

# The Return to Firm Investments in Human Capital

Rita Almeida\*  
The World Bank

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

June 13, 2008

## 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 detailed information on the duration of training, the direct costs of training, and several firm characteristics. Our estimates of the return to training are substantial (8.6%) for those providing training. Results suggest that formal job training is a good investment for these firms possibly yielding comparable returns to either investments in physical capital or investments in schooling.

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

**JEL classification codes:** C23, D24, J31.

---

\*We are grateful to the Editor and two anonymous referees for valuable comments which significantly improved the paper. 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), the 2005 Econometric Society World Congress, and the 2006 Bank of Portugal Conference on Portuguese Economic Development. We thank especially the comments made by Manuel Arellano, Ana Rute Cardoso, Pedro Telhado Pereira and Steve Pischke. Carneiro gratefully acknowledges the support of the Economic and Social Research Council for the ESRC Centre for Microdata Methods and Practice (grant reference RES-589-28-0001), and the hospitality of Georgetown University, and of the Poverty Unit of the World Bank Research Group. Corresponding author: [ralmeida@worldbank.org](mailto: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, and more than one half of life-time 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 a modern economy, a firm cannot afford to neglect investments in the human capital of its workers. 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.<sup>1</sup> 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 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, observed between 1995 and 1999, with detailed information on investments in training, its costs, and several firm characteristics.<sup>2</sup>

Most of the empirical work to date has focused on 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 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).<sup>3</sup> More importantly, the literature 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,

---

<sup>1</sup>An 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.

<sup>2</sup>We 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.

<sup>3</sup>For 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.

Barret and O’Connell, 1999, Dearden, Reed and Van Reenen, 2005, Ballot, Fakhfakh and Taymaz, 2001, Conti, 2005). This happens most probably due to lack of adequate data. As a result, and as emphasized by Mincer (1989) and Machin and Vignoles (2001), we cannot interpret the estimates in these papers as well defined rates of return.

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, which can be estimated. We do not distinguish whether the costs and benefits of training accrue mainly to workers or to the firm. Instead, we quantify the internal rate of return to training jointly for firms and workers.<sup>4</sup> This implies that, to obtain estimates of the foregone opportunity cost of training we will not take into account whether firms or workers support the costs of training.

The major challenge in this exercise are possible omitted variables and the endogenous choice of inputs in the production and cost functions. Given the panel structure of our data, we address these issues using the estimation methods proposed in Blundell and Bond (2000). In particular, we estimate the cost and production functions using a first difference instrumental variable approach, implemented with a system-GMM estimator. 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 input decisions in period  $t - 1$  are made without knowledge of the transitory shocks in the production and cost functions from period  $t + 1$  onwards.<sup>5</sup>

---

<sup>4</sup>Dearden, Reed and Van Reenen (2006) and Conti (2005) estimate the differential effect of training on productivity and wages. The former find that training increases productivity by twice as much as it increase wages, while the latter finds only effects of training on productivity (none on wages).

<sup>5</sup>This assumption is valid as long as there does not exist strong serial correlation in the transitory schocks

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, Frazer and Lowenstein, 2005, Ballot, Fakhfakh and Taymaz, 2001, Conti, 2005) our estimates of the effects of training on productivity are high: an increase in 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. 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).<sup>6</sup> Given this wedge between the benefits and the foregone output costs of training, ignoring the direct costs of training is likely to yield a rate of return to training that is absurdly high (unless the marginal product of labor function is convex, so that the marginal product exceeds the average product of labor).

Second, we estimate that, on average, foregone productivity accounts for less than 25% of the total costs of training. This finding shows that the simple returns to schooling intuition is inadequate for studying the returns to training, since it assumes negligible direct costs of human capital accumulation. In particular, the coefficient on training in a production function (or in a wage equation) is unlikely to be a good estimate of the return to training. Moreover, 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 (see also the calculations in Frazer and Lowenstein, 2005).

Our estimates indicate that, while investments in human capital have on average zero returns for training for all the firms in the sample, the returns for firms providing training are

---

in the data, and firms cannot forecast future shocks. Given the relatively short length of our panel our ability to test this assumption is limited. Dearden, Reed and Van Reenen (2005) apply an identical methodology (using industry level data for the UK) for a longer panel and cannot reject that second order serial correlation in the first differences of productivity shocks is equal to zero. In their original application, Blundell and Bond (2000) also do not find evidence of second order serial correlation using firm level data for the UK.

