THE INTERNAL RATE OF RETURN TO ON-THE-JOB TRAINING

Rita Almeida[*]

The World Bank

Pedro Carneiro

University College London, Institute for Fiscal Studies

and Center for Microdata Methods and Practice

November 2005

Abstract

In this paper we estimate the marginal internal rate of return to training using firm level data. Our data allows us to adequately account for the costs of training. We find that (unlike schooling) foregone output accounts for more than 75% of the total cost of training. This is not surprising given that in our sample firms train less than 1% of hours worked. We also quantify the effect of training on current and future output. We use a fixed effects instrumental variables estimator for dynamic panel data to estimate the production function (e.g., Blundell and Bond, 2000). In our preferred specification 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 about 0.5% and to smaller increases in future output since knowledge depreciates over time. These estimates imply that the private return to training among the firms providing training is 32%, and it is -9% for those providing no training. The former firms offer positive but low amounts of training, in spite of large returns. We conjecture that this is due to coordination problems between management and employees, and due to uncertainty in the returns to training.

Keywords: On-the-Job Training, Panel Data, Production Function, Rate of Return

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), the IZA/SOLE Meetings (Munich, 2004) the ZEW Conference on Education and Training (Mannheim, 2005), and the Econometric Society World Congress 2005, especially those of Steve Pischke. We also thank the comments of an anonymous reviewer. 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 on the firm (Heckman, Lochner and Taber, 1998), either through learning by doing or through on-the-job training. However, very little is known empirically about investments in on-the-job training. While there exist many public training programs for which we can find economic evaluations, most of the job training in the economy is done within firms and the evidence on their effectiveness is much more scarce.

Furthermore, like any other investment decision, firms will offer training programs as long as the expected benefits outweighed 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 (well defined) returns to training investments (one important exception is Mincer, 1989). Most of the literature focuses on quantifying the benefits of training for workers or firms with little or no concern with the costs of training (either with foregone productivity or direct costs). Most papers estimate the effects of on-the-job training on wages or productivity. In this paper we estimate the internal rate of return to formal on-the-job training.[1] Our data consists of 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 properly account for the costs of training.

Our estimated benefits are on par with or below most of the standard estimates of the literature. However, when we properly account for the costs of training we conclude that the implied estimates of the internal rate of return to training for the firms providing positive amounts of training are quite high. This implies that most of the estimates of the returns to training in the literature implicitly hide very large estimates of rates of return to training. Our estimated benefits for firms not currently providing training are generally quite low or negative. Our estimates are ex-post. We do not know the expected returns the firm faced at the time the firm made the decision of whether to invest or not invest.

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

Nevertheless, based on our estimates we argue that those firms deciding not to invest in training behave optimally given that return to training for them is on average so low. On the other end, those firms who offer training seem to be underinvesting given that their estimated average marginal return to training is very high. We conjecture that one possible reason for underinvestment may be coordination difficulties between the management and the workforce.

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

$$PDV = -C_t^T + \sum_{s=1}^{N} \frac{B_{t+s}}{(1+r)^s} = 0 \qquad (0.1)$$

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 is basically the marginal product of foregone production time. Our data is unusually rich for this exercise since it contains information on the duration of training, direct costs of training and training subsidies. In addition, we can also estimate the marginal productivity of labor and, therefore, the foregone productivity cost of training.

A major issue is to be able to identify the causal effect of training on output. In fact, we face the standard problem of how to identify parameters of the production function when there are missing inputs and inputs are chosen endogenously. Since we have access to a panel of firms, we will attempt to correct for this problem through a first difference strategy which uses past inputs as instruments for current differences in inputs, using the strategies developed by Arellano and Bond (1991), Arellano and Bover (1995) and Blundell and Bond (2000) for dynamic panel data models. This is also the approach taken by Dearden, Reed and Van Reenen (2000) in a recent paper on the effects of training on productivity and wages. From the production function we extract estimates of the marginal product of training and the marginal product of labor. We then estimate a direct cost of training function regressing the direct costs of training on polynomials in hours of training allowing for firm fixed effects. From this function we get the marginal direct cost of training.

