data. We use instead data for labor demand.

References

Acemoglu, D., Pischke, S. (1998) “Beyond Becker: Training in Imperfect Labor Markets”, Economic Journal, 109 (453), F112-F142.

Almeida, R., Carneiro P. (2006) “The Return to Firm Investment in Human Capital”, World Bank Policy Research Working Paper No. 3851

Arellano, M., Bond, S. (1991) “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations”, Review of Economic Studies, 58(2), 277-97.

Ballot, G, Fakhfakh, F., Taymaz, E. (2006) “Who benefits from training and R&D, the firm or the workers?” British Journal of Industrial Relation, 44(3), 473-495.

Barrett, A., O’Connell, P. (2001) “Does training generally work? The returns to in-company training”, Industrial and Labor Relations Review, 54(3), 647-662.

Bartel, A.P. (1994) “Productivity gains from the implementation of employee training programs”, Industrial Relations, 33 (4), 411-425.

Bartel, A. (1995) “Training, Wage Growth, and Job Performance: Evidence from a Company Database”, Journal of Labor Economics, 13(3), 401-425

Bartel, A. (2000) “Measuring the Employer’s Return on Investments in Training: Evidence from the Literature”, Industrial Relations, vol. 39, no. 3, pp. 502-24.

Bishop, J.H., 1994. The impact of previous training on productivity and wages. In: Lynch, L.M. (Ed.), Training and the Private Sector, International Comparisons. University of Chicago Press, Chicago, pp. 161?199.

Black, S., Lynch, E. (1995): Beyond the incidence: evidence from a national employers survey, Nationl Bureau of economic research Working Paper, n.5321

Black, S., Lynch, L. 1996. Human Capital Investment and Productivity. American Economic Review 86(2), 263-267.

Black, S., Lynch, E., 2001. How to compete: the impact of workplace Practices and Information technology on Productivity. The Review of Economics and Statistic 83(3), 434-445.

Blanchflower, D., Lynch, L., 1994. Training at work: a comparison of US and British youths. In: Lynch L., (ed.), International Comparisons of Private Sector Training, University of Chicago Press.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models, Journal of Econometrics, Vol. 87, pp. 115-143.

Blundell, R., Dearden, L., Meghir, C., 1996. The determinants of work-related training in Britain. Institute for Fiscal Studies, London.

Booth, A. (1991), Job-related formal training: who receives it and what is it worth?, Oxford Bulletin of Economics and Statistics, 53, pp.281-94.

Booth, A., Francesconi, M. and Zoega, G. (2003). "Unions, Work-related Training, and Wages: Evidence for British Men" Industrial and Labor Relations Review, 57(1), 68-91.

Brunello, G. (2002) "Is training more frequent when wage compression is higher? Evidence from 11 European countries", CESifo Working Paper, n.637.

Bassanini, A., Brunello, G., Booth, A., De Paola, M., Leuven, E. (2005) "Workplace Training in Europe", IZA Discussion Paper N. 1640.

Carriou, Y., Jeger, F. (1997) "La formation continue dans les entreprises et son retour sur investissement", Economie et Statistique, Vol. 303, pp. 45-58.

Conti, G. (2005) "Training, Productivity and Wages in Italy", Labour Economics 12(4), 557-76.

Dearden, L., Reed, H., Van Reenen, J. (2000) "Who Gains When Workers Train? Training and Corporate Productivity in a Panel of British Industries", IFS Working Paper 00-01.

Dearden, L., Reed, H., Van Reenen, J. (2006) "The Impact of Training on Productivity and Wages: Evidence from British Panel Data", Oxford Bulletin of Economics and Statistics, 68(4), 397-421.

Delame, E. and Kramarz, F. (1997) "Entreprises et formation continue", Economie et Prevision, Vol. 127, pp. 63-82.

Ichniowski, C., Shaw, K. and Prennushi, G. (1997) "The effects of human resource management practices on productivity", American Economic Review, Vol. 87, pp. 291-313.

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

Lynch, L. (1992) "Private sector training and the earnings of young workers", American Economic Review, 82, 299-312.

Tan, H., Chapman, B., Peterson, C., Booth, A. (1992) "Youth training in the US, Britain and Australia", Research in labour Economics 13, 63-99.

Unioncamere (2007) "Progetto Excelsior: Sistema Informativo per l'Occupazione e la Formazione", http://excelsior.unioncamere.net.

Zwick, T. (2005) "Continuing Vocational Training Forms and Establishment Productivity in Germany" German Economic Review 62(2), 155-184.

