



THE RETURN TO FIRM INVESTMENT IN HUMAN CAPITAL

Rita Almeida
Pedro Carneiro

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP21/05

The Return to the Firm Investment in Human Capital

Rita Almeida*
The World Bank

Pedro Carneiro
University College London, Institute for Fiscal Studies
and Center for Microdata Methods and Practice

December 2005

Abstract

In this paper we estimate the rate of return to firm investments in human capital in the form of formal job training. We use a panel of large firms with unusually detailed information on the duration of training, the direct costs of training, and several firm characteristics such as their output, workforce characteristics and capital stock. Our estimates of the return to training vary substantially across firms. On average it is -7% for firms not providing training and 24% for those providing training. Formal job training is a good investment for many firms and the economy, possibly yielding higher returns than either investments in physical capital or investments in schooling. In spite of this, observed amounts of formal training are very small.

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

JEL Classification codes: C23, D24, J31

*We thank conference participants at the European Association of Labor Economists (Lisbon, 2004), Meeting of the European Economic Association (Madrid, 2004), the IZA/SOLE Meetings (Munich, 2004), ZEW Conference on Education and Training (Mannheim, 2005), and the Econometric Society World Congress 2005. We thank especially the comments made by Manuel Arellano and 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

Europe is struggling to become a competitive and dynamic knowledge economy in the modern world. One key ingredient in its strategy is likely to be the investment in human capital. Although individuals invest in human capital over the whole life-cycle, more than one half of lifetime human capital is accumulated through post-school investments on the firm (Heckman, Lochner and Taber, 1998). This happens either through learning by doing or through formal on-the-job training. In spite of its importance, economists know surprisingly less about the incentives and returns to firms of investing in training compared with what they know about the individual's returns of investing in schooling.¹ Similarly, the study of firm investments in physical capital is much more developed than the study of firm investments in human capital, even though the latter may well be at least as important as the former in modern economies. In this paper we estimate the internal rate of return of firm investments in human capital. We use a census of large manufacturing firms in Portugal between 1995 and 1999 with detailed information on investments in training, its costs, and several firm characteristics.²

Most of the empirical work to date has focused on obtaining estimates of the return to training for workers using data on wages (e.g., Bartel, 1995, Arulampalam, Booth and Elias, 1997, Mincer, 1989, Frazis and Loewenstein, 2005). Even though this exercise is very useful, it has important drawbacks (e.g., Pischke, 2005). For example, with imperfect labor markets

¹An important part of the lifelong learning strategies are the public training programs. There is much more evidence about the effectiveness (or lack of it) of such programs compared with the available evidence on the effectiveness of the private on-the-job training.

²We will consider only formal training programs and abstract from the fact that formal and informal training could be very correlated. This is a weakness of most of the literature, since informal training is very hard to measure.

wages do not fully reflect the marginal product of labor, and therefore the wage return to training tells us little about the effect of training on productivity. Moreover, the “effect” of training on wages depends on whether training is firm specific or general (e.g., Becker, 1962, Leuven, 2004).³ As in the literature that focus on the effects of training on productivity, our parameter of interest is the return to training for employers and employees as a whole, irrespective of how these returns are shared between the two parties. In contrast, most of the literature on the wage returns to training focuses on the return to training for the individual employee. The few papers estimating the effects of training on productivity have little or no mention of the costs of training (e.g. Bartel, 1991, 1994, 2000, Black and Lynch, 1998, Barret and O’Connell, 1999, Dearden, Reed and Van Reenen, 2005). This happens most probably due to lack of adequate data. As a result, we cannot interpret the estimates in these papers as well defined rates of return.⁴

The data we use is unusually rich for this exercise since it contains information on the duration of training, direct costs of training to the firm as well as productivity data. This allow us to estimate both a production and a cost function and to obtain estimates of the marginal benefits and costs of training to the firm. In order to estimate the total marginal costs of training, we need information on the direct cost of training and on the foregone productivity cost of training. The first is observed in our data while the second is the marginal product of worker’s time while training. The major problems in this exercise are the treatment of omitted variables and the endogenous choice of inputs in the production and

³For example, Leuven and Oosterbek (2002, 2004) argue that they may be finding low or no effects of training because they are using individual wages as opposed to firm productivity.

⁴This shortcoming of the literature as been emphasized in Mincer (1989) and Machin and Vignoles (2001) among others.

cost functions. Given the panel structure of our data, we address these issues applying the methods developed in Blundell and Bond (2000), which build on Arellano and Bond (1991) and Arellano and Bover (1995). In particular, we estimate the cost and production functions using a first difference instrumental variable approach. By computing first differences we control for firm unobservable and time invariant characteristics. By using lagged values of inputs to instrument current differences in inputs (together with lagged differences in inputs to instrument current levels) we account for any correlation between input choices and transitory productivity or cost shocks. Our instruments are valid as long as the first differences of transitory shocks in the production and cost functions do not exhibit high order serial correlation. In our empirical application we allow for first order serial correlation but not for higher order serial correlation.

Several interesting facts emerge from our empirical analysis. First, in line with the previous literature (e.g., Pischke, 2005, Bassanini, Booth, De Paola and Leuven, 2005) our estimates of the effects of training on productivity are quite high: an increase in the amount of training per employee of 10 hours per year, leads to an increase in current productivity of 0.6%. 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 (e.g., Dearden, Reed and Van Reenen, 2005, Blundell, Dearden and Meghir, 1999). 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 at most 0.5% of output (assuming all training

occurred during working hours).⁵ With decreasing marginal product of labor (and because roughly 50% of training occurs outside normal working time) foregone productivity is much lower.

