DISCUSSION PAPER SERIES

IZA DP No. 1382

Firm-Level Social Returns to Education

Pedro S. Martins

November 2004

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Firm-Level Social Returns to Education

Pedro S. Martins
Queen Mary, University of London
and IZA Bonn

Discussion Paper No. 1382
November 2004

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

IZA Discussion Paper No. 1382
November 2004

ABSTRACT

Firm-Level Social Returns to Education*

Do workers benefit from the education of their co-workers? This question is examined first by introducing a model of on-the-job schooling, which argues that educated workers may transfer part of their general skills to uneducated workers and that this spillover is affected by the degrees of non-excludability, irreversibility and generality of those skills. We then conduct an empirical analysis drawing on a matched panel of Portuguese firms and their workers. Schooling endogeneity is tackled by considering firm fixed effects and instruments based on schooling lags and the lagged share of retirement-age workers. We find evidence of large firm-level social returns (ranging between 14% and 23% – and thus exceeding standard estimates of private returns) and of significant returns accruing to less educated workers but not to their more educated colleagues.

JEL Classification: J24, J31, I20

Keywords: education, spillovers, matched employer-employee data, wages, endogenous growth

Pedro S. Martins
Centre for Business Management
Queen Mary, University of London
Mile End Road
London E1 4NS
United Kingdom
Email: p.martins@qmul.ac.uk

* I thank, without implicating, Daron Acemoglu, Giulio Fella, Daniel Hamermesh, James Heckman, Jim Jin, Francis Kramarz, Robin Naylor, Pedro Portugal, Helena Skyt Nielsen, Jonathan Thomas, Ian Walker, Yoram Weiss and seminar participants at IZA (Munich), ESPE (Bergen), CAM (Copenhagen), EALE (Lisbon) and at the Universities of Aberdeen, St Andrews, London (Queen Mary) and Kent for their useful comments. Financial support from *Fundação para a Ciência e a Tecnologia* (SFRH/BD/934/2000 and POCTI/ECO/33089/99) and logistical support from *Banco de Portugal* are also gratefully acknowledged.

1. Introduction

While the labour economics literature has devoted considerable attention to the estimation of *private* relevance of education, relatively little is known about its social importance. However, from many different points of view, *social* returns to education are the key parameter to take into account. Indeed, a better understanding of whether education increases total output (and not simply of how much more the more educated earn) and if there are additional benefits above the private returns is of paramount importance in a number of areas.

One important example is a policy question frequently raised in the public debate: how should education costs be split between the student and the taxpayer? If, for instance, any returns to education that may exist are only private, then the case for public subsidies for education comes significantly eroded (see Gemmell, 1997). Another question, this time possibly of a more academic nature, concerns the importance of education for economic growth. While some of the endogenous growth literature argues that investment in education can sustain indefinitely positive growth rates of income per capita (Lucas, 1988), the empirical support of these views is far from clear (see the discussion in Krueger and Lindahl, 2001).

One explanation for the scarce empirical evidence about social returns to education lies on the demanding estimation strategy required. Firstly, as in many other areas of empirical research, one needs appropriate counterfactuals. In the case of private returns to education, this amounts to information about the earnings of very similar individuals but that have different levels of education: one original strategy based on this idea studies differences in earnings and schooling of twins (Ashenfelter and Krueger, 1994). In the case of social returns to education, this requirement may entail information about very similar firms but that happen to be located in cities with different levels of average human capital.

Secondly, one needs exogenous variation in education. For instance, in the twins approach, one may wonder why do individuals as similar as twins have different schooling levels. In particular, it is unclear whether some of those factors driving schooling variability across twins will not also be correlated with wages. In this case, methods based on instrumental variables can generate education variability which can then be used for identification purposes: for instance, Harmon and Walker (1995) draw on the increases of school leaving age in the UK to estimate private returns to education. In the case of social returns, and taking the previous example of firms in different cities, one would need variables that affect city-level education but not firm wages. However, even this IV approach is not free from criticism as the results obtained may be specific to the group affected or difficult to interpret without a more structural approach.

