

On-the-Job Training: Estimating Returns Using Firm Level Data*

Rita Almeida
The World Bank

Pedro Carneiro
University College London and Institute for Fiscal Studies

January 2005
Preliminary and Incomplete

Abstract

This paper fills in a major gap in the empirical literature on training by estimating the private return to training using firm level data. The literature has neglected the cost side of this investment and focused solely on quantifying the benefits of training. The coverage and quality of our data allows us to improve the existing literature in several dimensions. We find that, unlike schooling, the foregone output cost of training accounts for 75% to 90% of the total cost. This is not surprising given that in our sample firms train less than 1% of hours worked. Moreover, we estimate that more than 50% of the training takes place outside the working hours and is not associated with lower output. We also quantify the effect of training on current and future output. Even though disentangling correlation and causality bears difficulties, we use the Blundell and Bond (1998) estimator to address this issue. We find that increasing training per employee in 10 hours (approx. 0.5% of total hours worked/year), leads to an increase in current output of 1.5-4%. Changes in future output are smaller since knowledge depreciates over time due to worker turnover. These estimates imply that the median private return to training is higher than 200%. This number is high but it is implicit in the literature, since the estimates of benefits are usually very high. We also find significant heterogeneity in the returns, hourly benefits and costs of training.

Keywords: On-the-Job Training, Panel Data, Productivity

JEL Classification codes: C23, D24, J31

*We thank conference participants at the 16th European Association of Labor Economists (Lisbon, 2004), 19th Meeting of the European Economic Association (Madrid, 2004) and the IZA/SOLE Meetings (Munich, 2004). Corresponding author: ralmeida@worldbank.org. Address: 1818 H Street, NW MC 3-348, Washington, DC, 20433 USA.

1. Introduction

Individuals invest in human capital over the whole life-cycle. It is estimated that over one half of lifetime human capital is accumulated through post-school investments (Heckman, Lochner and Taber, 1998). An important but understudied component of post-schooling investments is on-the-job training. While there exist many public training programs, most of the job training in the economy is done within firms. Like any other investment decision, firms will offer training programs as long as the expected benefits outweigh the total costs of training. A crucial parameter to evaluate such investments and, therefore, the allocation of resources in the economy, is the return to training. Surprisingly, there is almost no research estimating the returns to training (one important exception is Mincer, 1989). Most of the literature focuses on quantifying the benefits of training for the workers or the firms, by estimating the effects of on-the-job training on wages or productivity. This paper fills in a major gap in the literature by estimating the private return to formal on-the-job training.¹

² We use the census of large firms operating in Portugal between 1995 and 1999. The coverage, quality and detail of our data allows us to improve the existing literature in several dimensions. In particular, it allows us to go beyond quantifying the benefits of training, which is only the first step in computing the overall private return to training.

To compute the rate of return of the investment in training we use the standard concept of internal rate of return of an investment. Let B_{jt+s} be the flow of benefits each period and let C_{jt}^T be the total cost of the investment. Assuming that the cost is all incurred in one period and that the investment generates benefits for N periods, the internal rate of return of the investment is given by the rate r that equalizes the present discounted value of net benefits (PDV) to zero:

$$PDV = -C_t^T + \sum_{s=1}^N \frac{B_{t+s}}{(1+r)^s} = 0$$

To estimate the total costs of training, C_t^T , we need information on the direct cost of training and on the foregone productivity cost of training. The first is rarely observed in firm data sets while the second requires the construction of the counterfactual scenario of what

¹We do not estimate the social return to training. To access the optimal amount of training in the economy policymakers require information on the social return to training, rather than the private return. In fact, the primary reason to justify state interventions in the training market is that the social returns exceed the private returns for various reasons (e.g. spillover effects or poaching of workers may cause workers/firms to under-invest in training). A different issue would also be to estimate the returns to the firm and worker separately. For this we would need to disentangle how costs and benefits are distributed between workers and firms. This depends on whether training is general or firm specific. In an extreme world, if training is completely firm specific the firm will support all the costs and receive all the benefits of training. If it is completely general then the worker supports all the costs and receives all the benefits. However, it can be shown that independently of who pays for the cost and who receives the benefits, the costs and benefits of training are the same. What is different is the distribution between workers and firms. Assuming that firms and workers can sign optimal contracts where they first choose the optimal amount of training, and then divide the benefits and costs, the amount of training will be the same independently of the type of training (Coase theorem). If workers are not getting the full benefits of training or if they are not seeing the full costs of training, they may want to over or under train, and the same for firms.

²We will consider only formal training programs. This is a weakness of most of the literature, since informal training is very hard to measure.

would be the firm's output had the workers not taken the training. Our data is ideal for this exercise since it collects information on the duration of training, direct costs of training, training subsidies and hours of training spent on the job. Moreover, it also has information that allow us to estimate the marginal productivity of labor and, therefore, the foregone productivity cost of training.

A major issue in identifying the string of training benefits, B_{t+s} , is to be able to identify all the changes in output that are caused by the investment in training. Disentangling correlation and causality bears difficulties. Several papers have tried to quantify this effect (e.g., Bartel, 1991, Black and Lynch, 1996) but they face two major obstacles. A regression of output on a measure of training, controlling for other inputs in production, is biased if there are unobserved firm characteristics correlated with the regressors and in particular with the training measure. For example, some firms might have an organizational structure that requires more training and is this capital that is boosting productivity. Another obstacle is related with the endogeneity of the training measure in the output equation. Unobserved shocks could raise productivity and induce changes in the training practices of the firm and in other inputs. We take advantage of the panel dimension of our firm sample to minimize these two problems by following an approach similar to the one that Dearden, Reed and Van Reenen (2000) use with industry level data. To address the first point we control for firm time invariant characteristics, unobserved to the econometrician, but that are potentially correlated with the decision to offer training. To address the second point we need an instrumental variable, correlated with training, but not with the transitory productivity shock. We use the Blundell and Bond (1998) system GMM estimator.

Finally, when computing the benefits of the investment in training we recognize that what determines current output is not the current flow but the stock of training.³ This implies that training increases output contemporaneously but also in future periods. However, the changes in output decrease over time as knowledge depreciates. In our empirical work, we compute a firm specific depreciation rate accounting for the fact that knowledge becomes obsolete (or workers simply forget lessons learned) and that each period new workers enter and old workers leave the firm taking with them firm specific knowledge.

Several interesting facts emerge from our work. The benefits estimated by system GMM seem to be very high: a increase in the amount of training per employee of 10 hours (approx. 0.5% of total hours worked/year), leads to an increase in current productivity of 1.5-4%. This estimate is in line with other estimates of the benefits of training in the literature (see Dearden et al, 2000, and Blundell et al, 1996). Using only first differences yields smaller point estimates: an increase in the amount of training by 10 hours per employee leads to an increase in current output of 0.2%. If marginal productivity of labor is constant (linear technology), an increase in the amount of training per employee by 10 hours would translate into foregone productivity costs of 0.5% of output. However, with decreasing marginal productivity of labor, foregone productivity is even lower. In our data the direct costs are a much more important component of the total costs of training (between 75-90%).

The median return in the sample is sensitive to the estimation method used. In our

³Note that most of the literature estimates the contemporaneous effect of training on output. In other words, they estimate B_t , neglecting all the future changes in output, B_{t+s} . Dearden, Reed and Van Reenen (2000) are one exception since they regress output on a measure of the training stock.

preferred specification, using the estimates of benefits generated with GMM methods, it is always above 200%.⁴ This estimate of the returns is high by any standard. However, they are implicit in the whole literature on the returns to private training, since the estimates of benefits are usually very high.⁵ ⁶ With such high returns, it is puzzling why firms train such a small proportion of the total hours of work (less than 1%). This finding suggests that either our estimates of the benefits of training are too large, or our estimates of costs are too small. The estimates of the effects of training on productivity can be too large if higher productivity firms also provide more training and if we do not account for this problem in an adequate way. Estimates of costs may be too small if our estimates of the marginal productivity of labor are downward biased. This would be the case if low productivity firms are labor intensive and have a lower marginal productivity of labor. We believe that the former is a more important source of bias than the latter, although we have no formal way to test this claim. However, estimates of the marginal product of labor are more robust to changes in the specification than estimates of the marginal product of training. We also have unusually good data on the direct costs of training in our data set which are a much more important component of the total cost of training than foregone productivity costs. Even if we are underestimating the marginal productivity of labor, foregone costs are likely to be a small part of total costs as the proportion of training hours that takes place outside the working period is above 50%. We also find significant heterogeneity in the returns to training. Moreover, this heterogeneity comes not only from the hourly benefits of training but also from the hourly costs of training.

The paper proceeds as follows. In section 2, we estimate the production function of the firm and compute the benefits of training. Section 3 estimates the foregone productivity cost of training and the total cost of training. Section 4 describes the data used. Section 5 computes the rates of return for the investment in training. Section 6 discusses the sensitivity of our results to different assumptions about production and cost functions. We conclude in section 7.

2. Training and Productivity

2.1. Estimating Production Functions

The most common approach used to estimate the effect of training on productivity is to specify a production function which includes training as another input (see e.g. Bartel,

⁴This rate is the return of the total investment. A different concept is that of marginal return to an extra unit of training provided. If there are decreasing returns to training in the production function, it is possible that this return is high but the marginal return is low. This is ongoing work. Some results in appendix D.