When computing the benefits of the investment in training we recognize that what determines current output is not the current flow training, but the stock of training.[2] This has implications both for the estimation of the production function and for the computation of returns. It implies that training increases output contemporaneously but also in future periods. However, human capital depreciates over time, and in our paper this can occur for two reasons. The first is genuine depreciation of knowledge. The second is worker turnover. We assume that workers that leave the firm take with them not only labor but also human capital, while workers who join the firm have no human capital of the type which is provided by company training. In our empirical work, we compute a firm specific depreciation rate accounting for knowledge depreciation and worker turnover.

Well defined rates of return to training are almost never computed in the literature. Papers such as Bartel (1994), Black and Lynch (1998) or Dearden, Reed and Van Reenen (2000) report either the effect of training on wages, the effect of training on productivity or both. In the returns to schooling literature similar quantities are often defined as rates of return to schooling, because of the assumption that direct costs are irrelevant and that individuals attend school full time. However, this is not a valid exercise in the context of our paper because training is usually undertaken simultaneously with work and spend only a small proportion of their time in training, which means that the cost of foregone work will be small. Furthermore, direct costs of training are substantial and likely to be an important component of total costs. Mincer (1989) makes this point very clear and Machin and Vignoles (2001) emphasize the need to adequately consider costs in this literature.

[2] Note 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.

As a first exercise, we compute implicit internal rates of return for some of the papers in the literature estimating the effects of training on wages or productivity. As best as we could, we computed the foregone productivity costs of training implicit in these papers. Since there was no information on the direct costs of training we assumed these were zero. The numbers we computed are of absurd size, most of them extremely high, due to the fact that small amounts of training are estimated to generate sizeable benefits in terms of increases in training or productivity. For example, the implicit internal rate of return in Dearden, Reed and Van Reenen (2000) is estimated to be, at the very least, 100%; for Bartel (1994), we estimate rates of return between 50% and 2600%; for Black and Lynch (1998), we estimate a return of -64%. The details of these calculations are available in the appendix. These are very simple calculations but they serve to make the point that most of the papers in this literature have not given adequate consideration to the plausibility of the size of the effects that they present. We could claim that their estimates are still plausible because of the neglect of direct costs, but then these need to come into the picture explicitly in order to assess such a claim. Careful consideration of costs is an essential part of this whole exercise, and the main concern of this paper.

Several interesting facts emerge from our empirical analysis. The estimated marginal productivity of training seems to be quite large: an increase in the amount of training per employee of 10 hours per year, leads to an increase in current productivity of 0.5-1.5%. Increases in future productivity are dampened by the rate of depreciation of human capital but are still substantial. In a rough comparison, this estimate is below other estimates of the benefits of training in the literature (see Dearden et al, 2000, and Blundell et al, 1996). If the marginal productivity of labor were 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 (given that 10 hours corresponds to 0.5% of total hours worked in a year). However, with decreasing marginal productivity of labor, foregone productivity is even lower. In our data, marginal direct costs account on average for about 77% of the total marginal costs of training.

We then proceed to estimate the internal marginal rate of return to training investments. We assume human capital depreciates by 17% a year[3] and we estimate average turnover for each firm. Returns vary across firms with different levels of productivity. In particular, we estimate that (across specifications) the returns to training for firms providing no amount of training are generally either negative or quite small in magnitude. However, the lower bound for our estimates of the returns to training for firms providing training is 29%. Our estimates of returns 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 of the same magnitude or higher than the ones we estimate, and estimates of the marginal product of labor are comparable.[4]

It is possible that we (and the whole literature) overestimate the benefits of training because we do not account properly for the fact that training (as well as all other inputs) is endogenous. However, we allow for time invariant firm effects in the production function and we instrument current differences in training with lagged values of training (and current levels of training with lagged differences). We further allow time variant productivity shocks to be first-order auto correlated (as in Blundell and Bond, 2000). Our instruments are valid instruments if the error terms in our estimating equation in first differences are not correlated over time. We test and reject the null of no first order auto-correlation of the error terms in first differences, but we do not reject the null of no second order auto-correlation. Therefore, all our instruments are lagged at least two periods. We have multiple instruments and therefore we can perform a standard overidentification test, which (for most specifications) we do not reject at conventional levels of significance. Our estimates of the direct cost of training function also account for firm fixed effects.