Second, we estimate that, on average, foregone productivity accounts for less than 25% of the total costs of training. This finding is of interest for two reasons. On the one hand, it shows that the simple returns to schooling intuition is inadequate for studying the returns to training. In particular, the coefficient on training in a production function is unlikely to be a good estimate of the return to training.⁶ On the other hand, without information on direct costs of training, estimates of the return to training will be too high since direct costs account for the majority of training costs.

Finally, our estimates of the internal rate of return to training vary across firms. While investments in human capital have on average negative return for those firms which do not provide training, returns for firms providing training are quite high (24%). Such high returns suggest that on-the-job training is a good investment for firms and for the economy as a whole, possibly yielding higher returns than either investments in physical capital or investments in schooling.

As a consequence, it is puzzling why these firm train on average such a small proportion of the total hours of work (less than 1%). One hypothesis is that suboptimal amounts of training may be the result of a coordination problem, as emphasized in Pischke (2005). 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.⁷ Unless investment decisions

⁵For an individual working 2000 hours a year, 10 hours corresponds to 0.5% of annual working hours.

⁶This is also likely to be a problem in wage regressions.

⁷This may be also due to the so called "poaching externality" (Stevens, 1994). See also Acemoglu and

are coordinated and decided jointly, inefficient levels of investment may arise. Furthermore, information problems and uncertainty in this investment in human capital may lead firms to invest small amounts in training even though the *ex post* average return is high. Even though under our current set up we cannot distinguish how much of the variability in returns across firms is due to heterogeneity and how much is due to uncertainty we find an enormous dispersion in the *ex post* returns to training which may be suggestive of the importance of uncertainty. For example, in our base specification the 5th percentile of the distribution of internal rates of return is -16% and the 95% percentile is 66%. Finally, it is possible that firms would like to invest more in their workers but they are unable to do so because they are constrained (e.g., credit constrained). In that case, investments in training are likely to be suboptimal. Unfortunately we cannot verify empirically the importance of each of these different hypotheses.

The paper proceeds as follows. Section 2 describes the data we use. In section 3, we present our basic framework for estimating the production function and the cost function. 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. Section 5 concludes.

2. Data

We use data from an annual survey collected by the Portuguese Ministry of Employment (*Balanço Social*) covering all the firms with more than 100 employees operating in Portugal.

The survey is mandatory and collects information on hours of training provided by the

Pischke (1998, 1999) for an analysis of the consequences of imperfect labor markets for firm provision of general training.

employers and on the direct training costs at the firm level. Other variables available at the firm level include the firm's location, ISIC 5-digit sector of activity, value added, number of workers and a measure of the capital, given by the book value of capital depreciation⁸, average age of the workforce and share of males in the workforce. It also collects several measures of the firm's employment practices such as the number of hires and fires within a year (which will be important to determine average worker turnover within the firm). We use information for manufacturing firms between 1995-1999. This gives us a panel of 1,500 firms (corresponding to 5,501 firm-year observations). On average, 53% of the firms in the sample provide some training.

Relative to other datasets that are used in the literature, the one we use has several advantages for computing the internal rates of return of investments in training. First, information is reported by the employer. This may be better than having employee reported information about past training if the employee recalls less and more imprecisely the information about on-the-job training. Second, training is reported for all employees in the firm, not just new hires. Third, the survey is mandatory for firms with more than 100 employees (34% of the total workforce in 1995). This is an advantage since a lot of the empirical work in the literature uses small sample sizes and the response rates on employer surveys tend to be low.⁹ Fourth, it collects longitudinal information for training hours, firm productivity and direct training costs at the firm level. More than 50% of the firms are observed at least

⁸We assume that depreciation is a linear function of the book value of the firm's capital stock : $Dep_t = \pi * K_t$.

⁹Bartel (1991) uses a survey conducted by the Columbia Business School with a 6% response rate. Black and Lynch (1997) use data on the Educational Quality of the Workforce National Employers survey, which is a telephone conducted survey with a 64% "complete" response rate. Barrett and O'Connell (2001) expand an EU survey and obtain a 33% response rate.

twice during the period 1995-1999.¹⁰

Table 1 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. We choose to report medians rather than means to avoid extreme sensitivity to extreme values. Firms that offer training programs and have a high training intensity 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 have a higher stock of physical capital. 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 6% and 3% in high and low training firms, versus 1% in non-training firms. The workforce in firms that offer training has a higher proportion of male workers.¹¹ 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 1.6 hours of training per employee per year while those that offer a large amount of training offer 19 hours of training. Even though the difference between the two groups of firms is large, the number of training hours even for high training firms looks very small when compared with the 2055 average

¹⁰Firms can leave the sample because they exit the market or because total employment is reduced to less than 100 employees.

¹¹Arulampalam, Booth and Bryan (2004) 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.

annual hours on-the-job for the (0.9% of total time on-the-job). High training firms spend almost 8 times more in training per employee than low training firms. These costs are 0.05% and 0.3% of value added respectively. This proportion is rather small, but is in line with the small amounts of training being provided.

In sum, firms train a rather small amount of hours. This pattern is similar to other countries in the south of Europe (Italy, Greece, Spain) as well as in Eastern Europe (e.g., Bassanini, Booth, De Paola and Leuven, 2005). We find a lot of heterogeneity among the firms that offer training, with low and high training firms being very different. Finally, firms spend a small proportion of their value added with formal training programs which is in line with training a small proportion of the working hours.