<sup>6</sup>For an individual working 2,000 hours a year, 10 hours corresponds to 0.5% of annual working hours.

quite high (8.6%). Such high returns suggest that on-the-job training is a good investment for firms that choose to undergo this investment, possibly yielding comparable returns to either investments in physical capital or investments in schooling.<sup>7</sup>

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 the census of large firms (more than 100 employees) operating in Portugal (Balanco Social). The information is collected with a mandatory annual survey conducted by the Portuguese Ministry of Employ. The data has information on hours of training provided by the 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 depreciation, 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

---

<sup>7</sup>As a consequence, it is puzzling why firms that choose to undergo this investment in training, train on average such a small proportion of the total hours of work (less than 1%). We conjecture that this could happen for different reasons but unfortunately we cannot verify empirically the importance of each of these hypotheses. First, it may be the result of a coordination problem (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. This may be also due to the so called "poaching externality" (Stevens, 1994). See also Acemoglu and Pischke (1998, 1999) for an analysis of the consequences of imperfect labor markets for firm provision of general training. Unless investment decisions are coordinated and decided jointly, inefficient levels of investment may arise. Second, firms can be constrained (e.g., credit constrained) and decide a suboptimal investment. Third, uncertainty in the returns of this investment may lead firms to invest small amounts even though the *ex post* average return is high, although what really matters for determining the risk premium is not uncertainty per se, but its correlation with the rest of the market. However, it is unlikely that uncertainty alone can justify such high rates of return. In our model uncertainty only comes from future productivity shocks, since current costs and productivity shocks are assumed to be known at the time of the training decision. The R-Squared of our production functions (after accounting for firm fixed effects) is about 85%, suggesting that temporary productivity shocks explain 15% of the variation in output. Since productivity shocks are correlated over time this is an overestimate for the uncertainty faced by firms.

(corresponding to 5,501 firm-year observations). On average, 53% of the firms in the sample provide some training. All the variables used in the analysis are defined in the appendix.

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.<sup>8</sup> Fourth, it collects longitudinal information for training hours, firm productivity and direct training costs at the firm level. Approximately 75% of the firms are observed for 3 or more years and more than 60% of the firms are observed for 4 or more years. For approximately 50% of the firms there is information for the 5 years between 1995-1999.<sup>9</sup>

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 training hours per employee are above the median (6.4 hours) for the firms that provide training. We report medians rather than means to avoid extreme sensitivity to extreme values. Firms that offer training programs and are defined as high training intensity firms 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

---

<sup>8</sup>Bartel (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. Ballot, Fakhfakh and Taymaz (2001) use information for 90 firms in France between 1981–1993 and 250 firms in Sweden between 1987–1993. One exception is Conti (2005). She uses a large panel of Italian firms between 1996-1999 but the analysis is done at the more aggregated industry level.

<sup>9</sup>Firms can leave the sample because they exit the market or because total employment is reduced to less than 100 employees.

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 1.3% in non-training firms. The workforce in firms that offer training has a higher proportion of male workers.<sup>10</sup> 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 annual hours on-the-job for the (0.9% of total time on-the-job). High training firms spend 9 times more in training per employee than low training firms. These costs are 0.01% 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 Southern 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 between firms offering training, with low and high training firms being very different. Finally, the direct costs of formal training programs are small (as a proportion of the firm's value added) 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. This is the relevant parameter for evaluating the rationale for additional investments in training, since firms compare the returns to alternative investments at the margin. Let  $MB_{t+s}$  be the marginal benefit of an additional unit of training in  $t$  and

---

<sup>10</sup> Arulampalam, 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.

$MC_t$  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, as in so much of the literature, that the firm's production function is semi-log linear and that the firm's stock of human capital determines the current level of output:

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

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,  $\mu_j$  captures time-invariant firm heterogeneity and  $\varepsilon_{jt}$  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 decision 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).

We apply the methods for estimation of production functions proposed 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 (essentially) a first difference instrumental variable approach, implemented with a GMM estimator. By computing first differences we control for firm unobservable and time invariant characteristics (much of the literature generally stops here). 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 transitory shocks in the production and cost functions are unknown two or more periods in advance. Bond and Soderbom (2005) provide a rationale for this procedure, which is based on the existence of factor adjustment costs. An alternative procedure could be based differences in input prices across firms (if they existed) such as, for example, training subsidies which apply to firm A but not firm B in an exogenous way, but these are unobserved in our data.

Given the evidence in Blundell and Bond (2000), we assume that the productivity shocks in equation (3.2) follow an AR(1) process:

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

where  $\varphi_{jt}$  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 a minimum distance estimator (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.