We conjecture that those firms not providing training are behaving optimally given that their returns are so low. However, for those providing training, if returns are as high as we estimate them to be it is puzzling why they train such a small proportion of the total hours of work (less than 1%). We conjecture that suboptimal amounts of training may be the result of a coordination problem. Given that the benefits of training need to be shared between firms and workers, each party individually only sees part of the total benefit of

---

3 See Lillard and Tan (1986). Our attempts to estimate this parameter produced very imprecise estimates and therefore we decided to use a number within the range of reasonable estimates in the literature.

4 Mincer (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 total returns, since firms and workers should share the rents generated by training.

training. Unless investment decisions are coordinated and decided jointly, inefficient levels of investment may arise. Furthermore, our estimates of returns are ex post, and there is considerable heterogeneity. We do not estimate ex ante expected returns faced by firms, nor do we estimate the risk of human capital investments. Information problems and uncertainty may also lead firms to invest small amounts in training even though the ex post average return is substantial. A third hypothesis is that the firm is constrained in its investment opportunities, for example due to lack of access to credit.

The paper proceeds as follows. In section 2, we present our basic framework for estimating the production function of the firm, and therefore compute the marginal benefits of training and the marginal product of labor. We also briefly discuss the direct cost function. Section 3 describes the data used. In section 4 we present our empirical estimates of the costs and benefits of training and compute the marginal internal rate of return for investments in training. We conclude in section 5.

2. Basic Framework

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, 1991).[5] 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

[5] An alternative approach to measure the effct 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 incraese percentualy 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.

the stock of human capital in the firm.[6] 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.[7] 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.

In most of the work on this topic the stock of human capital is measured using current or lagged training. We relax this assumption allowing the stock of human capital to depend on the past history of human capital investments, 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:

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

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, η_j captures firm heterogeneity and

6 Some 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).

7 Bartel (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.

ε_{jt} captures time-varying productivity shocks.

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 $s = 1,..,N$. 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. In particular, 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 ($h_{jt} = \frac{H_{jt}}{Nj_t}$), X_{jt} is the number of new employees in period t and E_{jt} the employees who leave the firm in period t.[8] $\omega_{jt}^{'}$ 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:

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

where $\phi_{jt} = \frac{N_{jt} - E_{jt}}{N_{jt+1}}$ and $0 \le \phi_{jt} \le 1$. 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:

8 We 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).

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

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 (Human capital) into the log-linearized version of (Production function final) for period $t \in$ gives:

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

where $\mu_{jt} = \gamma(1-\delta)^t \phi_{j1}...\phi_{jt-1} H_{j0} + \eta_j$.[9]

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.[10] 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

[9] From this equation in principle it would be possible to estimate the depreciation rate δ. However, our attempts to do so yielded very imprecise estimates and we chose to set δ equal to 17%, a number which is reasonable in the literature (e.g., Lillard and Tan, 1986).

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. Another advantage of this approach is that it also corrects for biases generated by measurement error in the variables in the model.

However, the first differencing approach only works if μ_{jt} is a firm fixed effect, which only occurs 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 $\mu_{jt} - \eta_j$ is depreciates every period at rate $(1-\delta)\phi_{jt}$.[11] Given δ, and ϕ_{jt} it is probably possible to separately identify η_j, a time variant fixed effect, from $\gamma(1-\delta)^t \phi_{j1}...\phi_{jt-1} H_{j0}$, a firm effect that depreciates every period at a fixed rate. For computational reasons, we ignore this problem and approximate this parameter by a time invariant firm effect.[12] Therefore, in practice we employ a differencing procedure that does eliminate the influence of η_j, but not of H_0, in the regression (unless the latter is well approximated by an invariant firm effect). The extent of the resulting bias implied by this restriction depends on the correlation between H_0 and the choice of inputs. In future work we plan to relax this constraint.[13]