⁵To cross check our numbers we compute the implicit rates of return to job training from selected papers in the literature that estimate the effect of training on productivity. Because these studies do not report direct costs of training we assume that direct costs are on average the same proportion of the total costs than in our data. Only considering the foregone productivity cost of training and assuming that all the training takes place during the working time, many of the implicit rates of return are implausibly high. Results reported in appendix.

⁶Mincer (1989), the only other paper we know that performs a similar calculation of returns, finds estimates around 17-25%. He computes the return using wages, so this number is a lower bound to the private returns.

1991).⁷ One of the problems in this literature is that the functional form and the outcome variable are often determined by data availability and, therefore, estimates are not directly comparable across papers. Moreover, the measures of training used also differ across papers (hours of training, number of trainees, indicators for the presence of training), and so do the outcome variables for the firm (sales, output, value added, value added per worker). Most of the literature uses the current or the one-year before level of training to measure the stock of human capital in the firm.⁸ The latter procedure is correct only if we assume that skills depreciate fully in one period. It is usually assumed that the production function is either log-linear in training (e.g. Barron et al., 1989, and Black and Lynch, 1998), or semi log-linear (e.g. Bartel, 1991, 1994, Dearden et al., 2000, and Ramirez, 1994),

Table 1 presents a summary of the variables used to measure productivity, training, the specification and data set used in different papers.⁹ The diversity of these measures creates a difficulty in comparing and interpreting the estimates across samples. However, across studies from the U.S. and Europe there is a robust positive correlation between a measure of formal on-the-job training and firm level productivity.

One of the shortcomings of this literature was the use of current or lagged training as a measure of the stock of human capital. We relax this assumption and derive the reduced form equation that will be estimated to quantify the effect of training on productivity. We assume that the production function is semi-log linear and that firm's stock of human capital determines the level of current output:

$$Y_{jt} = A_t K_{jt}^\alpha l_{jt}^{\beta_1} N_{jt}^{\beta_2} \exp(\gamma H_{jt} + \theta Z_{jt} + \varepsilon_{jt}) \quad (2.1)$$

⁷An alternative approach to measure the effect of training on productivity is to use wages (see Bartel, 1995, and Arulampalam et al., 1997). However, changes in the average wages within a firm may be a poor indicator of changes in the firm's productivity. For example, if training is firm specific then wages may not increase after training. See Acemoglu and Pishke (1998, 1999) for a model where institutions that foster wage compression provide an incentive for firms to invest in general training. Wage growth is likely to be a lower bound to changes in productivity that is likely to be overestimated as more able workers are more likely to have higher wage growth and are also more likely to receive training. Notice however that the lower bound is not in growth rates. For example, Dearden et al (2001) find that wages increase percentually by less than output but wages might be a lower bound even if this is not the case as they are measured in different units.

⁸Some papers propose regressing output growth on the period flow of training to overcome the difficulties associated with measuring the stock of human capital (see e.g. Barret and O'Connell, 1999).

⁹Bartel (1991) finds evidence for the US manufacturing firms that if a firm increases the proportion of occupations with training by 1 percentage point then current productivity increases by 0.17%. Bartel (1994) shows that firms that provided training to particular occupations, and did not do so before have higher productivity growth than other firms. And vice-versa, i.e, firms that gave some training and stopped doing so, registered decreases in productivity growth. The magnitude of the qualitative effect of training depends on the occupation group: it varies from 17% in managers to 60% in clerical occupations. Black and Lynch (1998) using data for the US, do not find evidence for a statistically significant effect of the number of trainees on current or future productivity level. They find that an increase in the proportion of total hours of training off the job by 1 percentage point increases productivity by 0.002%. Evidence for European countries, as that in Alba-Ramirez (1994) for Spain and Barrett and O'Connell (1999) for Ireland, has also pointed to a positive correlation between training and firm productivity. Dearden, Reed and Van Reenen (2000) find also evidence for the UK, that sectors in manufacturing that offer training receive subsequent increases in output. Their findings suggest that a 5 percentage point increase in industry training incidence leads to an increase in industry labor productivity by 4 per cent and a 1.6 percent increase in hourly wages.

where Y_{jt} is a measure of output in firm j at period t , K_{jt} is a measure of capital stock, L_{jt} is a measure of the labor input, H_{jt} is the aggregate stock of human capital in the firm, l_{jt} is the hours worked per employee, N_{jt} is the total number of employees in the firm and Z_{jt} is a vector of firm and workforce characteristics. Given that the production function is assumed to be identical for all the firms in the sample, ε_{jt} captures firm heterogeneity.

We need to obtain estimates for two components: $\frac{\partial Y_{jt}}{\partial H_{jt}}$ and $\frac{\partial H_{jt}}{\partial T_{jt-s}}$ for $s = 1, \dots, N$. To compute $\frac{\partial H_{jt}}{\partial T_{jt-s}}$ we need to make further assumptions on the functional form of the human capital production function. We assume that the total human capital in the firm depreciates for two reasons. On the one hand, skills acquired in the past become less valuable as knowledge becomes obsolete and workers forget past learning. This depreciation affects all the workers in the firm. We assume that one unit of knowledge at the beginning of the period depreciates at rate δ per period. On the other hand, skills depreciate because each period the worker turnover within the firm implies that, all else constant, new workers need to acquire skills and those who leave take with them firm specific knowledge. Therefore, we propose a human capital production function of the following form (abstracting from j):

$$H_{jt+1} = ((1 - \delta)h_{jt} + t_{jt})(N_{jt} - E_{jt}) + X_{jt}t_{jt} + \omega'_{jt}$$

where h_{jt} is the per capita human capital in period t ($h_{jt} = \frac{H_{jt}}{N_{jt}}$), X_{jt} is the number of new employees in period t and E_{jt} the employees who leave the firm in period t .¹⁰ ω'_{jt} is a firm time varying productivity shock. At the end of period t , the stock of human capital is given by the human capital of those $N_{jt} - E_{jt}$ workers that were in the firm in the beginning of the period (these workers have a stock of human capital and receive some training on top of that) plus the training of the X_{jt} new workers. This specification implies that the stock of human capital per employee is given by:

$$h_{jt+1} = (1 - \delta)h_{jt}\phi_{jt} + t_{jt} + \omega_{jt}$$

where $\phi_{jt} = \frac{N_{jt} - E_{jt}}{N_{jt+1}}$ and $0 \leq \phi_{jt} \leq 1$. Comparing this equation with the permanent inventory formula it is straightforward to see that the total skill depreciation in the model is given by $(1 - \delta)\phi_{jt}$. Assume each firm starts out with an initial stock of human capital given by H_{j0} . The human capital of the firm in any future period can be written as:

$$H_{jt} = (1 - \delta)^t \phi_{j1} \dots \phi_{jt-1} H_{j0} + \sum_{s=1}^{t-1} (1 - \delta)^{s-1} \phi_{jt-s} \dots \phi_{jt-1} t_{jt-s} + \sum_{s=1}^{t-1} (1 - \delta)^{s-1} \phi_{jt-s} \dots \phi_{jt-1} \omega_{jt-s} \quad (2.2)$$

H_{j0} is the firm's human capital the first period the firm is observed in the sample and it is unobservable in our data. Plugging (2.2) into the log-linearized version of (2.1) for period t gives:

¹⁰We assume that all entries and exits occur at the beginning of the period. We also ignore the fact that workers who leave may be of different vintage than those who stay. Instead we assume that they are a random sample of the existing workers in the firm (who on average have h_t units of human capital).

$$\ln Y_{jt} = \ln A_t + \alpha \ln K_{jt} + \beta_1 \ln l_{jt} + \beta_2 \ln N_{jt} + \gamma \sum_{s=1}^{t-1} (1-\delta)^{s-1} \phi_{jt-s} \dots \phi_{jt-1} t_{jt-s} + \theta Z_{jt} + \mu_{jt} + v_{jt} \quad (2.3)$$

where $\mu_{jt} = \gamma(1-\delta)^t \phi_{j1} \dots \phi_{jt-1} H_{j0}$, and $v_{jt} = \varepsilon_{jt} + \gamma \sum_{s=1}^{t-1} (1-\delta)^{s-1} \phi_{jt-s} \dots \phi_{jt-1} \omega_{jt-s}$. However, μ_{jt} becomes a firm fixed effect if skills fully depreciate ($\delta = 1$ or $\phi_{jt} = 0$ for all t) or if there is no depreciation ($\delta = 0$) and turnover is constant ($\phi_{jt} = \phi_j$). If δ and turnover are positive and time varying but smaller than 100%, then μ_{jt} is depreciates every period at rate $(1-\delta)\phi_{jt}$.¹¹ Since ϕ_{jt} is observed, assuming a value for δ we can estimate γH_{j0} . This corresponds estimating a model with firm dummies in every period with the restriction that the coefficients on these dummy variables fall at rate $(1-\delta)\phi_t$. For computational reasons, we proxy this effect by a time invariant firm effect.¹²

Estimating equation (2.3) by least squares has some problems. There is a possible misspecification due to the functional form chosen, even though there is no obvious solution for this (except being non-parametric which is not feasible with so many explanatory variables). The production function may also be misspecified in the sense that variables that determine the level of output are omitted and uncorrelated with other observable inputs but correlated with training. Another obstacle with using this approach is that the human capital measure, and therefore training, may be correlated with unobserved productivity shocks. Even though the inclusion of firm time invariant effects mitigates this problem (see Griliches and Mairesse, 1995), this assumption would be violated if, for example, transitory productivity shocks determine the decision of offering training. Moreover, the sign of the bias is indeterminate as it depends on the sign of the correlation between training and the productivity shock.¹³ Finding an exogenous variable correlated with training and uncorrelated to unobserved firm productivity would help identifying the direction of causality but, in practice, this is a very difficult task due to lack of good instrumental variables. We take advantage of the panel dimension of our firm sample to minimize these two problems by following an approach similar to the one that Dearden, Reed and Van Reenen (2000) use with industry level data. To address the first obstacle we control for firm time invariant characteristics, unobserved to the econometrician, but that are potentially correlated with the decision to offer training. To address the second obstacle we need an instrumental variable that is correlated with training but not with the transitory productivity shock. We use the Blundell and Bond (1998) GMM estimator to address this issue. This estimator exploits information in levels and in differences. It's validity is based on the assumption that the productivity shocks are an AR(1) process. An advantage of this approach is that it corrects for biases generated by measurement error in the variables in the model.