Finally, the estimation of social returns to education (unlike private returns) may additionaly have to deal with possible general equilibrium effects. For instance, if high- and low-skill workers are imperfect substitutes, then an increased supply of the former will affect the prices of both types of workers even if spillovers do not exist (see Ciccone and Peri, 2002).

On top of the estimation hurdles described above, the few studies available on the topic of social returns to education have not yet reached any stylised fact, not even on whether there are social returns, not to mention their specific magnitude. In particular, two of the most prominent papers, Acemoglu and Angrist (2000) and Moretti (2004a), find conflicting results, even if drawing on similar data sets for the same country (the U.S.) and employing similar methodologies, although considering slightly different time periods and units of analysis.

Acemoglu and Angrist (2000) draw on compulsory schooling laws (compulsory attendance and child labour laws) to identify the impact of average schooling on average wages in US states, finding insignificant external returns. On the other hand, Moretti (2004a) uses city demographic structures and the geographical presence of some colleges to find significant impacts of the share of graduates on the wages of workers in the same city, particularly on those workers with lower levels of schooling. These contrasting results also extend to other studies that examine the U.S. case but that take schooling as exogenous: Rauch (1993) finds positive, significant effects while Rudd (2000) documents insignificant effects.[1]

However, while the findings on *wage* social returns to education are mixed, more consistent evidence has been found for other social domains where education may also matter. Studies

[1] In a different study, using a new methodology, based on a “constant-composition” approach, Ciccone and Peri (2002) also find insignificant results.

focusing on productivity (Moretti, 2004b), crime (Lochner and Moretti, 2004), citizenship (Milligan et al., 2004), and intergenerational effects (Currie and Moretti, 2003) find positive and significant effects of education.[2]

It is to this emerging literature on social returns to education that the present paper contributes. On the theoretical side, we present a simple model of *on–the–job schooling*, a new concept which we introduce here. We argue that educated workers at a firm may transmit part of their education skills to their uneducated colleagues. This assumption is intuitively appealing as it is likely that most of the education externalities that affect productivity (and, subsequently, pay) occur within firms: it is probably precisely at the firm level that workers, including those of different skills or education levels, interact the most – and not at the city or region levels, as examined before. In any case, the strength of this schooling spillover is allowed to vary depending on some of the skill characteristics, namely its non-excludability, irreversibility and generality (defined below). Among other empirically testable results, our model shows that, at least for some labour market structures, the existence of spillovers leads to a stronger relationship between wages and education at the firm level than at the individual level.

We then empirically examine this and other implications of the model, using a panel of about 5,000 Portuguese firms and their workers, followed for up to nine years. Focusing on firms, we not only address the less aggregated level of analysis considered in the model, but we also sidestep possible general-equilibrium effects induced by imperfect substitutability. Moreover, the longitudinal dimension of the data allows us to implicitly control for unobserved differences across firms, as in the twins literature, and we benefit from within-firm variation of education driven by the vigorous educational expansion experienced in Portugal. Finally, we also use instruments (lagged schooling and the lagged share of workers approaching retirement age) to generate exogenous variation of education. This is particularly important as the model emphasises the endogeneity of firm-level education, even within each firm over different time periods.

[2] However, Behrman and Rosenzweig (2002), who also study intergenerational mobility, find in their U.S. data that the correlation between the schooling of mothers and their children can be explained by heritable ability and assortative mating.

Consistently with the model, we estimate social (firm-wide) returns to education significantly above those commonly obtained in studies of private (individual) returns. These social returns are particularly high when focusing at specific firm/job-levels cells, a less aggregated level of analysis where one would intuitively expect greater spillovers. Again as predicted by the model, we also find that the less-educated workers benefit from increases in their firms' average schooling levels, unlike their more educated counterparts.

The remaining of the paper is structured as follows. The next section describes the "on-the-job schooling" model. Section 3 presents the empirical methodology and the data. Section 4 describes the results, the robustness analysis and some extensions and discusses the implications of our findings. Section 5 concludes.

2. A Model of On-the-Job Schooling

This section presents a simple model that examines the implications of the admission of more skilled workers in a given firm and the reasons why such hiring may increase the average wages in a firm at a rate greater than that predicted by standard, individual-level estimates of private returns to education. The key assumption is that education affects the productivity not only of the educated individuals themselves (a direct, private effect) but also of those workers who have not invested in a higher educational level but who interact with colleagues that have (an indirect effect or spillover).