Blundell and Bond (2000) start from the equivalent of the following equation:

$$\ln Y_{jt} = \ln A_t + \alpha \ln K_{jt} + \beta_1 \ln l_{jt} + \beta_2 \ln N_{jt} + \gamma H_{jt} + \theta Z_{jt} + \mu_j + \varepsilon_{jt}$$

and assume that

$$\varepsilon_{jt} = \rho \varepsilon_{jt-1} + \varphi_{jt}$$

---

10 If when productivity is high firms allocate more resources to training, $\widehat{\gamma}$ 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 $\widehat{\gamma}$ will be underestimated (Dearden et al, 2000).

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

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

13 Our instrumental variables are not helpful fir this problem unless they are uncorrelated with H_0, which again is an unreasonable assumption.

where φ_{jt} is i.i.d. Then:

$$\ln Y_{jt} = \ln A_t + \alpha \ln K_{jt} + \beta_1 \ln l_{jt} + \beta_2 \ln N_{jt} + \gamma H_{jt} + \theta Z_{jt} + \mu_j + \\ + \rho \ln Y_{jt-1} - \rho \ln A_{t-1} - \rho a \ln K_{jt-1} - \rho\beta_1 \ln l_{jt} - \rho\beta_2 \ln N_{jt} - \rho\gamma H_{jt} - \rho\theta Z_{jt} \\ - \rho\mu_j - \varphi_{jt}$$

or

$$\ln Y_{jt} = \pi_0 + \pi_1 \ln K_{jt} + \pi_2 \ln l_{jt} + \pi_3 \ln N_{jt} + \pi_4 H_{jt} + \pi_5 Z_{jt} + \\ + \pi_7 \ln Y_{jt-1} + \pi_8 \ln K_{jt-1} + \pi_9 \ln l_{jt} + \pi_{10} \ln N_{jt} + \pi_{11} H_{jt} + \pi_{12} Z_{jt} \\ + \upsilon_j + \varphi_{jt}.$$

This formulation imposes some restrictions between the regression coefficients, usually called common factor restrictions. We first estimate the unrestricted model and them impose (using minimum distance) and test the validity of these restrictions, as is standard in this literature.

2.2. Computing the Marginal Benefits of Training and Marginal Foregone Productivity

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 proportion of males in the workforce, the average age 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 high education), time, region and sector effects. 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 period[14] and we estimate the turnover rate from the data. The average turnover in the sample is 14%.[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 $N_{t+1} = N_t + X_t - E_t$. 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

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

15 Although there is some sector variation in the turnover rates, all sectors report value above 20% per period.

average of the lead and lagged training values[16]. Total average skill depreciation in our sample is 25% per period.[17]

From the estimates of the production function above we compute the current marginal product of training and the marginal product of labor. We assume that the future marginal product of current training investments is equal to current marginal product of training minus depreciation. We measure training in hours per employee and we measure labor in number of workers. Therefore in order to get the marginal product of an hour of training per employee we divide the marginal product of an additional worker by the number of annual working hours per employee in the firm and we multiply the resulting quantity by the number of workers in the firm.

3. Direct 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 current output if they were not used in raising future output. Let the total training cost is given by $C_{jt}^{T} = C_{jt} + FP_{jt}$. In the previous section we saw how we compute FP_{jt}.

The data used in this paper allows us to compute estimates of the total training costs because 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). We the estimate a direct cost function by regressing direct costs of training on quadratic spline in total hours of training, where the knot point is 2500, roughly the median number of hours of training in the sample. We also include a time invariant firm effect in the cost function. Because we measured benefits in terms of hours of training per employee in order to get a comparable measure of marginal direct costs of training we multiple the marginal cost of training from the above function by the number of

---

16 This 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.

17 In results available on request, we perform some sensitivity analysis of the rates of return with respect to the skill depreciation.

workers in the firm (since we want the marginal cost of giving one extra hour of training to every worker).[18]

Once we have marginal benefits and marginal costs we can compute individual specific marginal internal rates of return, by solving equation (1.1) for each firm.