Zwick, T. (2006) "The impact of training Intensity on Establishment Productivity, Industrial Relations, 45(1), 26-46.

Table 1: Descriptive Statistics

Variable	Mean	Std. Dev.	Min.	Max.	N.Obs.
Training intensity	0.28	0.32	0	1	30603
Average training duration	8.85	21.39	0	250	24139
Log productivity	10.69	0.70	2.74	14.24	30262
Log capital per worker	9.98	1.53	1.8	15.36	30352
Executives share	0.02	0.05	0	1	33717
Clerks share	0.42	0.32	0	1	33763
Workers share	0.56	0.33	0	1	33742
R&D intensity	0.37	2.31	0	73.66	30446
Patents intensity	0.4	2.65	0	119.73	30446
Firms' age (years)	27.67	16.3	3	154	32703
Exporter (dummy)	0.41	0.49	0	1	33509
Listed (dummy)	0.01	0.09	0	1	33815
Group (dummy)	0.34	0.47	0	1	33815
Small (dummy)	0.42	0.49	0	1	33775
Medium (dummy)	0.40	0.49	0	1	33775
Large (dummy)	0.18	0.38	0	1	33775
Agriculture	0.01	0.07	0	1	33718
Mining	0.02	0.13	0	1	33718
Manufacturing	0.49	0.5	0	1	33718
Construction	0.06	0.23	0	1	33718
Trade	0.11	0.31	0	1	33718
Hotels and Restaurants	0.01	0.12	0	1	33718
Transport and Commerce	0.06	0.23	0	1	33718
Finance, Banking	0.01	0.08	0	1	33718
Business Services	0.14	0.35	0	1	33718
Education, Health	0.08	0.28	0	1	33718
Social Services	0.02	0.14	0	1	33718
Year 2002	0.19	0.39	0	1	33815
Year 2003	0.26	0.44	0	1	33815
Year 2004	0.28	0.45	0	1	33815
Year 2005	0.26	0.44	0	1	33815

Note: Source: AIDA and Excelsior. See section 4 and the data appendix for details on data sources and the definition and construction of individual variables.

Table 2: Effect of training on productivity: OLS, RE, FE

	OLS1	OLS2	Random Effects	Fixed Effects
Training intensity	0.112**	0.045**	0.030**	0.028*
	(9.306)	(3.871)	(3.007)	(2.443)
Capital per worker	0.230**	0.219**	0.285**	0.450**
	(41.433)	(38.455)	(24.841)	(16.632)
Executives share		1.322**	0.596**	0.094
		(10.424)	(5.047)	(0.623)
Workers share		-0.447**	-0.335**	-0.119**
		(-33.555)	(-19.302)	(-4.156)
R&D intensity		-0.006**	0.001	0.007
		(-4.003)	(0.360)	(1.109)
Patents intensity		0.006**	0.006*	0.001
		(4.861)	(2.412)	(0.189)
R^2	0.36	0.40	0.39	0.22
N. observations	26355	26312	26312	26312

Note: Dependent variable: Log-productivity (Value added over Employment).
t-statistics reported in brackets. Standard errors robust to heteroskedasticity and autocorrelation. * p<0.05, ** p<0.01. All regressions include firms' age and a full set of dummy variables for time, industry, regional, size, export, group and listed status.

Table 3: Effect of training on productivity: GMM

	DIF1	DIF2	DIF3	SYS1	SYS2
Training intensity	0.028*	0.044	0.074	0.074**	0.202
	(2.292)	(1.543)	(0.207)	(3.025)	(0.934)
Capital per worker	0.472**	0.513**	0.491**	0.251**	0.123*
	(14.021)	(5.157)	(3.500)	(7.523)	(2.174)
Executives share	0.196	0.272	0.274	1.299**	1.251**
	(0.967)	(0.819)	(0.800)	(6.987)	(6.568)
Workers share	-0.097**	-0.116	-0.119	-0.426**	-0.463**
	(-2.616)	(-1.851)	(-1.733)	(-16.934)	(-10.196)
R&D intensity	0.006	-0.078	-0.069	-0.014**	-0.007
	(0.856)	(-0.730)	(-0.658)	(-3.700)	(-1.911)
Patents intensity	0.003	0.023	0.018	0.009*	0.015**
	(0.462)	(0.509)	(0.461)	(2.204)	(2.774)
Hansen (overall)		0.98	0.94	0.32	0.53
Hansen (levels)				0.29	0.07
N. observations	14035	6666	6666	15306	15306

Note: Dependent variable: Log-productivity (Value added over Employment).
t-statistics reported in brackets. Standard errors robust to heteroskedasticity and autocorrelation. * p<0.05, ** p<0.01. All regressions include firms' age and a full set of dummy variables for time, industry, regional, size, export, group and listed status.