2.1. General Outline

The model allows the new skills obtained in higher schooling levels by the new workers to be transferred to other workers that were already in the firm. In a sense, new workers act as informal teachers (of on-the-job schooling) to unskilled workers. To this extent, skilled workers transfer to unskilled workers general knowledge that those (previously) unskilled workers can now use to achieve higher productivity levels in that or other firm.[3] This is thus another sense in which one may argue that "skills beget skills" (Carneiro and Heckman, 2003).

There are three main characteristics of this process of on-the-job schooling. The first aspect is about the *non-excludability* of the spillover: we consider that the educated cannot prevent the spillover from occurring (i.e. their schooling skills being passed on to less educated colleagues), at least to some degree. We argue this assumption is reasonable because if education is valuable for its productive skills (and not because of signalling issues), then the educated workers will presumably have to certify their acquisition of those skills by applying them in firms. This application of skills by the educated in their workplace typically takes place in such a way that it allows other workers, in particular the less educated colleagues, to learn from the educated. Educated workers therefore become *de facto* workplace teachers of their uneducated colleagues.

Non-excludability is also what prevents the educated workers from capturing the spillover created by their presence in the firm. In fact, those workers cannot credibly threaten not to disseminate their skills to other workers as that would imply working at a productivity level below that which is expected of an educated worker. A related point is to emphasise the difference between on-the-job schooling and on-the-job training: unlike the latter, on-the-job schooling does not rest on a *deliberate* decision taken by firm owners to improve the skills of some specific workers. On-the-job schooling is instead an *informal* process,[4] potentially affecting all workers in a firm, in particular the less educated ones – those that stand to benefit the most from such interactions.

Irreversibility is the second aspect of on-the-job schooling. This characteristic is important in that it creates a permanent spillover from the recently-hired skilled workers to the other workers in the firm. It contrasts to the case when productivity is affected simply because there are complementarities between different types of workers in the firm's technology. Under the

[3] We will focus on quantity differences in the schooling of different workers. The model could, however, also be adopted to consider quality differences (e.g. degrees in different subjects).

[4] Some businesspeople argue that most learning inside firms is of an informal type, via worker-to-worker interaction, rather than via formal methods, such as training sessions or even e-learning.

case of complementarities (or some firm-specific public good), once unskilled workers are not interacting with their skilled colleagues (or when the public good is removed), the unskilled marginal product would return to its previous, pre-interaction, lower level. On the other hand, in the present case of on-the-job schooling, if the interaction came to an end, the unskilled marginal product would still stay at its previous, higher level, since the spillover is irreversible.

The third and last aspect in the transmission process described here is that schooling is regarded as providing *general* skills. As argued in Acemoglu and Pischke (1999b), "schooling provides *general-purpose knowledge* by teaching conceptual tools and information, useful in a variety of occupations and industries" (page F140, our italics). Since these types of skills are widely used, across many different firms, their possession will increase the workers' outside options. Examples will obviously depend on the specific level and type of schooling considered but will include greater competence in numeracy and literacy, for instance. Specific examples may include computing, languages and management skills.

Another point to mention is that we allow education to present different degrees of these three aspects (non-excludability, irreversibility and generality) in different contexts. This will imply that the strength of the education spillover may vary across occupations and/or industries, for instance. Moreover, a precise clarification of the practical importance of spillovers is, of course, an empirical matter, which is examined below.

At this stage, it is also important to state that the productivity spillovers discussed here do not necessarily imply wage spillovers. Even if such productivity spillovers do indeed materialise, the extent to which wage spillovers will be observed depends, we argue, on the specific type of the labour market. In particular, if the labour market is sufficiently competitive (i.e., subject to few frictions), the outside option of those unskilled workers will improve in parallel to the importance of the new skills obtained from their educated colleagues. Alternatively, a similar result of improved outside options will apply if the unskilled workers have enough bargaining power so that the rents created by the spillover can be captured by those workers.[5]

---

[5] Different papers, using a variety of methods, document significant and, in many cases, substantial levels of rent sharing in many labour markets, including Blanchflower et al (1996) for the U.S., Hildreth and Oswald (1997) for the U.K., Arai (2003) for Sweden, and Kramarz (2003) for France. Martins