¹¹Just notice that $\frac{\mu_{t+1}}{\mu_t} = (1-\delta)\phi_t$.

¹²We have an unbalanced panel of approximately 5,500 observations of 1,500 firms. Using this approach would imply imposing approximately 4,000 restrictions on the regression.

¹³If when productivity is high firms allocate more resources to training, $\hat{\gamma}$ is overestimated. On the other hand, it might be when demand is low and labor is idle that firms invest more in training programs (because the opportunity cost of labor is lower), and $\hat{\gamma}$ will be underestimated (Dearden et al, 2000).

2.2. Computing the Benefits of Training

Empirically Y_{jt} is the firm's value added, K_{jt} is the book value of capital, l_{jt} is hours worked per employee, N_{jt} the total number of employees, Z_{jt} includes time varying firm and workforce characteristics as the size of the firm, the proportion of males in the workforce, the age distribution of the workforce, the distribution of the tenure on the job of the workforce, occupational distribution of the workforce, education of the workforce (measured by the proportion workers with low, medium and high education). t_{jt} is measured with the hours of training per employee.

We assume that depreciation has these two components (δ and ϕ). We assume that $\delta = 17\%$ per period¹⁴ and we estimate the turnover rate from the data. The average turnover in the sample is 86%.¹⁵ We observe the initial and the end of the period workforce as well as the number of workers who leave the firm. To compute the number of new hires we assume that $N_{t+1} = N_t + X_t - E_t$.¹⁶ Whenever the initial number of employees in the firm is unobserved we assumed that $\phi = 0$. Also, because the panel is unbalanced, we assumed that when the firm is unobserved between any two periods, the training provided is an average of the lead and lagged training values¹⁷. Average skill depreciation in our sample is 25% per period. We perform some sensitivity analysis of the rates of return with respect to the skill depreciation.

To obtain estimates of the return of the training investment we must simulate the effect of training on output in all future periods. The period $t + s$ gain of training the workforce in period t is given by the difference in output with and without training: $B_{jt+s} = Y_{t+s(T_t=T_t^*)} - Y_{t+s(T_t=0)}$, where T_t^* is the training that firm j effectively provided at period t . Let $AT_{jt} = \sum_{s=1}^{t-1} (1-\delta)^{s-1} \phi_{jt-s} \dots \phi_{jt-1} t_{jt-s}$. Then, the expressions for $Y_{jt+s(T_{jt}=T_{jt}^*)}$ and $Y_{jt+s(T_{jt}=0)}$ are given by: $Y_{jt+s(T_{jt}=T_{jt}^*)} = \exp [\ln Y_{jt} - \hat{\gamma} AT_{jt} + \hat{\gamma} AT_{jt} (1-\delta)^s \phi_j^s + \hat{\gamma} T_{jt}^* (1-\delta)^{s-1} \phi_j^{s-1}]$ and $Y_{jt+s(T_{jt}=0)} = \exp [\ln Y_{jt} - \hat{\gamma} AT_{jt} + \hat{\gamma} AT_{jt} (1-\delta)^s \phi_j^s]$ where ϕ_j is the average value for the ϕ_j 's that are observed in the sample. When skills do not fully depreciate, the training given today affects productivity in future periods. This effect can be computed for all future periods as long as we assume that the rate of skill depreciation is constant.¹⁸

3. Costs of Training

As argued in Becker (1962), training involves a direct cost (e.g., the cost of renting the equipment in training) and a value placed on the time and effort of trainees (foregone productivity). The latter are costs in the sense that they could have been used in producing

¹⁴Lillard and Tan (1986) estimate that depreciation is between 15% and 20% per year.

¹⁵Although there is some sector variation in the turnover rates, all sectors report value above 80% per period.

¹⁶There were only 35 firms in this case out of 5478 observations.

¹⁷This assumption is likely to have minor implications in the construction of the human capital variables because there were few of these cases (approx. 3% of the final sample). Moreover, most of the firms are unobserved for only one period.

¹⁸We assume that skill depreciation is constant and equal to the average depreciation that we observe in the sample (i.e. the average $(1-\delta)\phi_j$).

current output if they were not used in raising future output. Let the total training cost is given by $C_{jt}^T = C_{jt} + FP_{jt}$.

The data used in this paper allows us to compute estimates of the total training costs. On the one hand, there is information on total direct costs of training in a given year, C_{jt} . These costs include labor payments to trainers or training institutions, training equipment as books, or movies and cost related to the depreciation of training equipment (including buildings and machinery). It also includes grants the firm might receive from a third party. Since we focus on the private returns to training, we exclude from the direct costs any subsidies received by firms from a third party.

Because part of the on-the-job training takes place during the working hours, there is an output loss to the firm that is part of the total cost of training. This cost is given by the difference between potential output (maximum output had the firm devoted both the working and training hours to the production of output) and the output effectively produced at period t : $FP_{jt} = Y_t^* - Y_t$. Potential output is obtained simulating that the total hours of work are those that were effectively devoted to work plus the hours of training that were not worked due to training, i.e., $Y_{jt}^* = Y_{jt} + \beta [\ln(l_{jt} + l_{jt}^T) - \ln l_{jt}]$ where l_{jt} are the hours effectively worked and l_{jt}^T are the hours not worked due to training. Alternatively the foregone output of each training hour is given by: $FP_{jt} = \frac{\partial Y_{jt}}{\partial L_{jt}} \frac{\partial L_{jt}}{\partial T_{jt}}$, where $\frac{\partial Y_{jt}}{\partial L_{jt}}$ is simply the marginal productivity of labor, $\hat{\beta}$, that is estimated from equation (2.3) and $\frac{\partial L_{jt}}{\partial T_{jt}}$ is the % of training hours that take place during the working period.

4. The Data

The main source of data is the “Balanco Social”, an annual survey administered by the Portuguese Ministry of Employment covering every firm with more than one hundred employees operating in Portugal. It is based on a mandatory survey and it has information on different training measures, labor productivity, worker turnover, total wage bill and direct training costs at the firm level for the period 1995-1999. Details of this survey are given in appendix. This data set has several advantages relatively to the data used in other studies. First, it contains employer reports of training activities within the firm. Having information reported by the employer is better than asking the employee about past training if the employee recalls less information about on-the-job training.¹⁹ Second, training variables are reported for all employees in the firm, not just new hires. Third, the survey is mandatory for all the large firms in the country, representing 34% of the total workforce in 1995. Two problems with the empirical literature on this topic have been the small sample sizes and the low quality of the available training data. Response rates on surveys can be very low when surveys are not mandatory.²⁰ Fourth, it has a longitudinal dimension with time consistent

¹⁹However, firm level data is likely to be a greater advantage over worker level data for informal training. In this type of training the worker might consider as working time, the period in which he is actually being trained.

²⁰Bartel (1989) uses a survey conducted by the Columbia Business School with a 6% response rate. Lynch and Black (1997) use data on the Educational Quality of the Worforce National Employers survey, which is a telephone conducted survey with a 64% ”complete” response rate. Barrett and O’Connell (1999) expand an EU survey and obtain a 33% response rate.

information on training measures, total productivity, total wage bill and direct training costs at the firm level.²¹ More than 50% of the firms are observed at least twice during the period 1995-1999.²²

The structure of the final sample is given in Table A1 in the appendix. Table A2 in the appendix reports the descriptive statistics for the relevant variables in the analysis. We divide the sample according to whether the firm provides any formal training and, if it does, whether the yearly total training hours are above the median (1,489 hours) for the firms that provide training. Firms that offer training programs have a higher value added per employee and are larger than low training firms and firms that do not offer training. Total hours on the job per employee, either working or training do not differ significantly across types of firms. High training firms also invest more in physical capital measured by the book value depreciation of capital and have higher yearly labor costs per employee (approximately twice as large as labor costs per employee for firms that do not offer training).²³ In general, the workforce in firms that provide training is more educated and is older than the workforce in firms that do not offer training (the proportion of workers with bachelor or college degrees is 8% and 5% in high and low training firms versus 2% in non-training firms). The workforce in firms that offer training has a higher proportion of male workers and a higher average tenure.²⁴ These firms also tend to have a higher proportion of more skilled occupations such as higher managers and middle managers, as well as a lower proportion of apprentices. High and low training firms differ significantly in their training intensity. Firms with a small amount of training (defined as being below the median) offer on average 2.6 hours of training per employee per year while those that offer a large amount of training offer on average 33 hours of training.²⁵ Even though the difference between the two groups is large, it is surprising to find such small amounts of training overall (average yearly hours on the job 1,837 hours), even for the high training firms. Portuguese firms with more than 100 employees that offer more training than the median, train at most 1.8% of total time on-the-job, which is a rather small number. High training firms spend on average almost 5 times more in direct costs of training per employee than low training firms (41 euros per year and per employee for a low training firm versus 223 euros for a high training firm). As a proportion of value added these costs are 0.3% and 1.4% respectively. This proportion is rather small, but is in line with training such a small number of hours. In sum, the Portuguese data is in line with surveys from other countries in Europe and from the U.S.,

²¹Dearden, Reed and Van Reenen (2000) cover the period 1983-1996 but match the UK Labor Force Survey with industry level data, which might generate aggregation biases of unknown sign.