4. The Data

The main source of data is the "*Balanço 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 information on training measures, total

[18] Full estimates of the production and cost functions are available on request.

[19] However, 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.

[20] Bartel (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.

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]

Table 2 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

[21] Dearden, 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.

[22] The 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.

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

[24] Arulampalam, 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.

[25] In 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.

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., 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. Empirical Results

As our measure of the return to 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 (PV) to zero:

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

Panel A of Table 3 presents the coefficients on labor input (number of workers) and stock of training in the production function (equation (2.3)). The different columns present the results under different specifications. Column (I) has the lowest number of controls and column (II) has the highest number of controls.

The point estimates of the effect of training on output are sensitive to the set of controls that is used. However, the overall pattern is that as the set of controls becomes richer the point estimates on the training variable decrease, while variation in the point estimate of the coefficient on labor input does not follow a systematic pattern. The estimated benefits in all the columns of the table seem to be quite high: a increase in the amount of training per employee of 10 hours (approximately 0.5% of total hours worked per year), leads to an increase in current productivity of 0.5-1.5%. 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 product of labor, foregone productivity is even lower. Our preferred estimates are the ones of column (VIII) since those include the full set of controls. We present the remaining columns for completeness and analysis of the sensitivity of results.

For each column we present four sets of tests. The first one is a test of overidentifying restrictions (Hansen test). For most specifications (including VIII, our preferred one) we do not reject these restrictions, although this does not happen for all of them. The second test is a test of first order autocorrelation of the error term in first differences. For every specification we reject the null of no first order autocorrelation, and therefore all our instruments are included with a lag of 2 or higher. Across specifications, we do not reject the null of no second order autocorrelation. Finally, the last test is a test of the validity of the common factor restrictions, which we impose (using minimum distance) to get the estimates in panel A of table 3. Again, for most specifications, we do not reject the validity of the common factor restrictions.

One important 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. Median marginal costs of training are reported in panel B of table 3. On average, foregone productivity accounts for less than 25% of the total costs of training. Notice that marginal direct costs of training are much higher for firms offering no training than for firms offering positive amounts of training, suggesting that there are large fixed costs of setting up a training program in a firm. Panel B of Table 3 also presents, for each specification, the median estimates of the benefits of giving all employees one extra hour. Benefits are measured one period after the training is offered.

Our estimates of the median marginal internal rate of return for the whole sample ranges from 0 to 50%, it ranges from -9 to 23% for the firms who do not provide training, and it ranges from 29 to 139% for the set of firms offering training. Our preferred specification is in column VIII, which includes the full set of controls. Notice that the larger the set of controls included the smaller the return, suggesting that our returns may be overestimated. With the exception of columns (I) and (IV), the returns to training for firms not providing training are always below 5% and are negative in some cases. The reason these firms do not offer training may be precisely because they face low returns and therefore they may be acting rationally and optimally.

However, the returns for firms providing training are quite high, our lower bound being of 29% and our preferred estimate being 32%.With such high returns, it is puzzling why firms train such a small proportion of the total hours of work (less than 1%). Even though ours may be overestimates of the true internal rate of return (as suggested before), we believe that if such upward bias exists it should not be as high as to make these numbers decrease considerably. The reason is that we are being as careful as possible in the estimation of the production function, by including fixed effects and instrumenting current inputs with lagged inputs (with a large enough lag so that our results are robust to autocorrelation in the error terms in differences). Therefore, we conjecture that suboptimal amounts of training may be the result of a coordination problem. Given that the benefits of training need to be shared between firms and workers, each party individually only sees part of the total benefit of training. Unless investment decisions are coordinated and decided jointly, inefficient levels of investment may arise.

Furthermore, our estimates of returns are ex post, and there is considerable heterogeneity. We do not estimate ex ante expected returns faced by firms, nor do we estimate the risk of human capital investments. Information problems and uncertainty may also lead firms to invest small amounts in training even though the ex post average return is substantial. In fact, we find significant heterogeneity in the ex post returns to training, which may indicate also a large amount of uncertainty (although in our current framework we cannot distinguish ex ante heterogeneity from uncertainty). Focusing on our preferred specification, the 10th and 90th percentile of the distribution of overall returns to training are -13% and 86%, for firms providing training these quantities are -15% and 11%, and for those providing training we have -6% and 94%.