Since the model is estimated in first differences the assumption we need is  $E[(\varphi_{jt} - \varphi_{jt-1})X_{jt-2}] = 0$ , where  $X$  is any of the inputs we consider in our production function. Therefore, we allow the choice of inputs at  $t$ ,  $X_{jt}$ , to be correlated with current productivity shocks  $\varepsilon_{jt}$ , and even with the future productivity shock  $\varepsilon_{jt+1}$ , as long it is uncorrelated with the innovation in the auto-regressive process in  $t + 1$ , i.e.  $\varphi_{jt+1}$ , i.e., these shocks are not anticipated. In this case, inputs dated  $t - 2$  or earlier can be used to as instruments for the first difference equation in  $t$  (similarly,  $Y_{jt-1}$  can be instrumented with  $Y_{jt-3}$  or earlier).

Blundell and Bond (1998) point out that it is possible that these instruments are weak, and it may be useful to supplement this set of moment conditions with additional ones provided that  $E[(X_{jt-1} - X_{jt-2})(v_j + \varphi_{jt})] = 0$ , which is satisfied if  $E[(X_{jt-1} - X_{jt-2})v_j] = 0$ . When can this assumption be justified? Here we reproduce the discussion in Blundell and Bond (2000), which is as follows. Suppose we have the following model:

$$y_{it} = \alpha Y_{it-1} + \beta x_{it} + (\eta_i + e_{it}),$$

where  $y$  is output,  $x$  is input,  $\eta_i$  is the firm fixed effect, and  $e_{it}$  is the time varying productivity shock. Suppose further that  $x$  follows an AR(1) process:

$$x_{it} = \gamma x_{it-1} + (\delta\eta_i + u_{it}).$$

The absolute values of  $\alpha$  and  $\gamma$  are assumed to be below 1. After repeated substitution and first differencing of this equation, we obtain:

$$\Delta x_{it} = \gamma^{t-2} \Delta x_{i2} + \sum_{s=0}^{t-2} \gamma^s \Delta u_{it-s}.$$

Therefore, one way to justify  $E(\Delta x_{it} \eta_i) = 0$  would be to say that  $E(\Delta x_{i2} \eta_i) = 0$ . This, however, may be a quite unappealing assumption, since firms with a larger fixed effect may grow faster, especially in their early years. Instead, we assume that  $t$  is large enough for the firm to be in steady state, and the role of  $\Delta x_{i2}$  to disappear. In steady state, it is plausible to assume that the growth rate of the firm depends on the growth rate of productivity, rather than on the level of productivity. Actually, at least in the five years covered by our sample, firms do not seem to be on a path of sustained growth. Indeed, regressing current firm growth on past growth yields a negative coefficient, indicating that a year of firm growth is generally followed by a year of decline.<sup>11</sup>

The evidence in section 4 will show that using only the first set of instruments will raise problems of weak instruments in our sample. Therefore, we will use system-GMM in our preferred specification and will report the Sargan-Hansen test of overidentifying restrictions.<sup>12</sup>

In general, given the instrumental variables estimates of the coefficients, it is possible to test whether the first difference of the errors are serially correlated. Unfortunately, given the short length of the panel, we can only test for first order serial correlation of the residuals, which we reject almost by construction (since a series of first differences is very likely to exhibit first order serial correlation). The hypothesis that there exists higher order serial correlation (which would probably invalidate our procedure) is untestable in our data.<sup>13</sup>

---

<sup>11</sup> Available from the authors upon request.

<sup>12</sup> This approach has been implemented by others in the literature (e.g., Dearden, Reed and Van Reenen, 2005, Ballot, Fakhfakh and Taymaz, 2001, Zwick, 2004, Conti, 2005).

<sup>13</sup> Although we have 1,500 firms in our sample, the effect of training on productivity is identified with only approximately 61% of the sample, for whom we have three or more observations. The remaining firms are used to identify other parameters in the model, for which we do not need to instrument (e.g., year effects). There are five years of data in our panel but we can use at most four years for each firm because we use lagged training as our main explanatory variable (the first year of data is used only to construct the training stock). With three years of data it is not possible to test for serial correlation in the errors (since three years is the minimum number of years needed to identify the model), while with four years of data we can only test for first order serial correlation.

Hopefully this is not a big concern. Dearden, Reed and Van Reenen (2005) apply an identical method to analyze the effect of training on productivity (using industry level data for the UK over a longer period) and cannot reject that second order serial correlation in the first differences of productivity shocks is equal to zero. In their original application, Blundell and Bond (2000) also do not find evidence of second order serial correlation using firm level data for the UK.