In the two extreme cases of perfect competition and "full" rent sharing, unskilled workers (or, better, workers who were unskilled before the admission of skilled workers) will be able to extract higher wages from their current employers up to the level of their new marginal products. In this sense, productivity spillovers will imply wage spillovers.[6] However, in the other extreme case in which the labour market is monopsonistic (due to mobility costs, information constraints or other frictions), the outside option of unskilled workers will not change even if their effective productivity has increased. As discussed in Acemoglu and Pischke (1999a), what are technologically general skills become *de facto* firm-specific skills in a market with frictions. Therefore, employers will capture the entire surplus generated by the inflow of skilled workers and the wages of unskilled workers will not change.

After providing a non-technical outline of the key features of the model and highlighting the importance of labour market structures, we now describe the model in greater detail.

2.2. Model Description and Results

We consider a two-period, partial equilibrium framework, in which a single, representative firm chooses the quantity of educated workers it will hire in the second period. There are only uneducated and educated workers in this economy. The former are only available to be hired in the first period, while the educated workers can only be hired in the second period. In the first period, the firm therefore employs only uneducated workers (L_U), whose quantity is normalised at 1; in the second period, the firm chooses how many educated workers to hire (L_S).[7]

(2004) also finds evidence of rent sharing in the Portuguese labour market, using worker/firm spell fixed effects and exogenous variation in profits derived from interactions between exchange rates and export shares. Martins (2004) also finds rent sharing to be larger for workers with more tenure. See also Manning (2003) for an analysis of different theoretical and empirical aspects of monopsony.

[6] Moreover, these two labour market types are in this paper observationally equivalent. We will thus use the phrases "perfectly competitive" and "rent-sharing" interchangeably.

[7] These differences in workers' type over the two periods are motivated by the expansion of secondary and tertiary education systems observed throughout the Western world during the second half of the 20th century. For instance, in the case of Portugal, compulsory schooling increased from four to nine

Furthermore, the wages of the educated workers are exogenously defined at W_S – this is because we assume that the firm, our unit of analysis, will never be big enough to be able to affect wages in the whole economy. Unskilled wages are also exogenously defined at W_U ($W_U < W_S$) – this may be the minimum wage, for instance.[8] The product price is also normalised at 1.

The firm generates output according to a Cobb-Douglas production function with decreasing returns to scale: $Y = L_U^{\alpha} L_S^{\beta}$, $\alpha + \beta < 1$. However, we also allow the interaction between skilled and unskilled workers to lead to on-the-job schooling, whereby unskilled workers become more productive. The interaction is assumed to be as follows: $L_U^{'} = \gamma L_S , \left(\frac{1}{L_S} < \gamma < 1 \right)$.

This means that $L_U^{'}$, the *effective* quantity of educated workers (i.e. after taking into account their higher productivity) in the second period, will depend (positively) on the number of educated workers hired by the firm. Moreover, the bounds on γ (the spillover parameter) ensure that the inflow of educated workers will never decrease the effective size of the uneducated below their previous size in the first period (which was 1), i.e. no negative spillovers, and will also never increase the effective size of the uneducated above the size of the educated workers.[9]

We consider three types of labour-market/spillover scenarios: 1) no spillovers (denoted throughout the rest of the paper by the letter N); 2) spillovers, in a monopsonistic labour market (M); and 3) spillovers, in a perfectly competitive or rent-sharing labour market (P). As mentioned before, the difference between M and P is that in M the unskilled workers will not benefit from any wage increases as their productivity increases due to the spillover, while in P

years in a period of only two decades (from 1965 to 1986), implying clear differences in the characteristics of individuals leaving the education system over that time interval. This trend is now being upgraded to the tertiary/secondary levels.

[8] These two assumptions about wages imply that, in effect, we rule out the general equilibrium effects mentioned in the introduction. As mentioned before, this is justified by our consideration of a less aggregate level of analysis.

[9] The bounds on γ also imply that our analysis is implicitly restricted to certain ranges of the other parameters of the model, so that $L_S > 1$. For instance, very large gaps between W_U and W_S are ruled out.

their pay will be raised accordingly. Pay will stay unchanged in N as there are no spillovers in that case.