There is a third type of explanation that is usual in this type of situations. It is possible that firms would like to invest more in their workers but they are constrained (e.g., credit constrained). In that case, investments in training are likely to be suboptimal.

6. *Conclusion*

In this paper we estimate the marginal internal rate of 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 of the firm level survey allows us to improve the existing literature in several dimensions. We find that (unlike schooling) foregone output accounts for more than 75% of the total cost of training. This is not surprising given that in our sample firms train less than 1% of hours worked. We also quantify the effect of training on current and future output. Even tough disentangling correlation and causality is difficult, we use a fixed effects instrumental variables estimator for dynamic panel data to address the problem (e.g., Blundell and Bond, 2000). In our preferred specification 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 about 0.5% and to smaller increases in future output since knowledge depreciates over time. These estimates imply that the private return to training among the firms providing training is 32%, and it is -9% for those providing no training. The former firms offer positive but low amounts of training, in spite of large returns. We conjecture that this is due to coordination problems between management and employees, and due to uncertainty in the returns to training.

References

Arulampalam, W., Booth A., and Bryan, M. 2003. “Training in Europe”. University of Essex working paper.

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.

Alba-Ramirez, A.1994. “Formal Training, Temporary Contracts, Productivity and Wages in Spain”. *Oxford Bulletin of Economics and Statistics*, vol 56.

Barron, J., Black, D. and Lowenstein, M.1989. “Job Matching and On-The-Job Training” *Journal of Labor Economics*. vol 7.

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.

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.

- -. 1997. “How to Compete: The Impact of Workplace Practices and Information Technology on Productivity” *National Bureau Economic Research Working Paper* No. 6120.

- -. 1998. “Beyond the Incidence of Training: Evidence from a National Employers Survey” *Industrial and Labor Relations Review*, Vol.52, no.1.

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 Trainingof Adult Employees, ed. David Stern and Jozef Ritzen, Springer-Verlag.

- -.1994. “Productivity Gains From the Implementation of Employee Training Programs”. *Industrial Relations*, vol. 33, no. 4.

- -. 1995. “Training, Wage Growth, and Job Performance: Evidence from a Company Database”. *Journal of Labor Economics,* Vol. 13, No. 3, pp. 401-425.

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.

Blundell, R. and Stephen, B. (1998). “Initial Conditions and Moment Restrictions in

Dynamic Panel Data Models." *Journal of Econometrics* 87: 115-143.

Blundell, R. and Stephen, B. (2000), "GMM Estimation with Persistent Panel Data: An Application to Production Functions", *Econometric Reviews*, 19.

Blundell, R., Dearden, L., Goodman, A. and Reed, H. (1997), Higher Education, Employment and Earnings in Britain, London: Institute of Fiscal Studies.

Blundell, R., Dearden, L. and Meghir, C. (1996), The Determinants of Work Related Training in Britain, London: Institute of Fiscal Studies.

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.

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.

Boheim, R. and Booth, A.L. (2001), "The Impact of Bargaining Institutions on Employer-provided Training", University of Essex Working Paper.

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.

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.

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.

European Commission. 1994. "Continuing Training in Enterprises: Facts and Figures". Center for Training Policy Studies. University of Sheffield, England.

Griliches, Zvi. 1977. "Estimating the Returns to Schooling: Some Econometric Problems". *Econometrica* 45, no.1: 1-22.

Griliches, Zvi. 1994. "Productivity, R&D, and the Data Constraint", *The American Economic Review*, Vol. 84, No. 1., pp. 1-23.

Griliches, Zvi and Jacques Mairesse (1995), "Production Functions: The Search for

Identification", *NBER wp*. 5067.

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.

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.

Heckman, James, Lance Lochner and Petra Todd (2003), "Fifty Years of Mincer Regressions", *NBER wp* 9732.