Table 4: Effect of training on wages

	OLS	Random Eff.	Fixed Eff.	System GMM
Training intensity	0.017	0.014	0.020*	0.044*
	(1.790)	(1.561)	(2.039)	(2.129)
Capital per worker	0.127**	0.201**	0.423**	0.229**
	(21.389)	(16.986)	(14.879)	(6.493)
Executives share	1.176**	0.550**	0.010	1.023**
	(9.919)	(4.926)	(0.081)	(6.023)
Workers share	-0.361**	-0.279**	-0.121**	-0.319**
	(-30.662)	(-17.472)	(-4.326)	(-13.057)
R&D intensity	-0.002	0.001	0.001	-0.011**
	(-1.827)	(0.286)	(0.086)	(-4.066)
Patents intensity	0.003**	0.005*	0.001	-0.002
	(3.996)	(2.186)	(0.270)	(2.501)
N. observations	26206.00	26206.00	26206.00	15241

Note: Dependent variable: Log-productivity (Value added over Employment). t-statistics reported in brackets. Standard errors robust to heteroskedasticity and autocorrelation. * p<0.05, ** p<0.01. All regressions include firms' age and a full set of dummy variables for time, industry, regional, size, export, group and listed status.

Table 5: Effect of training on productivity: by occupation

	Skilled vs Unskilled	Occupation Shares
Trained White Collar	0.01	
	(0.33)	
Trained Blue Collar	0.13**	
	(3.92)	
Trained Executives		-0.28
		(-0.94)
Trained Clerks		0.03
		(0.73)
Trained Workers		0.12**
		(3.59)
Skill intensity	0.51**	
	(17.16)	
Executives share		1.39**
		(7.38)
Workers share		-0.45**
		(-14.67)
N. observations	15306.00	15306.00

Note: Dependent variable: labor productivity (value added over Employment). t-statistics reported in brackets. Standard errors robust to heteroskedasticity and autocorrelation. * p<0.05, ** p<0.01. All regressions include firms' age and a full set of dummy variables for time, industry, regional, size, export, group and listed status.

Table 6: Effect of training on productivity: by subsample

Industry	Services	Non-Services	Manufacturing
Training intensity	0.09*	0.05	0.06
	(2.29)	(1.76)	(1.68)
N. observations	6260	9046	7902

Region	North	Center	South
Training intensity	0.08**	0.12	-0.01
	(2.85)	(1.94)	(-0.11)
N. observations	11104.00	2490.00	1712.00

Size	Small	Medium	Large
Training intensity	0.08*	0.04	0.08
	(2.19)	(1.01)	(1.20)
N. observations	5750.00	6682.00	2874.00

Note: Dependent variable: labor productivity (Value added over Employment). t-statistics reported in brackets. Standard errors robust to heteroskedasticity and autocorrelation. * p<0.05, ** p<0.01. All regressions include firms' age and a full set of dummy variables for time, industry, regional, size, export, group and listed status.

Table 7: Effect of training duration: training intensity vs *effective* training

	OLS		FE		SYS	
Training intensity	0.017**		0.010*		0.019*	
	(4.289)		(2.256)		(2.256)	
Effective training		0.009*		0.005		0.022**
		(2.141)		(1.319)		(2.860)
N. observations	21540	19513	21540	19513	15306	14199

Note: Dependent variable: Log-productivity (Value added over Employment). Both training intensity and effective training are standardised. The sample is restricted to 2003-2005. t-statistics reported in brackets. Standard errors robust to heteroskedasticity and autocorrelation. * p<0.05, ** p<0.01. All regressions include firms' age and a full set of dummy variables for time, industry, regional, size, export, group and listed status.