We now examine the results of this model under each specific labour-market/spillover case:

2.2.1. Case 1 (no spillovers)

Profits in the second period are affected only by the choice of L_S and the production function becomes $Y_2 = L_U^{\alpha} L_S^{\beta} = L_S^{\beta}$ (since L_U=1). The firm's problem will be to find the optimal L_S that maximises profits: $Max_{L_S^N} \Pi_2 = L_S^{N\,\beta} - W_U - W_S L_S^N$. The first-order condition and its solution (the labour demand function) are:

$$(1) \qquad \frac{\partial \Pi_2}{\partial L_S^N} = 0 \Rightarrow L_S^N = \left(\frac{W_S}{\beta} \right)^{\frac{1}{\beta - 1}} = L_S^N (W_S ; \beta)$$

It is easily shown that labour demand (for L_S) falls with the educated wage and increases with β. In order to draw some empirically testable predictions of the model, we now derive some expressions for variables that are amenable to empirical examination and that can be related to other studies in the literature (in particular those about private returns to education) and to the different labour market types considered in this paper. First, the firm-level average wage and its partial derivative with respect to the number of educated workers will be:

$$(2) \qquad \overline{W}^N = \frac{W_U + W_S L_S^N}{1 + L_S^N} \Rightarrow \frac{\partial \overline{W}^N}{\partial L_S} = \frac{W_S - W_U}{(1 + L_S^N)^2}$$

The average wage then increases when more educated workers are hired (if their wage is bigger than that of the uneducated workers – which is true by assumption). This partial derivative corresponds to the sensitivity of the firm average wages to within or between changes in the levels of educated workers hired. (These changes in the levels of educated workers can be understood, for instance, as stemming from shocks that affect the optimal hiring of such workers.)

Moreover, average education in the firm (measured as the share of educated workers[10]) and its derivative with respect to L_S are given by:

$$(3) \qquad \overline{ed}^{N} = \frac{L_S^N}{1+L_S^N} \Rightarrow \frac{d\overline{ed}^{N}}{dL_S^N} = (1+L_S^N)^{-2}$$

The last two results [(2) and (3)] imply the following positive relationship between log mean wages, $\ln \overline{W}^{N}$, and average education, $\overline{ed}^{N}$:

$$(4) \qquad \frac{\partial \ln \overline{W}^{N}}{\partial \overline{ed}^{N}} = \frac{\partial \ln \overline{W}^{N}}{\partial \overline{W}^{N}} \frac{\partial \overline{W}^{N}}{\partial L_S^N} \frac{\partial L_S^N}{\partial \overline{ed}^{N}} = \frac{W_S - W_U}{\overline{W}^{N}}$$

The increase of log mean wages with respect to mean education is approximately determined by the percentage wage gap between educated and uneducated workers (i.e. by the standard private return to education). This partial derivative is comparable to those at the individual-level Mincer private returns to education studies, which will be taken as the benchmark against which we will evaluate our results. The formulation of equation 4 may, however, indicate that the firm-level returns to education would be less than that of individual-level studies, implying that our estimates of the external return to education, based on comparing firm- and individual-level returns to education would, if anything, be biased *downwards*.

2.2.2. Case 2 (positive spillovers, monopsonistic market)

[10] The expression can be adjusted to take into account the difference in years of education between the educated and the less educated groups. This does not, however, bring about any qualitative difference to our findings.

The firm's problem in this new case will be, again, to find the optimal L_S. Unlike before, now L_S has the additional benefit that it increases the productivity of the uneducated, as the production function becomes $Y_2 = (\gamma L_S)^\alpha L_S^\beta = \gamma^\alpha L_S^{\alpha+\beta}$. On the other hand, since in this case the labour market is assumed to be monopsonistic, the increased productivity of the uneducated will not affect their wages, which will stay unchanged. The new problem faced by the firm is then $Max_{L_S^M} \Pi_2 = \gamma^\alpha {L_S^M}^{\alpha+\beta} - W_U - W_S L_S^M$ and the first-order condition and its solution are:

$$(5) \qquad \frac{\partial \Pi_2}{\partial L_S^M} = 0 \Rightarrow L_S^M = \left[\frac{W_S}{\gamma^\alpha (\alpha+\beta)} \right]^{\frac{1}{\alpha+\beta-1}} = L_S^M (W_S; \alpha, \beta, \gamma)$$