Holzer, H., Block, R., Cheatham and Knott, J. 1993. "Are training subsidies for firms effective? The Michigan experience". *Industrial and Labor Relations*, vol 46.

Lalonde, R. (1986), "Evaluating the Econometric Evaluations of Training Programmes with Experimental Data", *American Economic Review*, 76, 4, 604-620.

Lillard, L. and Tan, H..1986. "Training: Who Gets It and What Are Its Effects on Employment and Earnings?" RAND Corporation, Santa Monica California.

Lynch, L. 1992. "Private-Sector Training and the Earnings of Young Workers". *The American Economic Review*, Vol. 82, No. 1, pp. 299-312.

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.

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.

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:

$$\ln Y = \alpha + \beta S + \varepsilon. \tag{6.1}$$

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

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

where β is the coefficient on schooling in regression (wage return)[26]. 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:

$$R = \frac{Y(s+1) - Y(s)}{\gamma Y(s)} = \\ = \frac{\beta}{\gamma}. \tag{6.2}$$

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

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 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 financed[27]. 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

[27] In 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).

training are smaller than β and $\frac{\beta}{\gamma}$ respectively.[28] 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.[29] This

---

[28] If human capital does not depreciate over time the return to training is given by expression (Return intro). An investment of $\gamma Y(s)$ today yelds 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 (Return intro). 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 $R=\frac{\beta}{\gamma}-1$.

[29] These 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.

table shows 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: $0.8*(0.15-0.10)=0.04$. This 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 teh % of direct costs on total costs is the same across occupations.

If $\beta_{Manager} > \beta = 0.12$ then $R\sqrt{b^2 - 4ac}$ 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:

$$R_{\text{Prof}} = \frac{0.27 - 0.01}{0.01} = 2600\%$$
$$R_{\text{c}} = \frac{0.55 - 0.007}{0.007} = 7700\%$$
$$R_{\text{Prod}} = \frac{0.2 - 0.01}{0.01} = 1900\%.$$

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:

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

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 $d \ln N = \left(d \ln T \right) \frac{T}{N} = 0.1 * 0.49 = 0.049$. 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 *Balanço Social*, an annual survey designed by the Portuguese Ministry of Employment covering every firm with more than one hundred employees operating in Portugal[30]. 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 programs[31], the total training hours and the total costs of training. Most of this information is available for two different types of training: internal and external to the firm, depending on whether training is offered inside the firm or in another location. Other variables available in the *Balanço Social* and collected at the firm level include the firm's regional location, ISIC five digit sector codes[32], total sales, value added, number of employees and a measure of capital, given by the book value of capital depreciation[33]. 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 A 1 reports the sample means for the proportion of firms providing training programs and for the training hours per employee and per trainee. On

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

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

[32] ISIC stands for International Sector Industry Classification.

[33] It 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 deprecition 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.

average, 53% of the firms in the sample provide some training[34]. 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.

[34] According to the EU report on continual training programs, the proportion if 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.

Table 1 (cont.)

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	Log Number Trainees Log Number Trainees 3 years before	Log of sales	Y(t) on T(t) and T(t-3)	-0.12 (-1.2) 0.09 (0.99)	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

Table 1 (cont.)

Review of the Literature using the "Production Function Approach"

	U.S.	Percentage of Training hours outside working hours			0.002 (2.1)	skills are important in recruiting, dummy for the use of TQM or benchmarking exports dummy, unianized establishment. Industry two-digit ISIC dummies.
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	
			Log Value Added per worker	Y(t) on T(t)	0.5 (3.1) WG	% 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)	1.3 (2.5) GMM	
		Junior Trainees per			-0.01 (-0.03)	Log size, Log K

Table 1 (cont.)