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 (e.g, Lillard and Tan, 1986). 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  $\delta$  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 (e.g., Ballot, Fakhfakh and Taymaz, 2001, Dearden, Reed and Van Reenen, 2005). 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  $t$ .<sup>14</sup> 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:

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

---

<sup>14</sup>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).

where  $\phi_{jt} = \frac{L_{jt} - E_{jt}}{L_{jt+1}}$  and  $0 \leq \phi_{jt} \leq 1$ . Our estimation procedure is robust to endogenous turnover rates since they can be subsumed as another dimension of the endogeneity of input choice.<sup>15</sup>

Under these assumptions, skill depreciation in the model is given by  $(1 - \delta)\phi_{jt}$ . We assume that  $\delta = 17\%$  per period in our base specification, although we will examine the sensitivity of our findings to this assumption. Our 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. This number is also close to the one used by Conti (2005) in her baseline specification (15%).<sup>16</sup> 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.<sup>17</sup>

---

<sup>15</sup>In 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.

<sup>16</sup>Alternatively, we could have estimated  $\delta$  from the data. Our attempts to do so yielded very imprecise estimates.

<sup>17</sup>Since we cannot observe the initial stock of human capital in the firm ( $h_0$ ), we face a problem of initial conditions. 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,  $\mu_{jt}$  becomes a firm fixed effect only 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  $\mu_{jt}$  depreciates every period at rate  $(1 - \delta)\phi_{jt}$ . If  $h_0$  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. Although it would be possible to estimate  $h_0$  by including in the production function a firm specific dummy variable whose coefficient decreases over time at a fixed and known rate  $(1 - \delta)\phi_t$ , this procedure would be quite demanding in terms of computation and data. For simplicity, we assume we can reasonably approximate the terms involving  $h_0$  with a firm fixed effect. This difficulty comes from trying to introduce some realism in the model through the consideration of stocks rather than flows of training, and the use of positive depreciation rates, both of which are sometimes ignored in the literature.

The semi-log linear production function we assume implies that human capital is complementary with other inputs in production ( $\frac{\partial^2 \ln Y}{\partial H \partial X} > 0$ , where  $X$  is any of the other inputs). However, we do not believe this is a restrictive assumption. In fact, it is quite intuitive that such complementarity exists since labor productivity and capital productivity are likely to be increasing functions of  $H$  (workers with higher levels of training make better use of their time, and make better use of the physical capital in the firm). The only concern would be that  $H$  and workers' schooling could be substitutes, not complements (workers' schooling is one the inputs in  $Z$ ). In this regard, most of the literature shows that workers with higher levels of education are more likely to engage in training activities than workers with low levels of education, indicating that, if anything, training and schooling are complements.

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 (ceteris paribus analysis: what would happens to future output keeping everything else constant, including the temporary productivity shock). 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  $t$ ).<sup>18</sup>

Given the concerns with functional form in the related wage literature, emphasized by Frazer and Lowenstein (2005), we estimated other specifications where we include polynomials in human capital in the production function. Since higher order terms were generally not significant we decided to focus our attention on our current specification.

---

<sup>18</sup>Alternatively, 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.

### 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 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.

Different firms face the same cost up to a level shift. We do not expect to see many differences in the marginal cost function across firms since training is probably acquired in the market (even if it is provided by the firm, it could be acquired in the market).<sup>19</sup> Therefore we model the direct cost function using levels of cost instead of log cost with a quadratic spline in the total hours of training provided by the firm to all employees, with several knots (using logs instead of levels gives us slightly lower marginal cost estimates). Initially we included a complete specification with knot points at the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th, percentiles of the distribution of (positive) training hours. However, in the estimation, the first six knot points systematically dropped from the specification due to strong collinearity (the distribution of training hours is fairly concentrated), and only the last three remained important. Therefore, in the final specification we include knots that correspond to the 90th, 95th and 99th percentiles of the distribution of training hours. Our objective with this functional form 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. This specification also makes it easier to capture potential fixed costs of training, that can vary across firms. In particular, we consider:

---

<sup>19</sup>Unfortunately, in our data we do not have any information on the content of the training programs that are offered in each firm. Still, we are fairly certain that the training measure captures hours of formal training (as opposed to informal training that occurs naturally on the job). We conjecture that the costs which the firm reports concern services that the firm can acquire in the market, or it would probably very difficult for a firm to quantify them.