3. Basic Framework

Our parameter of interest is the internal rate of return to the firm of an additional hour of training per employee. Let MB_{t+s} be the marginal benefit of an additional unit of training in t and MC_t be the marginal cost of the investment in training at t . Assuming that the cost is all incurred in one period and that the investment generates benefits in the subsequent N periods, the internal rate of return of the investment is given by the rate r that equalizes the present discounted value of net marginal benefits to zero:

$$\sum_{s=1}^N \frac{MB_{t+s}}{(1+r)^s} - MC_t^T = 0 \quad (3.1)$$

Training involves a direct cost and a foregone productivity cost. Let the marginal training cost be given by: $MC_t^T = MC_t + MFP_t$, where MC_t is the marginal direct cost and

MFP_t is the marginal product of foregone worker time. In the next sections we lay out the basic framework which we use to estimate the components of MC_t^T and MB_{t+s} . To obtain estimates for MFP_t and MB_{t+s} , in section 3.1 we estimate a production function and to obtain estimates for MC_t in section 3.2 we will estimate a cost function.

3.1. Estimating the Production Function

We assume that the firm's production function is semi-log linear and that the firm's stock of human capital determines the current level of output¹²:

$$Y_{jt} = A_t K_{jt}^\alpha L_{jt}^\beta \exp(\gamma h_{jt} + \theta Z_{jt} + \mu_j + \varepsilon_{jt}) \quad (3.2)$$

where Y_{jt} is a measure of output in firm j and period t , K_{jt} is a measure of capital stock, L_{jt} is the total number of employees in the firm, h_{jt} is a measure of the stock of human capital per employee 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, μ_j captures time-invariant firm heterogeneity and ε_{jt} captures time-varying firm specific productivity shocks.

The estimation of production functions is a difficult exercise because inputs are chosen endogenously by the firm and because many inputs are unobserved. Even though the inclusion of firm time invariant effects may mitigate these problems (e.g., Griliches and Mairesse, 1995), this will not suffice if, for example, transitory productivity shocks determine the de-

¹²Most of the papers estimating the effect of training on productivity assume that output is either log-linear in training (e.g. Barron, Black and Lowenstein, 1989, and Black and Lynch, 1998), or semi log-linear in training (e.g. Bartel, 1991, 1994, Dearden, Reed and Van Reenen, 2000, and Ramirez, 1994).

cision of providing training (and the choice of other inputs). Recently, several methods have been proposed for the estimation of production functions, such as Olley and Pakes (1996), Levinsohn and Petrin (2000), Akerberg, Caves and Frazer (2005) and Blundell and Bond (2000). In this paper we implement the latter.¹³ We control for time invariant firm characteristics that are potentially correlated with the decision to invest in training (and with the choice of other inputs) by estimating the model in first differences. To account for the potential correlation between the stock of training and current productivity shocks we use past measures of training (and past measures of other inputs) to instrument for current training (and the current use of other inputs). We implement this procedure in a GMM setting using the approach developed in Blundell and Bond (1998, 2000) which builds in Arellano and Bond (1991) and Arellano and Bover (1995). Our instruments are valid if productivity shocks in first differences are not too much correlated over time and if we have enough lags when constructing the instruments. For example, if productivity shocks in first differences are an AR(1) process we can only use two or more lags of the endogenous variables as instruments. In our empirical work we test and reject that productivity shocks do not exhibit first order correlation in first differences and therefore we use instruments lagged two periods¹⁴. Due to the shortness of the panel neither can we use extra lags nor can we test for higher order serial correlation of shocks. One advantage of this approach is that it also corrects for biases generated by measurement error in inputs.

Following Blundell and Bond (2000) we assume that the productivity shocks in equation

¹³Dearden, Reed and Van Reenen (2005) use a similar approach to estimate the effects of training on productivity using with industry level data for the UK.

¹⁴First order autocorrelation can be due, for example, to measurement error in output (e.g., Blundell and Bond, 2000).

(3.2) follow an AR(1) process:

$$\varepsilon_{jt} = \rho\varepsilon_{jt-1} + \varphi_{jt} \quad (3.3)$$

where φ_{jt} is for now assumed to be an i.i.d. process and $0 < \rho < 1$. Taking logs from equation (3.2) and substituting yields the following common factor representation:

$$\begin{aligned} \ln Y_{jt} = & \ln A_t + \alpha \ln K_{jt} + \beta \ln L_{jt} + \gamma h_{jt} + \theta Z_{jt} + \mu_j + \varphi_{jt} \\ & + \rho \ln Y_{jt-1} - \rho \ln A_{t-1} - \rho\alpha \ln K_{jt-1} - \rho\beta \ln L_{jt-1} - \rho\gamma h_{jt-1} - \rho\theta Z_{jt-1} - \rho\mu_j. \end{aligned} \quad (3.4)$$

Grouping common terms we obtain the reduced form version of the model above.

$$\begin{aligned} \ln Y_{jt} = & \pi_0 + \pi_1 \ln K_{jt} + \pi_2 \ln L_{jt} + \pi_3 h_{jt} + \pi_4 Z_{jt} + \\ & + \pi_5 \ln Y_{jt-1} + \pi_6 \ln K_{jt-1} + \pi_7 \ln L_{jt-1} + \pi_8 h_{jt-1} + \pi_9 Z_{jt-1} + v_j + \varphi_{jt}. \end{aligned} \quad (3.5)$$

subject to the common factor restrictions (e.g., $\pi_6 = -\pi_5\pi_1, \pi_7 = -\pi_5\pi_2$), where $v_j = (1 - \rho)\mu_j$.

We start by estimating the unrestricted model in equation (3.4) and then impose (and test) the common factor restrictions using minimum distance (Chamberlain, 1984). Empirically, we measure Y_{jt} with the firm's value added, K_{jt} with book value of capital and L_{jt} with the total number of employees. Z_{jt} includes time varying firm and workforce characteristics - the proportion of males in the workforce, a cubic polynomial in the average age of the workforce, occupational distribution of the workforce and the average education of the workforce (measured by the proportion workers with high education) - as well as time, region and sector effects. h_{jt} will be computed for each firm-year using information on the training

history of each firm and making assumptions on the average knowledge depreciation.