It can be shown that, as before, labour demand falls with the educated wage and increases with β. Moreover, the average wage will also be similar to that of case 1, the difference lying on the new specific hiring level of educated workers (L_S^M instead of L_S^N):

$$(6) \qquad \overline{W}^M = \frac{W_U + W_S L_S^M}{1 + L_S^M} \Rightarrow \frac{\partial \overline{W}^M}{\partial L_S^M} = \frac{W_S - W_U}{(1 + L_S^M)^2}$$

Moreover, firm-level average *formal* education, i.e. the share of educated workers, disregarding on-the-job schooling, and its derivative with respect to the number of educated workers are also given by similar expressions than before:

$$(7) \qquad \overline{ed}^M = \frac{L_S^M}{1 + L_S^M} \Rightarrow \frac{d\overline{ed}^M}{dL_S^M} = (1 + L_S^M)^{-2}$$

The two previous results imply again a similar relationship between log mean wages and education:

$$(8) \qquad \frac{\partial \ln \overline{W}^M}{\partial \overline{ed}^M} = \frac{\partial \ln \overline{W}^M}{\partial \overline{W}^M} \frac{\partial \overline{W}^M}{\partial L_S^M} \frac{\partial L_S^M}{\partial \overline{ed}^M} = \frac{W_S - W_U}{\overline{W}^M}$$

2.2.3. Case 3 (positive spillovers, competitive or rent-sharing market)

In this last case, the firm's optimal L_S will take into account the spillover benefit for uneducated workers (same production function as in case 2) but also their wage increase following their increased productivity. Therefore, the new maximisation problem becomes $Max_{L_S^P} \Pi_2 = \gamma^{\alpha} L_S^{P\,\alpha+\beta} - W_U(\gamma L_S^P) - W_S L_S^P$

Notice that the wage bill corresponding to the unskilled workers considered in the profit function takes into account their effective size, which is now more than one because the spillover makes them more productive (recall that, by assumption, $\gamma L_S > 1$). This specific new effective size depends on the spillover parameter (γ) and the number of educated workers hired (L_S^P).[11]

A more general formulation, not pursued here, would include a rent sharing parameter that indicated the share of the spillover surplus transferred to employers and less educated workers. Considering this parameter, the new wage for unskilled workers would be $W_U[\varphi(\gamma L_S) + (1-\varphi)]$, in which φ indicates the bargaining power of unskilled workers. In case 2 (monopsonistic labour market) φ was equal to zero. In the present type of labour market model (perfect competition or full bargaining power of workers), φ is equal to one.

In this case, the first-order condition and its solution are then:

(9) $$\frac{\partial \Pi_2}{\partial L_S} = 0 \Rightarrow L_S^P = \left[\frac{W_S + \gamma W_U}{\gamma^{\alpha}(\alpha+\beta)}\right]^{\frac{1}{\alpha+\beta-1}} = L_S^M(W_S, W_U; \alpha, \beta, \gamma)$$

It can be shown that, as before, labour demand falls with the educated wage and increases with β. Moreover, unlike before labour demand is now also affected (negatively) by the uneducated wage.

[11] It is important to notice that back-loading could not occur in this model. Back-loading would imply that uneducated workers, competing for the spillover, would take a pay cut in the first period and earn market wages (at a higher level of skill) in the second period. However, such uneducated workers would not know what would be the level of the spillover at the first period.