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

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  $I_{jt} > k_z$  ( $z = 1, 2, 3$ ),  $k_1 = 15,945$ ,  $k_2 = 32,854$ ,  $k_3 = 125,251$  (90th, 95th and 99th percentiles of the distribution of training hours),  $D_s$  are year dummies,  $\eta_j$  is a firm fixed effect and  $\xi_j$  is a time varying cost shock.<sup>20</sup>

We estimate the model using the Blundell and Bond (1998, 2000) system GMM estimator (first differencing eliminates  $\eta_j$  and instrumenting accounts for possible further endogeneity of  $I_{jt}$ ). We described this method in detail already, and again we believe that the identifying assumptions are likely to be satisfied by the cost function. We assume that  $E[(I_{jt-1} - I_{jt-2})(\eta_j + \xi_{jt})] = 0$  and  $E[(\xi_{jt} - \xi_{jt-1})I_{jt-k}] = 0$ ,  $k \geq 3$ . We choose  $k \geq 3$  rather than  $k \geq 2$  to increase the chances that the assumptions above hold.<sup>21</sup> We do not reject the test of overidentifying restrictions, and therefore that is the specification we use. 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  $t$ .

One last aspect with respect to the cost function concerns the choice of not modelling the temporary cost shock as an autoregressive process, as it was done for the production function. In fact, we started with such a specification. However, when we estimated the model the autoregressive coefficient was not statistically different from zero, and therefore we chose a simpler specification for the error term.

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  $\frac{\partial C_{jt}}{\partial I_{jt}} L_{jt}$ .

---

<sup>20</sup>We also estimated another specification, where we trimmed all the observations for which total hours of training were above 15,945 (90% percentile). In doing so we removed extreme observations. We then estimated a quadratic cost function as in (3.7) (but without the know points). The resulting estimates of marginal costs came out smaller, resulting in larger returns. We come back to this below.

<sup>21</sup>In fact, if we assume the above assumptions hold for  $k \geq 2$  we reject the test of overidentifying restrictions.



## 4. Empirical Results

Table 2 presents the estimated coefficients on labor and on the stock of training for alternative estimates of the production function. 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). For the latter specification we report the coefficients after imposing the common factor restrictions.<sup>22</sup> 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-Sargan) test. We can neither reject the overidentification restrictions nor the common factor restrictions.<sup>23</sup> Our preferred estimates are in column (3) because they account for firm fixed effects and endogenous input choice. Table A2 in the appendix reports the equivalent to the first stage regressions (or the reduced form regressions) for the specification in column (3), using system-GMM, for the main endogeneous variables of interest (sales, employment, capital and training stock). The reduced form regression for the first-difference equations (reported in Panel A) relates, for a given input ( $X$ ),  $\Delta X_{t-1}$  to the lagged levels,  $X_{t-3}$  and  $X_{t-4}$ . The reduced form regression for the level equations (reported in Panel B) relate  $X_{t-1}$  to  $\Delta X_{t-3}$  and  $\Delta X_{t-4}$ . For the first difference equation, the instruments are jointly significant for sales, employment, capital though not for the stock of training. This explains why the differenced-GMM estimator performs poorly in our model and why we have a problem of weak instruments. For the level equation, the instruments are jointly significant for employment, capital and for the stock of training, though not for sales. Again, this helps explaining why the system-GMM estimator, which exploits both sets of moment conditions, works well for our final specification. Even though our initial sample has 5,511 observations (firm-year), we can only estimate the effect of training on productivity for a smaller sample. This happens because we use lagged training to construct the stock

---

<sup>22</sup>Table A1 in the appendix reports the estimated coefficients for the full set of variables included in the regression with system-GMM. Columns (1) and (2) present the unrestricted and restricted models, respectively.

<sup>23</sup>We estimate the model using the `xtabond2` command for STATA, developed by Roodman (2005).

of training (and the first observation for each firm is not used in estimation) and because our preferred specification of the production function is estimated in first differences (and we lose one further observation per firm).<sup>24</sup>

Columns (1) and (2) are presented for comparison. In particular, column (2) corresponds to the most commonly estimated model in this literature (using either wages or output as the dependent variable). The instrumental variables estimate of the effect of training on value added in column (3) is well below the estimate in column (2). This may happen because firms train more in response to higher productivity shocks, generating a positive correlation between temporary productivity shocks and investments in training. Curiously, Dearden, Reed and Van Reenen (2005) also find that the first difference estimate overestimates the effect of training on productivity, although the difference between first difference and GMM estimates in their paper is smaller than in ours.

The estimated benefits in all the columns of table 2 seem to be quite high, even the system-GMM estimate. An increase in the amount of training per employee of 10 hours (approximately 0.5% of the total amount of hours worked in a year<sup>25</sup>) leads to an increase in current value-added which is between 0.6% and 1.3%. As far as this number can be compared with other estimates of the effect of training on productivity in the literature, our estimate is, if anything, smaller. 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). 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 (unless the marginal product of labor function is convex, so that the marginal product exceeds the average product of labor). As explained in the previous section, we will use the coefficient on labor input in column (3) of table 2 to quantify the importance of foregone productivity costs of training for each firm.