We assume that average 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 type of knowledge depreciation affects the human capital of all the workforce 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, average human capital in the firm depreciates because each period new workers enter the firm without training while workers leave the firm, taking with them firm specific knowledge. Using the permanent inventory formula for the accumulation of human capital yields the following law of motion for human capital (abstracting from j):

$$H_{jt+1} = ((1 - \delta)h_{jt} + i_{jt})(L_{jt} - E_{jt}) + X_{jt}i_{jt}$$

where H_{jt} is total human capital in the firm in period t ($H_{jt} = L_{jt}h_{jt}$), X_{jt} is the number of new workers in period t , E_{jt} is the number of workers leaving the firm in period t and i_t is the amount of training per employee in period t .¹⁵ At the end of period t , the stock of human capital in the firm is given by the human capital of those $L_{jt} - E_{jt}$ workers that were in the firm in the beginning of the period t (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:

¹⁵We 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).

$$h_{jt+1} = (1 - \delta)h_{jt}\phi_{jt} + i_{jt} \quad (3.6)$$

where $\phi_{jt} = \frac{L_{jt} - E_{jt}}{L_{jt+1}}$ and $0 \leq \phi_{jt} \leq 1$.

Under these assumptions, skill depreciation in the model is given by $(1 - \delta)\phi_{jt}$. We assume that $\delta = 17\%$ per period (although we will examine the robustness of our findings to this assumption).¹⁶ We estimate the turnover rate from the data since we have information on the initial and end of the period workforce as well as on the number of workers who leave the firm (average turnover in the sample is 14%). The average skill depreciation in our sample is 25% per period. We measure i_{jt} with the average hours of training per employee in the firm.¹⁷ Since we cannot observe the initial stock of human capital in the firm (h_0), we face a problem of initial conditions. Under some restrictions the effect of h_0 on firm productivity can be subsumed in the firm fixed effect of equation (3.5).¹⁸

¹⁶Our choice of 17% is based on Lillard and Tan (1986), who estimate an average depreciation in the firm is between 15% and 20% per year. Alternatively, we could have estimated δ from the data. Our attempts to do so yielded very imprecise estimates.

¹⁷In approximately 3% of the firm-year observations we had missing information on training although we could observe it in the period before and after. To avoid losing this information, we assumed the average of the lead and lagged training values. This assumption is likely to have minor implications in the construction of the human capital variables because there were few of these cases.

¹⁸More precisely, we can write:

$$h_{jt} = (1 - \delta)^t \phi_{j1} \dots \phi_{jt-1} h_{j0} + \sum_{s=1}^{t-1} (1 - \delta)^{s-1} \phi_{jt-s} \dots \phi_{jt-1} i_{jt-s}$$

where h_{j0} is the firm's human capital the first period the firm is observed in the sample (unobservable in our data). Plugging this expression into the production function gives:

$$\ln Y_{jt} = \ln A_t + \alpha \ln K_{jt} + \beta \ln L_{jt} + \gamma \sum_{s=1}^{t-1} (1 - \delta)^{s-1} \phi_{jt-s} \dots \phi_{jt-1} i_{jt-s} + \theta Z_{jt} + \mu_{jt} + \varepsilon_{jt}$$

where $\mu_{jt} = \gamma(1 - \delta)^t \phi_{j1} \dots \phi_{jt-1} h_{j0}$. 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 $0 < \delta < 1$ and $0 < \phi_{jt} < 1$, then μ_{jt} depreciates every period at rate $(1 - \delta)\phi_{jt}$. If h_{j0} is correlated with the future sequence of i_{jt+s} then the production function estimates will be biased, and our instrumental variable strategy will not address this problem. However, it is possible to estimate h_0 by including in the production function a

We are interested in computing the internal rate of return of an additional hour of training per employee in the firm. From the estimates of the production function we can directly compute the *current* marginal product of training (MB_{t+1}). We assume that *future* marginal product of current training ($MB_{t+s, s \neq 1}$) is equal to current marginal product of training minus human capital depreciation. To obtain an estimate for the MFP_{jt} , we must compute the marginal product of one hour of work for each employee. Since our measure of labor input is the number of employees in the firm, we approximate the marginal product of an additional hour of work for all employees by $\frac{MPL_{jt}}{(\text{Hours per Employee}_{jt})}L_{jt}$ (where MPL_{jt} is the marginal product of an additional worker in firm j and period t).¹⁹ Finally, since part of the training occurs outside the normal working hours and our data set includes information on this share for each firm, we need to transform the marginal product of one hour of work into the marginal foregone cost of one hour of training. In our data, only 52% (on average) of the training hours takes place during normal working hours. To estimate marginal foregone productivity we multiply the marginal product of labor by this proportion for each firm.

3.2. The Costs of Training for the Firm

In the previous section we described how to obtain estimates of the marginal product of labor and, therefore, of the foregone productivity cost of training. Here we focus on the direct costs of training. To estimate MC_t , we need data on the direct cost of training. These include labor payments to teachers or training institutions, training equipment such as books

firm specific dummy variable whose coefficient decreases over time at a fixed and known rate $(1 - \delta)\phi_t$. This procedure is quite demanding in terms of computation and data, and in the present version of the paper we assume we can be reasonably approximate the terms involving h_0 with a firm fixed effect.

¹⁹Alternatively, we could have included per capita hours of work directly in the production function. Because there is little variation in this variable across firms and across time, our estimates were very imprecise.

or movies, and costs related to the depreciation of training equipment (including buildings and machinery). Such information is rarely available in firm level data sets. Our data is unusually rich for this exercise since it contains information on the duration of training, direct costs of training and training subsidies.