²²The major reasons why firms leave the sample are related with firms exiting the market, reduction in employment below 100 employees, which implies that answering to this survey is not mandatory, and to changes in the address of the firm headquarter. This implies a change in the firm identifier so that the firm would still be in the sample but under a different firm identifier.

²³This labor costs excludes the labor costs related to training programs.

²⁴Arulampalam, Booth and Bryan (2003) also find evidence for European countries that training incidence is higher among men, and is positively associated with high educational attainment and a high position in the wage distribution.

²⁵In our empirical work we use as measure of training the number of hours of training per employee. This statistic is more informative than the total number of training hours because it controls for the fact that larger firms have a higher number of trainees. Other useful statistics would be the number of trainees and hours of training per trainee but the number of trainees is imperfectly measured in this data set.

with respect to the profile of the firms that train. There is a lot of heterogeneity among the firms that offer training, with the low training firms and the high training firms being very different. Finally, the Portuguese firms train a very small proportion of time and therefore also spend a small proportion of their value added with formal training programs.

5. The Returns to On-the-Job Training

To compute the rate of return of the investment in training we use the standard concept of internal rate of return of an investment. Let B_{jt+s} be the flow of benefits each period and let C_{jt}^T be the total cost of the investment. Assuming that the cost is all incurred in one period and that the investment generates benefits for N periods, the internal rate of return of the investment is given by the rate r that equalizes the present discounted value of net benefits (PDV) to zero:

$$PV = -C_t^T + \sum_{s=1}^N \frac{B_{t+s}}{(1+r)^s} = 0$$

Panel *A* of Table 2 presents the results of estimating the production function in equation (2.3) for the effect of the human capital measure and for hours worked on output. The different columns present the results under different assumptions about the estimation method and the rate of skill depreciation. Columns (1) assume a depreciation of 100%, columns (2), (4) and (5) assume that depreciation is a function of the firm's turnover and column (3) assumes that depreciation is 17% per year. Columns (1) to (3) estimate equation (2.3) with least squares in first differences while columns (4) to (5) estimate it with system GMM (Blundell and Bond).

The point estimates of the effect of training on output are very sensitive to assumptions on depreciation rates, autocorrelation of the errors and method used. Very small changes in translate in large changes in the estimates of total benefits. The point estimate in columns (1) to (3), using only first differences, implies that increasing the current training by one hour for all the employees in the firm increases output next period by 0.0002% to 0.0003%. This implies that an increase in the amount of training by 10 hours per employee leads to an increase in current output of approx. 0.2%. The benefits estimated by SYS-GMM, in columns (4) and (5) seem to be very high: a increase in the amount of training per employee of 10 hours (approx. 0.5% of total hours worked/year), leads to an increase in current productivity of 1.5-4%. This estimate is in line with other estimates of the benefits of training in the literature (see Dearden et al, 2000, and Blundell et al, 1996). If marginal productivity of labor is constant (linear technology), an increase in the amount of training per employee by 10 hours would translate into foregone productivity costs of 0.5% of output. With decreasing marginal productivity of labor, foregone productivity is even lower.

Another contribution of our paper is to quantify the importance of foregone costs of training. In the case of education, foregone earnings are a much more important component of total costs than tuition and other direct costs (full time students forego 100% of labor earnings to invest in schooling). Not much is known for the on-the-job training. Table 3 describes the costs of training in our sample. Not only there is heterogeneity in training incidence (see table *A1* in the appendix), but those firms that offer training programs spend

a median of 5 hours per employee training. This number is very small when compared to the total hours of work (0.3%). Moreover, approximately 60% of the hours do not imply that the firm is foregoing production because training is done outside the working hours. Therefore, the foregone productivity costs are likely to be small relatively to the direct costs of training.

Panel *B* of Table 3, presents, for each specification, the median estimates for the benefits of training all employees for one hour (one period after the training is offered), $\frac{B_{jt+1}}{h_{jt}}$, for the direct costs of training all employees for one hour, $\frac{C_{jt}}{h_{jt}}$, and for the foregone output cost of training all employees for one hour, $\frac{FP_{jt}}{h_{jt}}$. Reporting average values of these variables per hour trained is more informative than reporting the total training benefits/costs because these benefits/costs increase with the amount of training provided. For example, the benefits assuming 100% depreciation in column (1) implies that for the median firm, the output is 1116 euros higher, one period after the training, for each training hour given to all employees than what it would have been had the firm offered no training. All the values are in prices of 1995 and denominated in euros.

In column (1), assuming that skills fully depreciate during one period the median return to training is negative (-48%). This happens because average benefits are small (1116 euros) and occur only for one period, while the direct cost of training is more than twice the average benefits (2,436 euros). From these numbers it is clear that the assumption behind several models in the literature of full skills depreciation is very unrealistic. With our preferred specification of the depreciation rate (firm specific) and using the estimates of benefits generated from simple first difference methods, columns (2), our estimate of median rate of return increases to 10%.

In our data the direct costs are a much more important component of the total costs of training (approx. 75-90% of all costs). However, the direct costs are not high enough to offset the high estimates of the benefits of training that we obtain using SYS-GMM. The median rate of return (IRR) in columns (4) and (5) is always above 200% and therefore are high by any standard. However, they are implicit in the whole literature on the returns to private training, since the estimates of benefits are usually very high (and costs are never estimated). Even in Mincer (1989), the only other paper we know that performs a similar IRR calculation as we do, finds estimates around 17-25%. Given that Mincer (1989) uses wages his number is a lower bound to the returns using productivity. In most of the literature (including in our paper) the benefits of training seem to be high, implying IRR above 250%.

With such high returns, it is puzzling why firms train such a small proportion of the total hours of work (less than 1%). This finding suggests that either our estimates of the benefits of training are too large, or our estimates of costs are too small. The estimates of the effects of training on productivity can be too large if higher productivity firms also provide more training and if we do not account for this problem in an adequate way. Estimates of costs may be too small if our estimates of the marginal productivity of labor are downward biased. This would be the case if low productivity firms are labor intensive and have a lower marginal productivity of labor. We believe that the former is a more important source of bias than the latter, although we have no formal way to test this claim. However, estimates of the marginal product of labor are more robust to changes in the specification than estimates of the marginal product of training.

We also find significant heterogeneity in the returns to training. Figure 1 plots the distribution of the returns of training all the employees during one hour excluding the observations larger than percentile 99 and using the specification of column (2). The distribution is very heterogeneous: percentile 25 of the returns to training is -15% , the median is 10% while percentile 75 is 57% . This heterogeneity is a result not only of the heterogeneity existent in the benefits but also in the costs of training. There is a positive correlation between the first year average benefits of training and the average costs of training even though there is no systematic relation between any of these and the returns. In future work we look at the distribution of returns emphasizing differences across sectors, firm size and ownership nationality of the firm.

6. Sensitivity Analysis

6.1. Benefits

One can think about several reasons why the benefits of on-the-job training might be overstated. On the one hand, the benefits of training might be overstated if formal and informal training are positively correlated and both have positive effect on output. Unfortunately, we do not have any data on informal training and we have no way of testing this statement. On the other hand, if positive transitory shocks affect the decision to offer training programs the estimates of the effect of training on productivity are overstated even controlling for firm fixed effects. One solution to this problem could be finding an instrument for training. Unfortunately, we do not have an exogenous instrument available and using lagged training values as instrument is not possible because the decision of offer training is persistent.

In future work we plan to experiment with different specifications of the production function and examine how our estimates of costs, benefits and returns change.

6.2. Costs

The returns to training can be high if the training costs are underestimated. We estimate the foregone output cost of training by valuing the time spent training, that would be working time had the training program not been offered, with the marginal productivity of labor. However, if high productivity firms employ less labor there is a downward bias in the estimation of the marginal productivity of labor because labor is correlated with the residual in the productivity equation.²⁶ Assuming that labor markets are competitive, hourly wages equal the marginal productivity of an hour of work and the labor costs supported by the firm can be used to estimate the foregone output associated with training.²⁷ However, in the real world there are several situations where the cost of labor to the firm and the worker's

²⁶The models we estimate assume we have a fixed (non-random) coefficient and therefore there is only possibility of selection on levels (the residual of the production function). In this context it is natural to assume that labor intensive sectors, or at least sectors that employ a lot of workers, are of low productivity (this is factor neutral productivity). However, if we allow the coefficients of the production function to be random as well, then it is likely that firms that use an input very intensively have a higher productivity for that input. We ignore this possibility in our analysis and leave it for future work.