---

<sup>24</sup>However, it is reassuring that the results obtained using OLS on the sample of firms that is reported in columns (2) and (3) of table 2 would yield similar findings to the ones reported in column (1) of the same table.

<sup>25</sup>For an individual working 2,000 hours a year, 10 hours corresponds to 0.5% of annual working hours.

The results of estimating the direct training cost function in equation (3.7) are reported in table 3. These estimates are based on a larger set of firms than the ones reported in table 2 because we use as explanatory variable the current training, not the lagged. In other words, in our specification current training affects current costs of training and lagged training affects current productivity. Again, for comparison, 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.<sup>26</sup> Regarding the latter, one specification that works well, both in terms of the strength of the first stage relationships, and in terms of non-rejection of over-identifying restrictions, takes variables lagged 3 periods to instrument the first differences of the endogenous variables, and first differences lagged 2 periods to instrument for the levels. Table A3 in the appendix reports the reduced form equation equivalent to the first stage when using system-GMM. The significance of the instruments for hours of training in both in Panel A and B, give us confidence on these estimates using the system-GMM methodology. We test and reject that all coefficients on training are (jointly) equal to zero. We also test whether second order correlation in the first differenced errors is zero and do not reject the null hypothesis. Similarly, we do not reject the test of overidentifying restrictions for the cost function (P value reported in table 3).<sup>27</sup>

We proceed to compute the marginal benefits and marginal costs of training for each firm. On average, we estimate that foregone productivity accounts for less than 25% of the total costs of training. This finding is of great 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.<sup>28</sup> The reason is that, unlike the case of schooling, direct

---

<sup>26</sup>It is reassuring to see that, the results obtained using OLS on the sample of firms that is reported in columns (2) and (3) of table 3 would yield similar findings to the ones reported in column (1) of the same table.

<sup>27</sup>For ease of interpretation of the regression coefficients, Figure 1 in the appendix reports the graphical representation of the marginal cost of training with the three alternative methodologies reported in table 3. We plot the marginal cost up to the 90th percentile of the distribution of training hours (equivalent to 16,000 hours of training in the firm).

<sup>28</sup>As emphasized in Mincer (1989), this is likely to also be a problem in wage regressions.

costs cannot be considered to be negligible. Second, without data on direct costs estimates of the return to investments in training are of limited use given that direct costs account for the majority of training costs. Unfortunately it is impossible to assess the extent to which this result is generalizable to other datasets (in other countries) because similar data is rarely available. However, given the absurd rates of return implicit in most of the literature when one ignores direct costs (e.g., Frazer and Lowenstein, 2005), we conjecture that a similar conclusion most hold for other countries as well.

Finally, table 4 presents the estimates of the internal rate of return (IRR) of an extra hour of training per employee for an average firm in our sample, and the average return for firms providing training.<sup>29</sup> The results of tables 2 and 3 assume a rate of human capital depreciation ( $\delta$ ) of 17%. In columns (1)-(5) we display the sensitivity of our IRR estimates to different assumptions about the rate of human capital depreciation (the production function estimates underlying this table are reported in table A4 in the appendix). In our base specification, where we assume a 17% depreciation rate, the average marginal internal rate of return is  $-0.3\%$  for the whole sample. However, the average return is quite high ( $8.6\%$ ) for the set of firms offering training. As expected, the higher the depreciation rate the lower is the estimated IRR. In particular, under the standard assumption that  $\delta = 100\%$  (so that the relevant input in the production function is the training flow, not its stock), the average IRR for the marginal unit of training is negative, independently of taking the sample as a whole or only the set of training firms. For reasonable rates of depreciation (which in our view are the ones in the first three columns of the table) returns to training are quite high for the sample of firms that decide to engage in training activities, our lower bound being of  $6.7\%$  and our preferred estimate being  $8.6\%$  (ignoring the estimates where we assume a 100% depreciation rate).<sup>30</sup>

---

<sup>29</sup>In 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. Of course, misspecification of the production or cost functions will affect these estimates. One important reason to report returns both for the average firm in the sample, and for the average firm providing training, is that we are more confident in our estimates of the marginal direct costs of training for the latter group of firms. The former group of firms are in a corner solution, and it is probably hard to estimate the cost function at 0 hours of training.