We model the direct cost function with a quadratic spline in the total hours of training provided by the firm to all employees, with three knots. The knots correspond to the 90th, 95th and 99th percentiles of the distribution of training hours. Our objective is to have a more flexible form at the extreme of the function where there is less data, to avoid the whole function from being driven by extreme observations. In particular, we consider:

$$C_{jt} = \theta_0 + \theta_1 I_{jt} + \theta_2 I_{jt}^2 + \theta_3 D_{1jt} (I_{jt} - k_1)^2 + \theta_4 D_{2jt} (I_{jt} - k_2)^2 + \theta_5 D_{3jt} (I_{jt} - k_3)^2 + \sum \sigma_s D_s + \eta_j + \xi_j \quad (3.7)$$

where C_{jt} is the direct cost of training, I_{jt} is the total hours of training, D_{zt} is a dummy variable that assumes the value one when $I_{jt} > k_z$ ($z = 1, 2, 3$), $k_1 = 15,945$, $k_2 = 32,854$, $k_3 = 125,251$, D_s are year dummies, η_j is a firm fixed effect and ξ_j is a time varying cost shock. We estimate the model using the Blundell and Bond (1998, 2000) system GMM estimator (first differencing eliminates η_j and instrumenting accounts for possible further endogeneity of I_{jt}). Empirically, C_{jt} is the direct cost supported by the firm (it differs from the total direct cost of training by the training subsidies), and I_{jt} is the total hours of training provided by the firm in period t .

From the above estimates we obtain $\frac{\partial C_{jt}}{\partial I_{jt}}$. To obtain the marginal direct costs of an

additional hour of training for all employees in the firm we compute $\frac{\partial C_{jt}}{\partial I_{jt}} L_{jt}$.

4. Empirical Results

Table 2 presents the estimated coefficients on labor and on the stock of training for alternative estimates of the production function.²⁰ Column (1) reports the ordinary least squares estimates of the log-linear version of equation (3.2), column (2) reports the first differences estimates of the log-linear version of equation (3.2) and column (3) reports the system-GMM estimates of equation (3.5) (Blundell and Bond, 2000). For the latter specification we report the coefficients after imposing the common factor restrictions. We also present the p-values for two tests for the latter specification: one is a test of the validity of the common factor restrictions, the other is an overidentification (Hansen) test. We can neither reject the overidentification restrictions nor the common factor restrictions.²¹ Our preferred estimates are in column (3) because they account for firm fixed effects and endogenous input choice. Columns (1) and (2) are presented for completeness.

The estimated benefits in all the columns of table 2 seem to be quite high: a increase in the amount of training per employee of 10 hours (approximately 0.5% of the total amount of hours worked in a year²²) leads to an increase in current value-added between 0.6% and 1.3%. As far as they can be compared, this estimate is in line with (and if anything is smaller than) other estimates of the benefits of training in the literature (e.g., Dearden, Reed and Van Reenen, 2005, Blundell, Dearden and Meghir, 1999). If the marginal product of labor

²⁰The estimated coefficients for full set of variables included in the regression are presented in table A1 in the annex.

²¹We estimate the model using the `xtabond2` command for STATA, developed by Roodman (2005).

²²For an individual working 2000 hours a year, 10 hours corresponds to 0.5% of annual working hours.

were constant (linear technology), an increase in the amount of training per employee by 10 hours would translate into foregone productivity costs of at most 0.5% of output (if all training occurred during working hours). With decreasing marginal product of labor (and because roughly 50% of training occurs outside normal working time) foregone productivity is much lower. Given that the impact of training on productivity lasts for more than just one period, ignoring direct costs would lead us to implausibly large estimates of the return to training. As explained in the previous section, we use the coefficient on labor input in table 2 to quantify the importance of foregone productivity costs of training for each firm.

The results of estimating the direct training cost function in equation (3.7) are reported in table 3. Again, for completeness we report the estimates for different methods. Column (1) estimates the equation in levels with ordinary least squares, column (2) estimates the equation in first differences with least squares and column (3) estimates equation with system-GMM. The latter are again our preferred estimates since they account for firm fixed effects and for the correlation between training and transitory cost shocks.

On average, foregone productivity accounts for less than 25% of the total costs of training. This finding is of great potential interest for two related reasons. First, it shows that a simple returns to schooling intuition is inadequate for studying the returns to training. In particular, it is unlikely that we can just read the return to training from the coefficient on training in a production function.²³ Second, without data on direct costs estimates of the return to investments in training lack some credibility given that direct costs account for the majority of training costs. Unfortunately it is impossible to assess the extent to which this result is

²³As emphasized in Mincer (1989), this is likely to also be a problem in wage regressions.

generalizable to other datasets (in other countries) because similar data is rarely available.

Table 4 presents the estimates of the internal rate of return of an extra hour of training for per employee for an average firm in our sample, the average return for firms providing training and the average return for firms not providing training.²⁴ In columns (1)-(5) we display the sensitivity of our results to different assumptions about the rate of human capital depreciation. The production function estimates underlying this table are reported in table A2. They differ across columns because the construction of the human capital measure depends on the rate of skill depreciation we assume (as explained in section 3.1). In our base specification, where we assume a 17% depreciation rate, the average marginal internal rate of return is 9% for the whole sample. The average return is negative (-7%) for firms not providing training and quite high (24%) for the set of firms offering training.