²⁷In this case the wage equals to the worker's outside option in the market if the worker is equally productive everywhere. Firms will hire up to the point where wages equal the marginal product of the worker.

marginal product differ. If there are unions that set up wages above the marginal product of labor or if firms train workers when labor is idle (see Dearden et al., 2000), i.e., when marginal productivity is low, then wages are higher than the marginal product of labor. Also, if a firm faces a negative demand shock and if there are hiring and firing costs, it may be optimal for the firm not to fire its employees and therefore to have too much labor, and a low marginal product of labor, so that wages rise above the marginal product. By a similar argument, in times of high demand, it may be optimal for the firm not to hire as much as in the absence of hiring and firing costs and, therefore, to have too little labor, so that wages are lower than the marginal product of labor. Wages may also be below the marginal product of the worker if there is firm specific human capital that raises the marginal product above the worker's market wage. In this case, the firm pays the worker's outside option, which is the market wage, and might be below the marginal product of labor. In this section we estimate the foregone output costs of on-the-job training using data for wages but it is unclear whether these are over or underestimates of the true foregone costs of training.

The data set used in this paper has information on total labor costs for the firm in a given period.²⁸ In our sample, the average cost of labor per employee is 10,000 euros per year while the average hourly cost per employee is 5 euros per hour. This implies that the labor cost for one hour of work for all employees in a firm is on average 1,700 euros. Combining this cost with the proportion of training time that is done during working hours generates a foregone cost of labor that is 616 euros for the median firm. This is approximately five times the cost obtained using the marginal productivity of labor from the production function approach. This can happen for one of the reasons explained above. It may also be driven by aggregation of wages and productivity of very heterogenous workers.²⁹ Results for the rates of return are work in progress. It is also work in progress the estimation of a cost function that allow us to obtain estimates for the marginal cost of training.

7. Conclusion

This paper fills in a major gap in the empirical literature on training by estimating the private return to training using firm level data. The literature has neglected the cost side of this investment and focused on quantifying the benefits of training. The coverage and quality

²⁸Labor costs include wages, regular subsidies (e.g., food subsidies), irregular subsidies (e.g. holidays), fringe benefits, compulsory legal labor costs (e.g. social security payments), training costs and other costs (e.g., garment and transport costs).

²⁹Suppose a firm has workers of different productivities because they have different skills (or work in different occupations) and therefore they earn different wages. Then the average wage in a given firm is $w = \sum \frac{H_k}{H} w_k$, where k is an index of the type of worker, w_k is the hourly wage of workers of type k , H_k is hours of work of workers of type k and H is total hours of work ($H = \sum_{k=1}^K H_k$). The true production function is $Y = f(H_1, H_2, \dots, H_k)$ but we estimate $Y^* = f(H) = f(\sum_{k=1}^K H_k)$ instead. Let $k = 1$ be the low productivity workers. If most of the variation in hours across time and across firms is given by variation in hours of low productivity workers (H_1) then $\frac{\partial Y^*}{\partial H} \simeq \frac{\partial Y}{\partial H_1}$, the marginal productivity of labor for low productivity workers. However, $\frac{\partial Y}{\partial H_1} = w_1 < w$. In future work we use information for hours of work and hours of training, and wages for different occupational groups.

of the firm level survey allows us to improve the existing literature in several dimensions. We find that, unlike schooling, the foregone output cost of training accounts for 75% to 90% of the total cost. This is not surprising given that in our sample firms train less than 1% of hours worked. Moreover, we estimate that more than 50% of the training takes place outside the working hours. We also quantify the effect of training on current and future output. Even though disentangling correlation and causality is difficult, we use the Blundell and Bond (1998) estimator to address this issue. We find that increasing training per employee in 10 hours (approx. 0.5% of total hours worked/year), leads to an increase in current output of 1.5-4% and to smaller increases in future output since knowledge depreciates over time. These estimates imply that the private return to training is higher than 200%. This number is high but it is implicit in the literature, since the estimates of benefits are usually very high. We also find significant heterogeneity in the returns, hourly benefits and costs of training.

One important extension of our work is to consider different types of heterogeneity. We want to incorporate heterogeneity in different types of workers, and consequently, heterogeneity in different types of training. We also want to conduct a much better analysis of heterogeneity in the returns to training across firms. This would allow us to analyze the relationship between training patterns and returns to training across different firms. We would also like to explore the link between training, productivity and wages. In particular, we would like to examine how the costs and benefits of training are shared between workers and firms, and how does this division conform to standard economic theories about on-the-job training.

References

- [1] Arulampalam, W., Booth A., and Bryan, M.. 2003. "Training in Europe". University of Essex working paper.
- [2] Arulampalam, W., Booth A., and Elias, P..1997. "Work-related Training and Earnings Growth for Young Men in Britain" *Research in Labor Economics*,16, 119-147, Fall.
- [3] Alba-Ramirez, A..1994. "Formal Training, Temporary Contracts, Productivity and Wages in Spain". *Oxford Bulletin of Economics and Statistics*, vol 56.
- [4] Barron, J., Black, D. and Lowenstein, M..1989. " Job Matching and On-The-Job Training" *Journal of Labor Economics*. vol 7.
- [5] Becker, G. 1962. "Investment in Human Capital: A Theoretical Analysis". *The Journal of Political Economy*, vol. 70, No. 5, Part 2: Investment in Human Beings, pp.9-49.
- [6] Black, S. and Lynch, M. 1996. "Human-Capital Investments and Productivity (in Technology, Human Capital, and the Wage Structure)". *American Economic Review Papers and Proceedings*, vol. 86, No. 2, pp. 263-267.
- [7] - -. 1997. "How to Compete: The Impact of Workplace Practices and Information Technology on Productivity " *National Bureau Economic Research Working Paper No. 6120*.
- [8] - -. 1998. "Beyond the Incidence of Training: Evidence from a National Employers Survey ". *Industrial and Labor Relations Review*, Vol.52, no.1.
- [9] Bartel, A..1991. "Formal Employee Training Programs and Their Impact on Labor Productivity: Evidence from a Human Resources Survey ". *Market Failure in Training? New Economic Analysis and Evidence on Training of Adult Employees*, ed. David Stern and Jozef Ritzen, Springer-Verlag.
- [10] - -.1994. "Productivity Gains From the Implementation of Employee Training Programs". *Industrial Relations*, vol. 33, no. 4.
- [11] - -. 1995. "Training, Wage Growth, and Job Performance: Evidence from a Company Database". *Journal of Labor Economics*, Vol. 13, No. 3., pp. 401-425.
- [12] Barrett, A. and O'Connell, P.. 2001. " Does Training Generally Work? The Returns to In-Company Training". *Industrial and Labor Relations Review*, 54 (3), 647-662.
- [13] Blundell, R. and Stephen, B. (1998). "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics* 87: 115-143.
- [14] Blundell, R., Dearden, L., Goodman, A. and Reed, H. (1997), *Higher Education, Employment and Earnings in Britain*, London: Institute of Fiscal Studies.
- [15] Blundell, R., Dearden, L. and Meghir, C. (1996), *The Determinants of Work Related Training in Britain*, London: Institute of Fiscal Studies.

- [16] 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*, 20, 1, 1-24.
- [17] Blundell, R., Dearden, L., Meghir, C. and Sianesi, B. (1999), *The Returns from Education and Training to the Individual, the Firm and the Economy: A Review of the Theoretical and Empirical Literature*, Technical Report to the DfEE Skills Task Force, Institute for Fiscal Studies: London, 202pp.
- [18] Boheim, R. and Booth, A.L. (2001), "The Impact of Bargaining Institutions on Employer-provided Training", University of Essex Working Paper.
- [19] Booth, A. (1991), "Job-related formal training: who receives it and what is it worth?", *Oxford Bulletin of Economics and Statistics*, vol. 53, pp. 281-94.
- [20] Booth, A.L. and Zoega, G. (2000), "Why do Firms Invest in General Training: 'Good' Firms and 'Bad' Firms as a source of Monopsony Power", University of Essex Working Paper.
- [21] Dearden, L., Reed, H., Van Reenen, J. 2000. "Who gains when workers train? Training and corporate productivity in a Panel of British Industries". Center for Economic Policy Research, Discussion Paper, No. 2486.
- [22] European Commission. 1994. "Continuing Training in Enterprises: Facts and Figures". Center for Training Policy Studies. University of Sheffield, England.
- [23] Griliches, Zvi. 1977. "Estimating the Returns to Schooling: Some Econometric Problems", *Econometrica* 45, no.1: 1-22.
- [24] Griliches, Zvi. 1994. "Productivity, R&D, and the Data Constraint", *The American Economic Review*, Vol. 84, No. 1., pp. 1-23.
- [25] Griliches, Zvi and Jacques Mairesse (1995), "Production Functions: The Search for Identification", NBER wp. 5067.
- [26] Heckman, J., LaLonde, R. and Smith, J.. 1999. "The Economics and Econometrics of Active Labor Market Programs" in *Handbook of Labor Economics*, vol. 3A, Orley Ashenfelter and David Card, eds. Amsterdam: Elsevier.
- [27] Heckman, James, Lance Lochner and Christopher Taber (1998), "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Earnings with Heterogeneous Agents", *Review of Economic Dynamics*, 1, no 1, pp 1-58.
- [28] Heckman, James, Lance Lochner and Petra Todd (2003), "Fifty Years of Mincer Regressions", NBER wp 9732.
- [29] Holzer, H., Block, R., Cheatham and Knott, J.. 1993. "Are training subsidies for firms effective? The Michigan experience". *Industrial and Labor Relations*, vol 46.