Review of the Literature using the "Production Function Approach"

Alba - Ramirez	Collective Bargaining in Large Firms	Junior Employees				Log av hours worked, rate capacity utilization
(OB, 1994)	Firm Level Data		Log Value Added	Y(t) on T(t)		foreign and public dummy variables, tech change dummy
	Spain	Senior Trainees per	per worker		0.30 (2.01)	% Y exported, fraction temporary workers
		Senior Employees				% temporary contracts in newly hired workers
						% training in newly hired temporary contracts
						% apprenticeship in newly hired temporray workers
						% Workers in high level managers, low level managers
						and clerical workers.
		Trainees per	Log Sales	Y(t+2)-Y(t) on T(t)	0.099(1.8)	
		Employees	per Employee			
Barrett and O'Connell	Eurostat data for 1993 and extension for 1995.					Change in Employment, Investment,
(ILRR, 1999)	Firm Level Data	Training Days per	Log Sales	Y(t+2)-Y(t) on T(t)	0.014 (2.2)	Broad Sector Controls.

Table 1 (cont.)

Review of the Literature using the "Production Function Approach"

Ireland	Employee	per Employee		
	Training Costs per Labor Costs	Log Sales per Employee	Y(t+2)-Y(t) on T(t)	0.005 (0.8)

Table 2 (cont.)

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

Table 2 (cont.)

Medians of Some Variables by Training Intensity

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
Apprenteces	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.

Table 3
Returns, Benefits and Costs of an hour of training for all employees

	I	II	III	IV	V	VI	VII	VIII
A. Production Function Estimates								
Human Capital Measure	0.0015	0.0009	0.0010	0.0005	0.0011	0.0007	0.0009	0.0005
	(0.0007)**	(0.0005)*	(0.0004)***	(0.0004)	(0.0004)***	(0.0003)**	(0.0004)**	(0.0003)*
Log Number of Employees	0.7591	0.7185	0.8581	0.7898	0.6142	0.6645	0.6985	0.7140
	(0.1873)***	(0.1466)***	(0.1609)***	(0.1301)***	(0.1236)***	(0.1099)***	(0.1288)***	(0.1104)***
Amortization	Y	Y	Y	Y	Y	Y	Y	Y
Shares in Different Occupations	N	N	N	N	Y	Y	Y	Y
Average Age, Year, Region and Sector Controls	Y	Y	Y	Y	Y	Y	Y	Y
Share Highly Educated	N	Y	N	Y	N	Y	N	Y
Share Male	N	N	Y	Y	N	N	Y	Y
Observations	3.790	3.790	3.787	3.787	3.790	3.790	3.787	3.787
P-Value Hansen Over-Identification Test	0.393	0.375	0.022	0.024	0.483	0.667	0.358	0.291
P-Value Test for AR(1) in First Differences	0.035	0.038	0.035	0.036	0.032	0.030	0.024	0.025
P-Value Test for AR(2) in First Differences	0.168	0.238	0.115	0.157	0.159	0.163	0.128	0.137
P-Value Common Factor Restrictions	0.128	0.072	0.764	0.278	0.797	0.372	0.382	0.361

Table 3

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

	I	II	III	IV	V	VI	VII	VIII
B. Yearly Benefits, Total Costs of Training and Average Returns to Training								
Median Marginal Benefits (whole sample)	960	576	640	320	704	448	576	320
If Hours of Training = 0	601	360	401	200	441	280	360	200
If Hours of Training > 0	1.548	929	1.032	516	1.135	722	929	516
Median Marginal Direct Cost (whole sample)	863	863	863	863	863	863	863	863
If Hours of Training = 0	1.042	1.042	1.042	1.042	1.042	1.042	1.042	1.042
If Hours of Training > 0	623	623	623	623	623	623	623	623
Median Marginal Foregone Output (whole sample)	279	264	315	290	225	244	256	262
If Hours of Training = 0	132	163	194	179	139	150	158	162
If Hours of Training > 0	439	415	496	457	335	384	404	413
Median Marginal IRR (whole sample)	*49%*	*20%*	*23%*	*0%*	*32%*	*11%*	*21%*	*0%*
If Hours of Training = 0	*21%*	*3%*	*5%*	*-9%*	*10%*	*-2%*	*3%*	*-9%*
If Hours of Training > 0	*139%*	*76%*	*79%*	*29%*	*106%*	*56%*	*77%*	*32%*

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

*** Significant at 1%, ** Significant at 5%, * Significant at 10%.

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.