The negative returns for firms not providing training is a constant feature across the columns in this table. We conjecture that these firms do not offer training 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 17% and our preferred estimate being 24% (ignoring the estimates where we assume a 100% depreciation rate). With such high returns, it is puzzling why firms train such a small proportion of the total hours of work (less than 1%²⁵). One hypothesis is that suboptimal amounts of training may be the result of a coordination problem, as emphasized in Pischke (2005). Given that the benefits of training need to be shared between firms and workers, each party individually

²⁴In this paper heterogeneity in returns across firms does not come from a random coefficients specification, but from non-linearity in training and labor input in the production and cost functions.

²⁵From table 1 we can see that, in firms providing high amounts of training, hours trained per employee per year are on average 19, while hours worked per employee per year are above 1800.

only sees part of the total benefit of training. Unless investment decisions are coordinated and decided jointly, inefficient levels of investment may arise. Furthermore, information problems and uncertainty in this investment in human capital may lead firms to invest small amounts in training even though the *ex post* average return is high. Even though under our current set up we cannot distinguish how much of the variability in returns across firms is due to heterogeneity and how much is due to uncertainty (as in, for example, Carneiro, Hansen and Heckman, 2003), we find an enormous dispersion in the *ex post* returns to training which may be suggestive of the importance of uncertainty. For example, in our base specification the 5th percentile of the distribution of internal rates of return is -16% and the 95% percentile is 66%. Finally, it is possible that firms would like to invest more in their workers but they are unable to do so because they are constrained (e.g., credit constrained). In that case, investments in training are likely to be suboptimal. Unfortunately we cannot verify empirically the importance of each of these different hypotheses.

5. Conclusion

In this paper we estimate the internal rate of return of firm investments in human capital. We use a census of large manufacturing firms in Portugal between 1995 and 1999 with unusually detailed information on investments in training, its costs, and several firm characteristics. Our parameter of interest is the return to training for employers and employees as a whole, irrespective of how these returns are shared between these two parties.

We document the empirical importance of adequately accounting for the costs of training when computing the return to firm investments in human capital. In particular, unlike

schooling, direct costs of training account for about 75% of the total costs of training (foregone productivity only accounts for 25%). Therefore, it is not possible to read the return to firm investments in human capital from the coefficient on training in a regression of productivity on training. Data on direct costs is essential for computing meaningful estimates of the internal rate of return to these investments.

our estimates of the internal rate of return to training vary across firms. While investments in human capital have on average negative returns for those firms which do not provide training, we estimate that the returns for firms providing training are quite high, our lower bound being of 17% and our preferred estimate being 24%. Such high returns suggest that company job training is a sound investment for firms and for the economy as a whole, possibly yielding higher returns than either investments in physical capital or investments in schooling. Therefore, it is puzzling why these firms train on average such a small proportion of the total hours of work (less than 1%). We suggest three possible explanations: 1) coordination failures between employers and employees; 2) uncertainty in the returns to training; 3) credit constraints. Unfortunately we cannot assess the empirical importance (if any) of each of these hypotheses.

References

- [1] Acemoglu, D. and J. Pischke, 1998, "Why Do Firms Train? Theory and Evidence", *Quarterly Journal of Economics*, 113.
- [2] - -, 1999, "The Structure of Wages and Investment in General Training", *Journal of Political Economy*, 107.
- [3] Akerberg, D., K. Caves and G. Frazer, 2005, "Structural Estimation of Production Functions", UCLA Working Paper.
- [4] Alba-Ramirez, A., 1994, "Formal Training, Temporary Contracts, Productivity and Wages in Spain", *Oxford Bulletin of Economics and Statistics*, vol. 56.
- [5] Arulampalam, W., A. Booth and M. Bryan, 2004, "Training in Europe". *Journal of the European Economic Association*, April-May, 2.
- [6] Arulampalam, W., A. Booth and P. Elias, 1997, "Work-related Training and Earnings Growth for Young Men in Britain", *Research in Labor Economics*, 16.
- [7] Arellano, M. and P. Bond, 1991, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", *Review of Economic Studies*, 58.
- [8] Arellano, M. and O. Bover, 1995, "Another Look at the Instrumental-Variable Estimation of Error-Components Models", *Journal of Econometrics*, 68.
- [9] Barron, J., D. Black and M. Lowenstein, 1989, "Job Matching and On-The-Job Training", *Journal of Labor Economics*, vol 7.
- [10] Bassanini, A., A. Booth, M. De Paola and E. Leuven, 2005, "Workplace Training in Europe", IZA Discussion Paper 1640.
- [11] 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.
- [12] Black, S. and L. Lynch, 1997, "How to Compete: The Impact of Workplace Practices and Information Technology on Productivity", National Bureau Economic Research Working Paper No. 6120.
- [13] - -, 1998, "Beyond the Incidence of Training: Evidence from a National Employers Survey", *Industrial and Labor Relations Review*, Vol.52, no.1.
- [14] 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 Ritzén, Springer-Verlag.
- [15] - -, 1994, "Productivity Gains From the Implementation of Employee Training Programs", *Industrial Relations*, vol. 33, no. 4.