- [30] Lalonde, R. (1986), “Evaluating the Econometric Evaluations of Training Programmes with Experimental Data”, *American Economic Review*, 76, 4, 604-620.
- [31] Lillard, L. and Tan, H..1986. “Training: Who Gets It and What Are Its Effects on Employment and Earnings?” RAND Corporation, Santa Monica California.
- [32] Lynch, L.. 1992. “Private-Sector Training and the Earnings of Young Workers”. *The American Economic Review*, Vol. 82, No. 1, pp. 299-312.
- [33] Machin, S. and Van Reenen, J. (1998) “Technology and Changes in Skill Structure: Evidence From Seven OECD Countries”, *Quarterly Journal of Economics*, 113, 1215-44.
- [34] Machin, S. and Vignoles, A. (2001), “The economic benefits of training to the individual, the firm and the economy”. mimeo. Center for the Economics of Education, UK.
- [35] Mincer, Jacob (1989), “Job Training: Costs, Returns and Wage Profiles”, NBER wp 3208.

Appendix

A. THE RETURNS TO SCHOOLING AND TRAINING

Similarly to schooling, on the job training is an investment in the individual's human capital. Therefore, we borrow the framework from the returns to schooling literature to compute the returns to job training. In the schooling literature the return to a year of schooling can be estimated from a regression of log wages on years of schooling:

$$\ln Y = \alpha + \beta S + \varepsilon. \quad (7.1)$$

The reason is that schooling is generally a full time activity and most of the cost of schooling is foregone earnings. Suppose an individual has s years of schooling and is considering whether to go to school one more year or not. If he decides to take one more year of schooling he cannot work during that year and he foregoes the annual wage he would have earned if had he entered the labor market with s years of schooling. Then the return to his investment is

$$\begin{aligned} R &= \frac{Y(s+1) - Y(s)}{Y(s)} = \\ &= \beta. \end{aligned}$$

where β is the coefficient on schooling in regression (7.1)³⁰. Assume now that s is years of job training and that an individual is considering whether to enrol in training one more year. Unlike schooling, training is a part time activity and, therefore, the foregone earnings cost of training is not $Y(s)$. If an individual spends a proportion γ of his working time in training (and if there are no direct costs of training), the returns to training would be:

$$\begin{aligned} R &= \frac{Y(s+1) - Y(s)}{\gamma Y(s)} = \\ &= \frac{\beta}{\gamma}. \end{aligned} \quad (7.2)$$

From the point of view of a firm deciding whether to provide training to its workers, foregone productivity can be computed as the product of the marginal product of worker's time and the amount of working time spent in training activities. This simple point, made by Mincer (1989), has rarely been picked up in the literature. One of the reasons is because the information on the percentage of time spent training within the working period is rarely available in surveys. In our sample the average training time that would not be working time had the training not taken place is approximately 50%. Therefore, assuming that all the training takes place during working time would lead a significant overestimation of the costs of training.

There is another important difference between standard analyses of schooling and our analysis of the returns to on-the-job training. In the estimation of the returns to schooling direct investment costs are usually neglected (one important exception is Heckman, Lochner

³⁰Heckman, Lochner and Todd (2003) discuss under what conditions we can interpret β as an internal rate of return to schooling.

and Todd, 2003). This procedure is justified by the fact that direct costs are small in absolute value because most of the education is publicly provided. Furthermore, even when these costs are substantial (such as college tuition in private universities) they are thought to be much smaller than the foregone earnings cost of schooling and therefore not worth considering. This assumption is more unattractive for the case of private training. Although some private training is subsidized, most of it is privately financed³¹. Moreover, because foregone productivity is smaller in the case of training, direct investment costs potentially account for a substantial fraction of total training costs. In our empirical work the direct costs of training turn out to be quantitatively very important, both in absolute terms and relative to the foregone productivity cost of training.

Finally, another important difference between the returns to schooling literature and our paper concerns the assumption on the depreciation rate of human capital. It is usually assumed that there is no depreciation rate for human capital acquired in school. However, if depreciation rates are positive then the estimated returns to schooling and training are smaller than β and $\frac{\beta}{\gamma}$ respectively.³² In our empirical work we consider different assumptions on the depreciation rates. In particular, we relate the depreciation rate with the rate of worker turnover within the firm.

B. IMPLICIT RATES FROM THE LITERATURE

One of the points of our paper is that we need to correctly estimate the costs of training in order to compute meaningful returns to training. By using the data available in several papers we can compute estimates of the implicit training costs in those papers. We can then compute the corresponding rate of return as net benefits divided by the total costs of training. To exemplify our argument, we compute the implicit returns to training for three papers in this literature: Dearden, Reed and Van Reenen (2000), Bartel (1994) and Black and Lynch (1998).

We assume that all costs occur in one initial period. The total benefit is given by the present discounted value of all future increases in output generated by the training that is offered in period t . Given that we do not have information on the direct costs of training for the other samples we compute two different returns. On the first one we focus only on the forgone output costs and, therefore, understate the total costs of training. We may be overstating the opportunity costs of training by assuming that all the training takes place during working hours. On the second return we assume that direct costs are the same proportion of total costs as in our data.

Table A1 presents estimates of returns to training implicit in the three examples we chose. The details of the construction of these estimates are available below.³³ This table shows

³¹In our sample 10% of the firms do not support any cost of training. The average percentage of costs of training supported by the firm is 75% (100% for the median firm).

³²If human capital does not depreciate over time the return to training is given by expression (7.2). An investment of $\gamma Y(s)$ today yields an increase in output of $Y(s+1) - Y(s) = \beta Y(s)$ in every year in the future. Therefore: $PV = -\gamma Y(s) + \frac{\beta Y(s)}{r}$, where PV is the present value of the investment in training. The r that solves $PV = 0$ (internal rate of return) is given by equation (7.2). In the extreme case that depreciation is 100% per year the benefits of training only occur in one year. Then present value of the investment is given by: $PV = -\gamma Y(s) + \frac{\beta Y(s)}{1+r}$. Solving for r we get that $R = \frac{\beta}{\gamma} - 1$.

³³These studies estimate production functions. We use the coefficient on training to compute the benefits of training and the coefficient on labor to compute the foregone productivity costs of training.

that once we properly account for the costs of training the implied estimates of returns to training implied from these studies are generally of an implausible magnitude.

B1. Dearden, Reed and Van Reenen (2000)

The specification of the production function in this paper is the following:

$$\ln \left(\frac{VA}{N} \right) = \alpha + \beta \left(\frac{T}{N} \right) + \gamma \ln \left(\frac{H}{N} \right) + X\theta + \varepsilon$$

where VA is value added, T is number of trainees, H is hours worked, N is number of employees and X are other controls. In their preferred specification: $\beta = 0.8$ and $\gamma = 0.42$. From their table of sample statistics, where they divide firms into high and low training firms, it is possible to obtain the following figures:

	Low T	High T
$E \left[\ln \left(\frac{VA}{N} \right) \right]$	2.76	3.19
H	39.1	40.2
$E \left(\frac{T}{N} \right)$	0.05	0.15

Replicating the exercise on their paper (page 49) suppose we increase $\frac{T}{N}$ from 10% to 15%. Then the effect on $\frac{VA}{N}$ is: $0.8 * (0.15 - 0.10) = 0.04$. These is the benefit of training.

What about the costs of this exercise? We cannot measure direct costs but we can say something about foregone productivity. Assume that workers, when they train, they train full time (this will give us an upper bound on the foregone productivity costs of training). In table 3.2 of their paper, we interpret “average hours worked” as average hours worked per employee per week: $\frac{H}{N} = 0.4$. By definition:

$$-\Delta H = \Delta T * \frac{H}{N}$$

because the change in total production hours per week (ΔH) has to equal the change in the number of trainees (individuals who stop working and start training: ΔT) multiplied by the amount of hours each one worked ($\frac{H}{N}$). Then:

$$\begin{aligned} -\Delta \left(\frac{H}{N} \right) &= -\frac{\Delta H}{N} \\ &= \frac{\Delta T}{N} * \frac{H}{N} \\ &= 0.05 * \frac{H}{N} \end{aligned}$$

which implies that:

$$d \ln \left(\frac{H}{N} \right) = \frac{\Delta \left(\frac{H}{N} \right)}{\frac{H}{N}} = -0.05.$$

Using their estimate for γ the effect on VA is:

$$\begin{aligned} d \ln \left(\frac{VA}{N} \right) &= \gamma * d \ln \left(\frac{H}{N} \right) \\ &= -0.42 * 0.05 \\ &= -0.02. \end{aligned}$$

Foregone productivity is equal to 2% of $\frac{VA}{N}$. This is the cost of training.

Finally the return is:

$$\begin{aligned} R &= \frac{Benefit - Cost}{Cost} \\ &= \frac{0.04 - 0.02}{0.02} \\ &= 100\% \end{aligned}$$

This estimate of returns assumes that: i) the effects of training last only for one period (100% depreciation rate); ii) workers who train, full time, and there is no joint production of training and output; iii) there are no direct costs of training. Relaxing assumptions i) and ii) leads to an increase in R while relaxing iii) leads to a decrease in R .