Under these new circumstances, the average wage and its partial derivative with respect to the quantity of educated workers become different expressions:

$$\text{(10)} \qquad \overline{W}^{P} = \frac{W_U \gamma L_S^P + W_S L_S^P}{1 + L_S^P} \Rightarrow \frac{\partial \overline{W}^{P}}{\partial L_S^P} = \frac{W_S + \gamma W_U}{(1 + L_S^P)^2},$$

while average (formal) education in the firm and its derivative remain similar to their equivalents in the other labour markets:

$$\text{(11)} \qquad \overline{ed}^{P} = \frac{L_S^P}{1 + L_S^P} \Rightarrow \frac{d\overline{ed}^{P}}{dL_S^P} = (1 + L_S^P)^{-2}$$

The results in (10) and (11) imply the following relationship between log mean wages and education:

$$\text{(12)} \qquad \frac{\partial \ln \overline{W}^{P}}{\partial \overline{ed}^{P}} = \frac{\partial \ln \overline{W}^{P}}{\partial \overline{W}^{P}} \frac{\partial \overline{W}^{P}}{\partial L_S^P} \frac{\partial L_S^P}{\partial \overline{ed}} = \frac{W_S + \gamma W_U}{\overline{W}^{P}} = \frac{1 + L_S^P}{L_S^P}$$

A similar relationship between $\ln \overline{W}_U^{P}$, the log of the average wages of the uneducated workers (after spillovers and pay rises), and, $\overline{ed}^{P}$, average education, can also be derived:

$$\text{(13)} \qquad \frac{\partial \ln \overline{W}_U^{P}}{\partial \overline{ed}^{P}} = \frac{\partial \ln \overline{W}_U^{P}}{\partial \overline{W}_U^{P}} \frac{\partial \overline{W}_U^{P}}{\partial L_S^P} \frac{\partial L_S^P}{\partial \overline{ed}} = \frac{(1 + L_S^P)^2}{L_S^P}$$

We now present the following results, obtained from comparing our findings across different labour market/spillover models:

Result 1: $L_S^M > L_S^P$

This inequality can be immediately derived from comparing equations (1) and (5). Provided that $\gamma>0$ (which is true by assumption), more educated workers will be hired under monopsony (case 2) than under competition (case 3). This result holds not only in absolute terms (when considering L_S) but also in relative terms (when considering $\overline{ed}$). (Moreover, it can also be shown that the bigger is γ, the bigger the educated hiring difference across the two market structures.)

The intuition behind Result 1 is straightforward but important. While the firm's marginal benefit of hiring educated workers (L_S) is the same under cases 2 and 3, the marginal cost is greater in case 3. This is because in the latter case the employer has to pay the productivity spillover γW_U (in the form of higher wages to unskilled workers) while in case 2 the unskilled wages are ignored as they are simply sunk costs, since the wage bill of unskilled workers is not affected by how many educated workers are hired.

Result 2: $$\frac{\partial \ln \overline{W}_U^P}{\partial \overline{ed}^P} > \frac{\partial \ln \overline{W}^P}{\partial \overline{ed}^P}$$

This follows immediately from a comparison of (13) and (12). The return to firm-level education, as measurable by a firm-level Mincer equation, is bigger for the subset of the uneducated than for the entire workforce. Moreover, under our assumption that the wages of the educated workers are exogenous and not affected by the number of educated workers in the firm, the model predicts no impact of changes in firm-level education on the wages of the educated workers. However, the result that the wages of less educated workers are positively affected by the inflow of educated workers to a firm can be understood as indicating positive social returns to education.

Result 3: $$\frac{\partial \ln \overline{W}^P}{\partial \overline{ed}^P} > \frac{\partial \ln \overline{W}^M}{\partial \overline{ed}^M}$$

A proof of this result is presented in the Appendix.[12] The result says that average wages are more sensitive to average education in a perfect competition setting than in a monopsonistic labour market.

Result 4: $$\frac{\partial\left(\frac{\partial \ln \overline{W}^{P}}{\partial \overline{ed}^{P}}\right)}{\partial \gamma} > 0$$

The proof of this result is also presented in the Appendix. This inequality means that the sensitivity of average wages with respect to average education increases with the spillover parameter, in the case of a perfectly competitive labour market.

Result 5: $$\frac{\partial^2 \ln \overline{W}_U^{P}}{\partial \overline{ed}^{P^2}} > 0$$

This result follows immediately from calculating the partial derivative of (13) with respect to $\overline{ed}^{P}$. The positive sign of the second partial derivative can be interpreted as *increasing* social returns to education, i.e. the wage spillovers to the uneducated in a competitive market increases more than proportionately with the inflow of more educated workers to the firm. Notwithstanding this, a caveat to bear in mind is that while the uneducated benefit increasingly more as there are more educated workers, there are also increasingly fewer uneducated workers in relative terms.