<sup>30</sup>The estimate goes up to  $12.8\%$  when we consider an alternative cost function where we trim all obser-

One criticism to our approach could be that depreciation rates could vary across firms, and we are only capturing this variation through heterogeneity in the turnover rate, and turnover is probably does not represent all heterogeneity in depreciation rates. For example, it would not capture the incidence of the maternity leave period on the workforce, unless the mother leaves the firm permanently. Moreover, it is possible that the rate of skill depreciation is correlated with training decisions, if firms with high rates of depreciation invest less in training. This problem is hard to address, since depreciation rates enter in two important places: the construction of training stocks, which are an input in the firm production function; and the computation of the future marginal benefits of an additional unit of training today. Take the case where depreciation rates are negatively correlated with training, because they reduce the firm’s incentives to invest. In this case the stock of training would be larger than we estimated it to be for those firms providing high amounts of training (since they would have low depreciation), and they would be lower than our estimates for firms providing little training (the opposite would happen if depreciation and training were positively correlated, which could be the case if firms with high levels of depreciation tried to overcompensate it by training more, or if firms with a high levels of training ended up with a many high skilled workers who would be very mobile in the labor market). In reality, this is almost as if we had a random coefficient in training in the production function (if we used our current measures of stock of training), and, as is well known, the IV estimates could become very hard to interpret in this case. Furthermore, the IV “bias” relatively to an average effect of training on output would be unpredictable. Still, suppose it was possible to get an unbiased estimate of the average benefit of training. We would still have the problem of allowing the schedule of marginal benefits across periods to be different across firms with different levels of depreciation. Again, if those firms providing training have the lowest depreciation rates, the variation in returns we estimate would be understated.

Another criticism is related to the possible complementarity between the average ability in the workforce and training. On the one end, firms whose workers have higher levels of ability

---

variations above the 90th percentile. We feel more confident about leaving all the data in and modelling the tails of the distribution of hours in a flexible way, but present this alternative estimate for completeness.

could engage in more training activities. On the other end, even within a firm, managers could provide training to the most able workers for whom the returns are the highest, and then worry about training for everyone else in the firm. Regarding the first concern, since our estimation strategy explores the variation in levels, we would be mainly worried about changes in training stocks that are correlated with changes in the unobserved skills of the workforce (given that all permanent effects should be handled by the fixed effect). The remaining changes in unobserved skills are treated as unforecastable productivity shocks and the instrumental variable strategy that we explore in the system-GMM methodology would address them. Nevertheless, the second concern is trickier. It implies that the effect of training varies across firms, because it would depend on the type of workers that are selected to undertake training in each firm. In this case, the instrumental variable approach would not address this concern and it is unclear exactly which parameter we would be estimating in such a case.

## 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 invest-

ments 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 substantial, our lower bound being of 6.7% and our preferred estimate being 8.6%. Such high returns suggest that company job training is a sound investment for firms that do train, possibly yielding comparable returns to either investments in physical capital or investments in schooling.

## 6. Data Appendix

The data used is the census of large firms conducted by the Portuguese Ministry of Employment in the period 1995-1998. We restrict the analysis to manufacturing firms. All the firms are uniquely identified with a code that allows us to trace them over time. This data collects information on balance sheet information, employment structure and training practices. All the nominal variables in the paper were converted to euros at 1995 prices using the general price index and the exchange rate published by the National Statistics Institute.

In the empirical work, we use information for each firm on total value added, book value of capital depreciation, total hours of work, total number of employees, total number of employees hired during the year, total number of employees that left the firm during the year (including quits, dismissals and deaths), average age of the workforce, total number of males in the workforce, total number of employees with bachelor or college degrees, total number of training hours, total costs of training, firm's regional location and firm 5-digit ISIC sector code.

We define value added as total value added in the firm, employees is the total number of employees at the end of the period, Hours work is the total hours of work in the firm (either working or training), Capital depreciation is the book value of capital depreciation <sup>31</sup>, Share of high educated workers is the share of workers with more than secondary education in the firm, Age of the workforce is the average age of all the employees in the firm, Share males in the workforce is the share of males in the total number of employees in the firm, Training hours per employee is the total number of hours of training provided by the firm (internal

---

<sup>31</sup>We assume that depreciation is a linear function of the book value of the firm's capital stock :  $Dep_t = \pi * K_t$ .

or external) divided by the total number of employees, Training hours per working hour is the total number of training hours provided by the firm (internal or external) divided by the total hours of work in the firm, Direct cost per employee is the total training cost supported by the firm (include, among others, the wages paid to the trainees or training institutes and the training equipment, including books and machinery) divided by the total number of employees, Average worker turnover is the total number of workers that enter and leave the firm divided by the average number of workers in the firm during the year, Average number of workers in the firm during the year is the total number of workers in the beginning of the period plus the total number of workers at the end of the period divided by two.