- [16] - -, 1995, "Training, Wage Growth, and Job Performance: Evidence from a Company Database", *Journal of Labor Economics*, Vol. 13, No. 3.
- [17] - -, 2000, "Measuring the Employer's Return on Investments in Training: evidence from the Literature", *Industrial Relations*, 39(3).
- [18] Barrett, A. and P. O'Connell, 2001, "Does Training Generally Work? The Returns to In-Company Training", *Industrial and Labor Relations Review*, 54 (3).
- [19] Blundell, R. and S. Bond, 1998, "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics* 87.
- [20] Blundell, R. and S. Bond, 2000, "GMM Estimation with Persistent Panel Data: An Application to Production Functions", *Econometric Reviews*, 19.
- [21] Blundell, R., L. Dearden and C. Meghir, 1996, "Work-Related Training and Earnings", Institute of Fiscal Studies.
- [22] Booth, A., 1991, "Job-related formal training: who receives it and what is it worth?", *Oxford Bulletin of Economics and Statistics*, vol. 53.
- [23] Carneiro, P., K. Hansen and J. Heckman, 2003, "Estimating Distributions of Counterfactuals with an Application to the Returns to Schooling and Measurement of the Effect of Uncertainty on Schooling Choice", *International Economic Review*, 44, 2.
- [24] Chamberlain, G., 1984, "Panel Data", Handbook of Econometrics, eds. Z. Griliches and M. Intriligator, Vol. 2.
- [25] Dearden, L., H. Reed and J. Van Reenen, 2005, "Who gains when workers train? Training and corporate productivity in a Panel of British Industries", *Oxford Bulletin of Economics and Statistics*, forthcoming.
- [26] Frazis, H. and G. Lowenstein, 2005, "Reexamining the Returns to Training: Functional Form, Magnitude and Interpretation", *The Journal of Human Resources*, XL, 2.
- [27] Griliches, Z. and J. Mairesse, 1995, "Production Functions: The Search for Identification", NBER wp 5067.
- [28] Leuven, E., 2004, "The Economics of Private Sector Training", *Journal of Economic Surveys*, forthcoming.
- [29] Leuven, E. and H. Oosterbek, 2004, "Evaluating the Effect of Tax Deductions on Training", *Journal of Labor Economics*, Vol. 22, No. 2.
- [30] - -, 2005, "An alternative approach to estimate the wage returns to private sector training", working paper.
- [31] Levinsohn, J. and A. Petrin, 2000, "Estimating production functions using inputs to control for unobservables". NBER wp 7819.

- [32] Lillard, L. and H. Tan, 1986, "Training: Who Gets It and What Are Its Effects on Employment and Earnings?", RAND Corporation, Santa Monica California.
- [33] Machin, S. and A. Vignoles, 2001, "The economic benefits of training to the individual, the firm and the economy", mimeo, Center for the Economics of Education, UK.
- [34] Mincer, J., 1989, "Job Training: Costs, Returns and Wage Profiles", NBER wp 3208.
- [35] Pischke, J., 2005, "Comments on "Workplace Training in Europe" by Bassanini et al.", working paper, LSE.
- [36] Olley, S. and A. Pakes, 1996, "The dynamics of productivity in the telecommunications equipment industry", *Econometrica*, 64.
- [37] Roodman, D., 2005, "Xtabond2: Stata module to extend xtabond dynamic panel data estimator", Statistical Software Components, Boston College Department of Economics.
- [38] Stevens, M., 1994, "A Theoretical Model of On-the-Job Training with Imperfect Competition", *Oxford Economics Papers*, 46.

Table 1: Medians of Main Variables by Training Intensity

	No Training Firms	Low Training Firms	High Training Firms
Value added / Employees	141,143	17,720	26,000
Employees	157	176	308
Hours work / Employees	2,043	2,044	2,056
Book Value Capital Depreciation	248,035	595,765	1,562,635
Share high educated workers	0.01	0.03	0.06
Average age workforce	37	39	41
Share males workforce	0.4	0.6	0.7
Occupations:			
Share top managers	0.01	0.02	0.03
Share managers	0.02	0.02	0.04
Share intermediary workers	0.04	0.05	0.05
Share qualified workers	0.41	0.42	0.43
Share semi-qualified workers	0.20	0.20	0.21
Share non-qualified workers	0.04	0.05	0.03
Share apprentices	0.03	0.02	0.001
Training hours / Employees	-	1.66	19.0
Training hours / Hours work	-	0.0008	0.009
Direct Cost / Employee	-	10	85
Direct Cost / Value Added	-	0.001	0.003
Nb observations	2,578	1,461	1,462

Source: *Balanço Social*

Nominal variables in Euros (1995 values). "Low training firms" are firms with at most the median annual hours of training (1,489) and "High training firms" are firms with at least the median annual hours of training. Employees is the total number of employees in the firm. Total Hours/Employees is annual hours of work per employee, Capital's Depreciation is the capital's book value of depreciation, "Share low educated workers" is the share of workers with at most primary education, Average age is the average age of the workforce (years), Share males is the share of males in the workforce, Training hours/Employee is the annual training hours per employee in the firm, Training hours / Hours work is the share training hours in total hours at work, Direct Cost/Employee is the cost of training per employee and Direct Cost / Value Added is the cost of training as a share of value added.

Table 2: Production Function Estimates

Dependent variable:	Log Real Value Added per Employee	Log Real Value Added per Employee	Log Real Value Added per Employee
Method:	OLS- Levels (1)	OLS-First Differences (2)	SYS-GMM (3)
Training Stock	0.0006 (0.0002)***	0.0013 (0.0002)***	0.0006 (0.0003)*
Log Employees	0.79 (0.01)***	0.56 (0.057)***	0.77 (0.11)***
Observations	4,327	2,816	2,816
P-Value Over-Identification Test	-	-	0.26
P-Value Common Factor Restrictions	-	-	0.52

The table presents estimates of the production function assuming that (time invariant) human capital depreciation in the firm is 17%.

Column (1) presents the estimates with ordinary least squares, column (2) with first differences and column (3) with SYS-GMM.

Standard errors in parenthesis, *** Significant at 1%, ** Significant at 5%, * Significant at 10%. All specifications include the following variables (point estimates not reported): log capital stock, share occupation group, share low educated workers, share males workforce, cubic polynomial in average age workforce, year dummies, region dummies and 2-digit sector dummies.