B2. Bartel (1994)

The estimated production function that we take from this paper is the following:

$$\ln\left(\frac{Y}{N}\right) = \alpha + \beta T + \gamma \ln N + X\theta + \varepsilon$$

where Y is output, T is an indicator for the presence of training in the firm, N is number of workers and X are other controls. We can rewrite this equation as

$$\ln Y = \alpha + \beta T + (\gamma + 1) \ln N + X\theta + \varepsilon.$$

When T corresponds to managerial training the estimates are the following: $\beta = 0.18$ and $\gamma + 1 = 0.12$. $T = 1$ means that more than 10% of the workforce (N) is trained (i.e., $d \ln N = 0.10$)

Assuming that a manager trains full time (no work while in training), the effect of one manager's work on output is 0.12 (N is just aggregate labor), training has a 100% depreciation rate and that the direct costs of training are 90% (average in our data) of the total training costs, then:

$$R = \frac{0.18 - 0.012 - 0.1}{0.012 + 0.1} = 50\%.$$

Here we also make the assumption that the % of direct costs on total costs is the same across occupations.

If $\beta_{Manager} > \beta = 0.12$ then R decreases. However, if depreciation rates are lower than 100% and/or managers train only part time, then all else constant, R should be higher than 360%. Using other occupations besides managers we estimate:

$$\begin{aligned} R_{Prof} &= \frac{0.27 - 0.01}{0.01} = 2600\% \\ R_c &= \frac{0.55 - 0.007}{0.007} = 7700\% \\ R_{Prod} &= \frac{0.2 - 0.01}{0.01} = 1900\%. \end{aligned}$$

B3. Black and Lynch (1998)

The specification estimated in this paper is the following:

$$\ln S = \alpha + \beta \ln T + \gamma \left(\frac{O}{H} \right) + \theta \ln H + X\eta + \varepsilon$$

where S is sales, T is number of trainees, O is off-work hours of training, H is hours of work and X are other controls. They estimate that: $\beta = 0.09$, $\gamma = 0.0002$ and $\theta = 0.47$.

For O we can assume there is no foregone productivity cost. However, we still need a measure of foregone productivity because not all training is done in off-work hours. In principle we can assume that:

$$H = h * N$$

where h is hours of work per worker and N is number of workers. From appendix D in their paper, $\frac{T}{N} = 0.49$ in 1993. Assuming that workers work full time (ignore O for now) then: $dT = dN$ (for a given size of the labor force). Then:

$$d \ln T = \frac{dT}{T} = \frac{dN}{T} = (d \ln N) * \frac{N}{T}.$$

Suppose that we increase T by 10%: $d \ln T = 0.1$. Then S increases by $0.1 * 0.08 = 0.009$. We also know that $d \ln N = (d \ln T) \frac{T}{N} = 0.1 * 0.49 = 0.049$. Assuming that h does not vary in this exercise, $d \ln H = d \ln N = 0.049$. Therefore, the reduction in sales is: $0.47 * 0.05 = 0.025$.

Assuming a depreciation rate for training of 100% per period we obtain:

$$R = \frac{0.009 - 0.025}{0.025} = -0.64 = -64\%.$$

In this exercise we ignored off-work hours of training so our costs are too high. If we also used a depreciation rate below 100% the estimated return would be higher. In particular, if the depreciation rate was 0% then $R = \frac{0.009}{0.025} = 0.36 = 36\%$.

C. THE DATA

The main source of data is “Balanco Social”, an annual survey designed by the Portuguese Ministry of Employment covering every firm with more than one hundred employees operating in Portugal³⁴. This is the first mandatory survey collecting longitudinal firm level data on training practices, productivity, wage bill and direct training costs. This paper covers the period 1995-1999.

The training information concerns only formal on-the-job training, i.e., structured training provided by the firm that is offered at the firm or at other location. Examples of formal training may include seminars, lectures, workshops, audio-visual presentations. In this survey there is yearly information for the number of training programs provided by the firm, the number of trainees involved in these programs³⁵, the total training hours and the total costs of training. Most of this information is available for two different types of training:

³⁴Public Administration is not included but state owned firms are. The survey accounts for approximately one third of the total private employment.

³⁵There is information for the yearly number of trainees but not for the number of workers enrolled. These will differ as long as the same worker participates in more than one training program per year.

internal and external to the firm, depending on whether training is offered inside the firm or in another location. Other variables available in the “Balço Social” and collected at the firm level include the firm’s regional location, ISIC five digit sector codes³⁶, total sales, value added, number of employees and a measure of capital, given by the book value of capital depreciation³⁷. Some worker characteristics available at the firm level include average age and tenure of the workforce and the proportion of males, as well as several measures of the firm’s employment practices (such as number of hires, fires and proportion of fixed term contracts in the firm).

The original data is composed of 2,923 firms. Due to the well known problems of estimating productivity in non-manufacturing we restrict the analysis to manufacturing (a total of 1,500 firms). Table A1 reports the sample means for the proportion of firms providing training programs and for the training hours per employee and per trainee. On average, 53% of the firms in the sample provide some training³⁸. Conditional on offering training, large firms, i.e., firms with more than 400 employees, train on average 24 hours per employee while smaller firms train approximately 15 hours. In general, sectors where there is a higher proportion of firms providing training also train more hours of training per employee and have a workforce with higher wages and years of schooling.

D. MARGINAL RETURNS TO TRAINING

An alternative indicator of the returns to training is the return on the marginal increase in training values. While average returns give information on the return of the total investment in training, the marginal return tell us the return obtained for increasing training for one hour to all employees in the firm. We report the marginal returns along with marginal benefits and costs for those firms that offer training programs in X ³⁹. The magnitude of the rates of return is higher than for the average returns because the marginal costs of training are lower than average benefits of training.

The marginal benefit of increasing training in one hour for all employees for those firms that already provide some training is obtained by computing $\hat{\gamma} * Y_{t+1}(T_t=T_t^*)$, where $\hat{\gamma}$ is the estimate obtained in Table 6 for the corresponding depreciation level and t is the period when training is provided. The marginal effects of an extra training hour for all employees are very similar to the average effects suggesting that the production function is close to linear in training over the range of values of training we observe. To estimate marginal costs of training we need to estimate direct cost functions.

The foregone productivity cost of training is the foregone output associated with spending

³⁶ISIC stands for “International Sector Industry Classification”.

³⁷It is a function of the book value of the firm’s capital stock. It depends on the value of capital and the book keeping value methods used to depreciate capital in the firm’s accounts. If the book value depreciation capital is linear: $BV_t = \pi * K_t$. It is a very imperfect measure of capital, but the only one available in this data set.

³⁸According to the EU report on continual training programs, the proportion of firms offering training in Portugal in the early nineties is smaller (13%) than for the EU average (57%). Their survey covers only continual training (while here I cover also up-front training, i.e., the training the worker receives when he is hired) and covers both small and large firms. It is a robust finding in the literature that smaller firms are more likely to offer training.

³⁹In ongoing work we compute the marginal return for those firms that do not offer training programs. We use the estimates from the cost function to compute marginal costs assuming that the fixed effects account for the selection of firms.

part of the working time in training. Because part of the training hours are spent off-the-job, an increase in training by one hour per employee in a year implies a decrease in annual working hours by less than one hour. Therefore, the marginal cost (excluding direct costs) of an extra hour of training equals the marginal productivity of labor. In most of the firms in the sample the training is either fully spent on-the-job (40%) or off-the-job (34%). The estimate of the marginal foregone productivity generated by increasing training by one hour for all the employees, is given by, $\frac{\partial Y_t}{\partial t} = \pi \hat{\beta} \frac{Y_t}{L_t}$, where π is the proportion of hours not worked due to training in the total hours of training in the firm, β is the coefficient of the working hours in the output regression and $\frac{Y_t}{L_t}$ is the average output per hour worked.

To compute the marginal cost of training we need to specify a cost function. Let C_{jt} be the total direct cost of training for firm j in period t . We assume the following functional form for costs of training:

$$C_{jt} = \varphi + \theta_1 h_{jt} + \theta_2 h_{jt}^2 + \phi_1 h_{jt} N_{jt} + \phi_2 h_{jt}^2 N_{jt} + \delta N_{jt} + \lambda X_{jt} + \rho_j + \nu_{jt} \quad (7.3)$$

where h_{jt} is the average hours of training per employee, N_{jt} is the total number of employees in the firm, X_{jt} is a set of firm characteristics including firm characteristics such as regional location and sector of activity of firm j and year dummy variables. All these variables are observed in our data set. ρ_j is a firm time invariant effect and ν_{jt} is an unobserved firm time varying shock affecting the cost function⁴⁰. ρ_j may be correlated with h_{jt} . With this functional form, the marginal cost of an hour of training for *all* the employees in the firm is a function of the training offered and of the initial size of the firm⁴¹. Equation (7.3) is estimated by least squares for the sample of all firms that offer training programs. One of the problems with estimating equation (7.3) using least squares for the sample of firms that offer training is the endogeneity of the training variable, i.e., the number of hours of training provided might be correlated to the time varying shocks in the cost function. We assume that including the firm fixed effect will account for this problem⁴². Regression results are ongoing work. 0.5cm

⁴⁰An example of a positive time varying shock for firm j might be the construction by local authorities of a training complex close to the firm headquarters, so that the cost of renting a training room falls from period t onwards for firm j .