## 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] Ballot, G. and F. Fakhfakh and E. Taymaz, 2001, "Firms' human capital, R&D and performance: a study on French and Swedish firms", *Labour Economics*, vol. 8(4), pages 443-462.
- [10] Barron, J., D. Black and M. Lowenstein, 1989, "Job Matching and On-The-Job Training", *Journal of Labor Economics*, vol 7.
- [11] Bassanini, A., A. Booth, M. De Paola and E. Leuven, 2005, "Workplace Training in Europe", IZA Discussion Paper 1640.
- [12] 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.
- [13] 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.
- [14] - -, 1998, "Beyond the Incidence of Training: Evidence from a National Employers Survey", *Industrial and Labor Relations Review*, Vol.52, no.1.
- [15] Bartel, A., 1991, "Formal Employee Training Programs and Their Impact on Labor Productivity: Evidence from a Human Resources Survey", *Market Failure in Training? New Economic Analysis and Evidence on Training of Adult Employees*, ed. David Stern and Jozef Ritzen, Springer-Verlag.

- [16] - -, 1994, "Productivity Gains From the Implementation of Employee Training Programs", *Industrial Relations*, vol. 33, no. 4.
- [17] - -, 1995, "Training, Wage Growth, and Job Performance: Evidence from a Company Database", *Journal of Labor Economics*, Vol. 13, No. 3.
- [18] - -, 2000, "Measuring the Employer's Return on Investments in Training: evidence from the Literature", *Industrial Relations*, 39(3).
- [19] Barrett, A. and P. O'Connell, 2001, "Does Training Generally Work? The Returns to In-Company Training", *Industrial and Labor Relations Review*, 54 (3).
- [20] Blundell, R. and S. Bond, 1998, "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Journal of Econometrics* 87.
- [21] Blundell, R. and S. Bond, 2000, "GMM Estimation with Persistent Panel Data: An Application to Production Functions", *Econometric Reviews*, 19.
- [22] Blundell, R., L. Dearden and C. Meghir, 1996, "Work-Related Training and Earnings", *Institute of Fiscal Studies*.
- [23] Bond, Steve and M. Söderbom, 2005, "Adjustment costs and the identification of Cobb Douglas production functions". IFS Working Papers W05/04, *Institute for Fiscal Studies*.
- [24] Booth, A., 1991, "Job-related formal training: who receives it and what is it worth?", *Oxford Bulletin of Economics and Statistics*, vol. 53.
- [25] 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.
- [26] Chamberlain, G., 1984, "Panel Data", *Handbook of Econometrics*, eds. Z. Griliches and M. Intriligator, Vol. 2.
- [27] Conti, G., 2005. "Training, productivity and wages in Italy", *Labour Economics*, vol. 12(4), pages 557-576.
- [28] 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.
- [29] Frazis, H. and G. Lowenstein, 2005, "Reexamining the Returns to Training: Functional Form, Magnitude and Interpretation", *The Journal of Human Resources*, XL, 2.
- [30] Griliches, Z. and J. Mairesse, 1995, "Production Functions: The Search for Identification", *NBER wp 5067*.
- [31] Heckman, J. and E. Vytlacil, 2005, "Structural Equations, Treatment Effects, and Econometric Policy Evaluation", *Econometrica*.

- [32] Leuven, E., 2004, “The Economics of Private Sector Training”, *Journal of Economic Surveys*, forthcoming.
- [33] Leuven, E. and H. Oosterbek, 2004, “Evaluating the Effect of Tax Deductions on Training”, *Journal of Labor Economics*, Vol. 22, No. 2.
- [34] - -, 2005, “An alternative approach to estimate the wage returns to private sector training”, working paper.
- [35] Levinsohn, J. and A. Petrin, 2003, “Estimating production functions using inputs to control for unobservables”, *Review of Economic Studies*, April, 2003, Vol. 70(2), No. 243, pp. 317-342.
- [36] Lillard, L. and H. Tan, 1986, “Training: Who Gets It and What Are Its Effects on Employment and Earnings?”, RAND Corporation, Santa Monica California.
- [37] 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.
- [38] Mincer, J., 1989, “Job Training: Costs, Returns and Wage Profiles”, NBER wp 3208.
- [39] Pischke, J., 2005, “Comments on “Workplace Training in Europe” by Bassanini et al.”, working paper, LSE.
- [40] Olley, S. and A. Pakes, 1996, “The dynamics of productivity in the telecommunications equipment industry”, *Econometrica*, 64.
- [41] Roodman, D., 2005, “Xtabond2: Stata module to extend xtabond dynamic panel data estimator”, Statistical Software Components, Boston College Department of Economics.
- [42] Stevens, M., 1994, “A Theoretical Model of On-the-Job Training with Imperfect Competition”, *Oxford Economics Papers*, 46.
- [43] Zwick, T., 2004, “Employee Participation and Productivity”, *Labour Economics*, 11, 715-740.