Table 8: Composition of the sample

	Year				
	2002	2003	2004	2005	Total
Size					
1-49	29.8	43.3	46.5	46.3	42.4
50-99	21.6	18.1	17.0	17.9	18.4
100-249	26.7	21.2	20.4	19.7	21.6
250-499	12.6	10.1	9.2	9.1	10.1
500-	9.2	7.3	7.0	7.0	7.5
Industry					
Agriculture	0.5	0.4	0.5	0.5	0.5
Mining	1.9	1.6	1.6	1.4	1.6
Manufacturing	53.5	48.2	46.9	47.4	48.6
Construction	4.5	5.6	5.8	6.1	5.6
Trade	9.3	11.5	11.6	11.4	11.1
Hotels and Restaurants	1.5	1.4	1.4	1.3	1.4
Transport and Commerce	6.0	5.7	5.8	5.9	5.9
Finance, Banking	0.5	0.6	0.7	0.7	0.6
Business Services	12.0	14.4	15.1	14.6	14.2
Education, Health	8.4	8.5	8.4	8.6	8.5
Community, Social	1.9	2.0	2.2	2.2	2.1
Region					
North-West	39.0	36.7	36.3	36.3	36.9
North-East	32.7	33.3	32.9	32.8	33.0
Center	15.8	16.5	16.8	16.7	16.5
South	12.4	13.6	14.0	14.2	13.6

Note: Source: AIDA and Excelsior (see section 4 for details on data construction). The table reports the percentage of firms in the corresponding sub-sample, by year and in total.

Table 9: Sectoral distribution, selected variables

	Value added		Output		Employment	
Sector	Popul.	Sample	Popul.	Sample	Popul.	Sample
Agriculture	2.89	0.22	1.84	0.45	3.23	0.28
Mining	2.54	9.23	3.01	5.19	1.07	2.75
Manufacturing	20.23	47.14	33.76	50.14	27.54	42.00
Construction	6.31	3.33	6.74	3.28	7.60	3.48
Trade	12.95	11.64	13.60	24.06	11.95	13.06
Hotels restaurants	4.05	1.07	3.73	0.71	4.71	2.33
Transport	8.33	12.04	7.97	6.16	6.72	15.55
Banking, insurance	4.98	0.57	4.13	0.40	3.51	0.52
Business services	23.49	10.94	15.36	7.92	11.22	13.27
Education, Health	11.11	1.85	6.73	0.81	18.00	4.00
Social, personal s.	3.12	1.96	3.11	0.89	4.46	2.77

Note: Source: Excelsior, Aida, ISTAT (Italian National Accounts).

Table 10: Structure of sample

Frequency	Percent	Cumulative	Pattern
3441	30.94	30.94	1111
2722	24.47	55.41	.111
895	8.05	63.45	..11
855	7.69	71.14	111.
726	6.53	77.67	1.11
724	6.51	84.18	.11.
440	3.96	88.13	11.1
401	3.61	91.74	11..
356	3.20	94.94	1..1
563	5.06	100.00	(other patterns)
11123	100.00		

Note: Source: AIDA and Excelsior.

Table 11: Training decision, by subsample

	Train.	No Train.		Train.	No Train.
Year			*Industry*		
2002	78.8	21.2	Agriculture	70.0	30.0
2003	70.1	29.9	Mining	79.0	21.0
2004	71.6	28.4	Manufacturing	70.5	29.5
2005	67.4	32.6	Construction	69.5	30.5
Size			Trade	73.5	26.5
1-49	64.0	36.0	Hotels, restaurants	68.8	31.2
50-99	66.5	33.5	Transport, Commerce	69.1	30.9
100-249	71.5	28.5	Finance, Banking	73.9	26.1
250-499	93.7	6.3	Business services	72.2	27.8
500-	97.2	2.8	Education, Health	72.4	27.6
Region			Social Services	73.2	26.8
North-West	73.2	26.8			
North-East	72.4	27.6			
Center	69.8	30.2			
South	66.5	33.5	*Total*	71.5	28.5

Note: Source: AIDA and Excelsior (see section 4 and the data appendix for details on data construction).

Table 12: Training intensity, by subsample

	Mean	Median		Mean	Median
Year			*Industry*		
2002	0.29	0.18	Agriculture	0.24	0.11
2003	0.26	0.14	Mining	0.41	0.33
2004	0.30	0.18	Manufacturing	0.24	0.12
2005	0.28	0.15	Construction	0.28	0.16
Size			Trade	0.29	0.19
1-49	0.30	0.17	Hotels, restaurants	0.27	0.16
50-99	0.21	0.10	Transport, comm.	0.26	0.13
100-249	0.22	0.12	Finance, banking	0.37	0.28
250-499	0.35	0.27	Business services	0.35	0.25
500-	0.41	0.36	Education, health	0.40	0.34
Region			Community social s.	0.32	0.22
North-West	0.27	0.16			
North-East	0.27	0.16			
Center	0.29	0.17			
South	0.32	0.18	*Total*	0.28	0.16

Note: Source: AIDA and Excelsior (see section 4 and the data appendix for details on data construction).