⁴¹

$$\frac{\partial C_{jt}}{\partial h_{jt}} = \hat{\theta}_1 + 2\hat{\theta}_2 h_{jt} + \hat{\phi}_1 \bar{L}_{jt} + 2\hat{\phi}_2 h_{jt} \bar{L}_{jt}^2. \quad (7.4)$$

⁴²Another problem with estimating equation (7.3) is that firms that provide training might be very different from firms that do not provide training (sample selection problem). In particular, firms that provide training are likely to have a lower cost (or a higher benefit) of training than those who don't. We will not deal with this problem in this version of the paper, and assume it can also be mitigated by including a firm fixed effect.

**Table 1:
Review of the Literature using the "Production Function Approach"**

Authors	Survey, Unit of Analysis and Country	Training Measures	Productivity Measure	Specification	Point Estimate for the Training Variable (T ratio in brackets)	Other Controls
Bartel (IR, 1994)	Columbia Business School Survey. Firm and Occupation Level Data U.S.	Dummy if firm implemented a training program for an occupation group.	Log sales per worker	Y(t) on T(t)	Managers: 0.06 (.64) Professionals: 0.02 (.02) Clerical: - 0.001 (-.01) Production: -0.01 (-.09)	Age of the firm, % workers unionized dummies for other personnel policies Industry two-digit ISIC dummies.
		Dummy if firm implemented a training program for an occupation group.	Log sales per worker	Y(t+3)-Y(t) on T(t+3)-T(t)	Managers: 0.18 (2.4) Professionals: 0.31 (3.3) Clerical: 0.6 (4.2) Production: 0.25 (2.6)	
Black and Lynch (AER PP, 1998)	Educational Quality of the Workforce National Employers Survey Firm Level Data U.S.	Log Number Trainees			-0.12 (-1.2)	Log K, Log Materials Log Hours, Number plants, Age equipment and age firm, Log av. education workforce Computer, Teamwork and Supervisor training Tenure, Dummy if grades/communication skills are important in recruiting, dummy for the use of TQM or benchmarking exports dummy, unianized establishment. Industry two-digit ISIC dummies.
		Log Number Trainees 3 years before	Log of sales	Y(t) on T(t) and T(t-3)	0.09 (0.99)	
		Percentage of Training hours outside working hours			0.002 (2.1)	
Dearden, Reed and Van Reenen (CEPR DP, 2000)	Labor Force Survey and Annual Census of Production Three-digit SIC Level Data U.K.	Trainees per Employees	Log Value Added per worker	Y(t) on T(t)	0.5 (2.2) OLS	% turnover in t-1, log capital per employee, Log hours per employee, Log share R&D in Y in t-1 Industry proportion of males, age distribution in workforce sector occupational distributio, % of small firms in the sector. Regional, time and tenure dummies.
			Log Value Added per worker	Y(t) on T(t)	0.5 (3.1) WG	
			Log Value Added per worker	Y(t) on T(t)	1.3 (2.5) GMM	
Alba - Ramirez (OB, 1994)	Collective Bargaining in Large Firms Firm Level Data Spain	Junior Trainees per Junior Employees			-0.01 (-0.03)	Log size, Log K Log av hours worked, rate capacity utilization foreign and public dummy variables, tech change dummy % Y exported, fraction temporary workers % temporary contracts in newly hired workers % training in newly hired temporary contracts % apprenticeship in newly hired temporary workers % Workers in high level managers, low level managers and clerical workers.
		Senior Trainees per Senior Employees	Log Value Added per worker	Y(t) on T(t)	0.30 (2.01)	
Barrett and O'Connell (ILRR, 1999)	Eurostat data for 1993 and extension for 1995. Firm Level Data Ireland	Trainees per Employees	Log Sales per Employee	Y(t+2)-Y(t) on T(t)	0.099(1.8)	Change in Employment, Investment, Broad Sector Controls.
		Training Days per Employee	Log Sales per Employee	Y(t+2)-Y(t) on T(t)	0.014 (2.2)	
		Training Costs per Labor Costs	Log Sales per Employee	Y(t+2)-Y(t) on T(t)	0.005 (0.8)	

Table 2:
Returns, Benefits and Costs of an hour of training for all employees

	(1)	(2)	(3)	(4)	(5)
A. Production Function Estimates					
Human Capital Measure	0.0002 (0.0001)	0.0002 (0.0001)	0.0003 (0.0001)	0.0015 (0.0009)	0.0039 (0.0031)
Log hours worked p.e.	0.111 (0.01)	0.111 (0.01)	0.109 (0.01)	0.3672 (0.30)	0.48 (0.58)
Estimation Method	First Diff.	First Diff.	First Diff.	SYS-GMM	SYS-GMM
Skill Depreciation	100%	Firm Specific	17%	Firm Specific	Firm Specific
Year controls	Y	Y	Y	Y	Y
Allow Autocorrelation	N	N	N	N	Y
Observations	3,737	3,719	3,633	2,739	2,740
B. Yearly Benefits, Total Costs of Training and Average Returns to Training					
Benefits	1,116	1,105	1,607	7,493	14,963
Direct Cost	2,436	2,436	2,436	2,436	2,436
Foregone Output	134	131	120	405	530
as % total cost	11.1%	10.9%	10.3%	21.0%	24.0%
Return (IRR)	-48%	10%	57%	249%	490%

Source: author's calculations based on data from "Balanco Social".

Note: Return (IRR) is the median rate of return of the investment in training, Benefits is the median difference between the potential output and the effective output in the first period after training (in euros), Direct Cost is the median direct cost of training (in euros), Foregone Output is the median foregone output (in euros), Foregone Output as % of Total costs is the median share of the foregone output cost of training in the total costs of training.

**Table 3:
The Costs of Training**

	Mean	Median
Total Hours of Training per Employee	13.2	5.8
% Hours Trained in Total Hours	0.007	0.003
% Hours Trained outside Working Hours in Total Hours Trained	0.52	0.6
% Direct Costs of Training in Total Training Costs	0.95	0.9

Source: "Balanco Social"

Note: The table reports the mean and median of the variables reported conditional on firms offering some training. Total Training Costs are computed assuming a depreciation rate of 17% per period.

Table A1 :
Simulating the Rates of Return of the Training Investment in the Literature - Some Examples

Authors	Productivity Measure	Training Measure	Training Benefits	Training Costs	Returns
Bartel (IR, 1994)	Log sales per worker	Dummy if firm implemented a training program for managers.	18% increase in future productivity.	0.012 decrease in productivity.	1400%
Black and Lynch (AER PP, 1998)	Log of sales	Log number of Trainees	Increase of T in 10% increases future sales by 0.09%.	An increase of Trainees in 10% decreases current sales by 2.5%.	-64%
Dearden, Reed and Van Reenen (CEPR DP, 2000)	Log Value Added per worker	Trainees per Employees	An increase in T/N from 0.1 to 0.15 increases value added by 0.04	An increase in T/N from 0.1 to 0.15 increases costs in 0.02 of VA	100%

Source: Authors calculations. We assume that direct costs of training are the same proportion of total costs as in the Portuguese data.

Table A2: Balance of the Data

1.1 Yearly Observations	
1995	1,088
1996	1,155
1997	1,147
1998	1,072
1999	1,049

1.2 Panel Structure	
Firms observed for 1 year	213
Firms observed for 2 years	177
Firms observed for 3 years	192
Firms observed for 4 years	222
Firms observed for 5 years	696
Total Number of Firms	1,500

Source: "Balanco Social"

Table A3:
Medians of Some Variables by Training Intensity

	No Training Firms	Low Training Firms	High Training Firms
Value Added	1,934,465	3,460,467	9,313,104
Value Added p.e.	11,113	17,704	26,040
Employees	157	176	308
Total Hours p.e.	1,774	1,796	1,835
Capital Depreciation	248,035	595,769	1,563,233
Labor Costs p.e.	7,232	9,958	13,242
share low educated	0.86	0.77	0.63
share high educated	0.01	0.03	0.06
Av. Age	37	39	41
Share Males	0.4	0.6	0.7
Tenure	6	7	8
Higher Managers	0.01	0.02	0.03
Middle Managers	0.02	0.02	0.04
Intermediary Staff	0.04	0.05	0.05
High Qual. Prof.	0.41	0.42	0.43
Semi Qual. Prof	0.21	0.20	0.22
Non-Qual. Prof.	0.04	0.05	0.03
Apprentices	0.03	0.02	0.002
Hours Training p.e.	-	2	19
Training Hours / Total Hours	-	0.001	0.01
Direct Cost p.e.	-	17	158
Direct Cost / VA	-	0.001	0.005
Observations	2,578	1,461	1,462

Source: "Balanço Social" .

Note: All nominal variables are in Euros (1995 values). Low training firms are firms offering less than 1,489 total hours of training and High training firms are firms offering more than 1,489 total hours of training. Employees is the total number of employees in the firm. Total Hours p.e. is the yearly number of hours on-the-job (working or training) per employee, Capital Depreciation is the book value of capital depreciation, Labor Costs p.e. is total cost of labor supported by the firm excluding training expenditures per employee, share low educated is the share of workers with at most primary education, share high educated is proportion of workers with bachelor or college degrees, Av. Age is the average age of the workforce, Share Males is the proportion of male workers in the workforce, Tenure is the average tenure of workers, Higher Managers-Apprentices are the proportions of each occupational group in the total workforce of the firm, Hours Training p.e. is the total hours of training provided by the firm per employee.

Training Hours / Total Hours is the proportion of training hours in the total hours spent at work, Direct Cost p.e. is the total direct (monetary) cost of training per employee and Direct Cost / VA is the total direct cost of training as a proportion of the firm's value added.