Table 3: Estimates of the Cost Function

Dependent variable:	Real Training Cost	Real Training Cost	Real Training Cost
Method:	OLS- Levels (1)	OLS-First Differences (2)	SYS-GMM (3)
Training Hours/1000	2046.3 (227.0)***	901.6 (331.1)***	7107.3 (4693.7)
(Training Hours/1000)^2	-57.8 (12.9)***	-19.1 (16.7)***	-223.2 (248.8425)
D1*(Training Hours/1000 -16)^2	115.3 (20.5)***	39.8 (25.1)	187.1 (339.0)
D2*(Training Hours/1000 -33)^2	-68.9 (8.7)***	-27.4 (9.8)***	53.0 (99.4)
D3*(Training Hours/1000 -125)^2	11.6 (.61)***	7.1 (0.68)***	-21.9 (8.9)**
Observations	5,511	3,908	5,511
P-Value F-test all slopes=0	0.00	0.00	0.00

The table presents the estimates of the cost function. Column (1) presents the estimates with ordinary least squares, column (2) with first differences and column (3) with SYS-GMM. Standard errors in parenthesis, *** Significant at 1%, ** Significant at 5%, * Significant at 10%. D1 is a dummy variable equal to 1 when total annual training hours in the firm is higher than 15,000, D2 is a dummy variable equal to 1 when total annual training hours in the firm is higher than 33,000 and D3 is a dummy variable equal to 1 when total annual training hours in the firm is higher than 125,000.

Table 4: Marginal Return of a Training Hour for All Employees

Depreciation Rate:	5%	10%	17%	25%	100%
	(1)	(2)	(3)	(4)	(5)
All Firms in Sample	14%	10%	9%	1%	-28%
Firms not providing training	0%	-4%	-7%	-14%	-64%
Firms providing training	27%	22%	24%	17%	4%

Table reports the average marginal internal rate of return for different assumptions on the (time invariant)

human capital depreciation in the firm. Marginal benefits and marginal costs were obtained with the SYS-

GMM estimates in columns (3) of table 2 and column (3) of table 4, respectively.

Table A1: Production Function Estimates

Dependent variable:	Log Real Value Added per Employee	Log Real Value Added per Employee
	SYS-GMM Unrestricted Common Factors	SYS-GMM Restricted Common Factors
Value Added per Employee -1	0.243 [0.113]**	-
Training Stock t	0.001 [0.001]	0.0006 [0.0003]*
Training Stock $t-1$	-0.001 [0.001]	-
Log Employees t	0.734 [0.241]***	0.7718 [0.117]***
Log Employees $t-1$	-0.149 [0.242]	-
Log Capital Stock t	0.132 [0.120]	0.2476 [0.045]***
Log Capital Stock $t-1$	0.06 [0.111]	-
Occupations:		
Share top managers t	5.0660 [6.131]	3.7722 [3.102]*
Share top managers $t-1$	-2.5100 [5.017]	-
Share managers t	4.2640 [6.654]	4.9432 [2.87]*
Share managers $t-1$	-1.2060 [5.179]	-
Share intermediary workers t	5.0550 [7.091]	5.9298 [3.110]*
Share intermediary workers $t-1$	-1.0770 [5.411]	-
Share qualified workers t	4.5500 [6.612]	5.0089 [2.877]*
Share qualified workers $t-1$	-1.2810 [5.227]	-
Share semi-qualified workers t	4.2190 [6.666]	4.828 [2.881]*
Share semi-qualified workers $t-1$	-1.1040 [5.272]	-
Share non-qualified workers t	3.8750 [6.365]	4.8915 [2.879]*
Share non-qualified workers $t-1$	-0.9260 [5.079]	-
Share apprentices t	3.2520 [6.329]	4.8873 [2.920]*
Share apprentices $t-1$	-0.1990 [4.986]	-
Share High Educated workers t	1.4930 [1.161]	2.3461 [0.561]***
Share High Educated workers $t-1$	0.1220 [0.414]	-
Share males workforce t	-1.09 [1.375]	0.8308 [0.331]***
Share males workforce $t-1$	1.772 [1.320]	-
Observations	2,816	2,816
Autocorrelation Coefficient	-	0.1256 [0.057]***

Columns (1) and (2) present the estimates of equation (3.3) and (3.4) in the text, respectively, with SYS-GMM, assuming that (time invariant) human capital depreciation in the firm is 17%. Standard errors in parenthesis, *** Significant at 1%, ** Significant at 5%, * Significant at 10%. The regressions also include year, region, sector dummies and a cubic polynomial on average age workforce.

Table A2: Production Function Estimates: Sensitivity to Different Depreciation Rates

Dependent variable:	Log Real Value Added per Employee				
Depreciation Rate:	5%	10%	17%	25%	100%
	(1)	(2)	(3)	(4)	(5)
Training Stock	0.0005 (0.0003)*	0.0005 (0.0003)*	0.0006 (0.0003)*	0.0006 (0.0003)*	0.0013 (0.0008)
Log Employees	0.75 (0.11)***	0.76 (0.11)***	0.77 (0.11)***	0.78 (0.12)***	0.85 (0.14)***
Observations	2,816	2,816	2,816	2,816	2,816
P-Value Over-Identification Test	0.26	0.26	0.26	0.26	0.33
P-Value Common Factor Restrictions	0.54	0.51	0.52	0.54	0.42

The table presents the SYS-GMM estimates of equation (3.4) in the text for different assumptions on the (time invariant) human capital depreciation in the firm. Standard errors in parenthesis, *** Significant at 1%, ** Significant at 5%, * Significant at 10%. All specifications include the following variables (point estimates not reported): capital stock, share occupation group, share low educated workers, share males workforce, cubic polynomial in average age, year dummies, region dummies and 2-digit sector dummies.