2.3. Other Comments

The different labour market structures that we consider will also lead to different policy implications, given the different wage spillovers that arise in each case. Under monopsonistic labour markets, employers will be residual claimants of the surpluses generated by the

[12] The comparison of the two partial derivatives is not immediately obvious as, while the numerator in (12) is much bigger than that of (8), the ranking of the denominators in unclear.

admission of skilled workers. Therefore, since employers will be able to fully internalise the externalities from their hiring choices, the social optimal level of on-the-job schooling (i.e. that that maximises total output) will be achieved. There is consequently no need for government intervention, at least from the efficiency point of view.[13]

However, if labour markets are competitive or if workers have sufficient bargaining power, employers will no longer be able to benefit from the productivity spillovers that may be generated by the admission of educated workers. This will constrain employers to hire socially suboptimal levels of skilled workers, in which the gap between the optimal and the realised levels will increase with the strength of the productivity spillover. Government intervention may then be important in correcting this inefficiency. Such public policies would make skilled workers cheaper from the point of view of employers, for instance by paying subsidies to employers that hire skilled workers or to the skilled workers themselves – precisely the opposite of standard wage subsidies policies.

On another level, it is interesting to notice that some of the implications of this model contrast with those that arise in studies of general training. In particular, it is found in this literature that, in competitive markets, firms should never pay for investments in general training (Becker, 1975). If they do, then workers can move to other firms and benefit from higher wages without having incurred the investment costs for their higher levels of human capital. Conversely, in non-competitive markets, firms may pay at least part of the costs of such investments in general training (Acemoglu and Pischke, 1999b). This will occur, for instance, if firms take an increasing share of the worker's productivity as the worker's skill level increases: in this case, the subsidisation of general skills can be profitable for the firm.

However, we have found that in the case of on-the-job schooling, which we also characterise as an investment in general human capital, firms will pay for these skills only in competitive markets. In markets with frictions, on the other hand, it will be the workers themselves that implicitly pay for the skills, as they earn the same wages while their productivity increases due

[13] Our approach also implies an additional social cost of (voluntary or involuntary) unemployment, as individuals out of work will be foregoing opportunities to benefit from on-the-job schooling, on top of any possible additional depreciation of their schooling and training skills.

to the spillovers. Without frictions, employers will not be able to keep pay unchanged, as then workers would move to other firms that paid them their new productivity levels. This interesting reversal result follows from the non-excludability assumed for on-the-job schooling: if less educated workers had to pay directly to have access to this type of schooling, then we would be back to the results described for the standard investments in general training.

Finally, before we present our empirical analysis, we also wish to compare the externality described here with that of Acemoglu's (1996). This influential model presents an economy with constant returns to scale and complementarities between human and physical capital. This economy is then shown to generate social increasing returns to human capital from market interactions involving matching frictions between workers and firms. These increasing social returns arise as the return to a worker's human capital will depend *positively* on the investment in human capital made by the remaining workers. The intuition is that the interplay between frictions and complementarities will make firms invest more in physical capital when the workers' investment in human capital increases. In this way, each workers' marginal product increases when other workers invest in their own human capital.

While the externalities in Acemoglu's paper can be characterised as of a strictly pecuniary nature – there is no technologically-imbedded spillover and the externality simply arises from the interaction between different agents – our model exhibits both pecuniary and technological externalities. Indeed, here the external effects emerge from (and are, in some cases, amplified by) market interactions between firms and workers as the firm maximises profits by hiring the optimal number of skilled workers. Nevertheless, there is also a technological externality, as the spillover mechanism from educated to less educated workers is at the root of the externality in the first place.

There are some additional differences between the two models. In our paper, the externalities stem only from the decisions taken by firms – which are subject to *decreasing*, not constant, returns to scale – as to how many educated workers to hire in a given period, and not from the decisions taken by firms and workers as in Acemoglu's paper. On the other hand, in our paper the spillovers arise only from the educated and benefit only the less educated, unlike in Acemoglu (1996), where each worker benefits from the human capital investments of all other workers. This asymmetry in the direction of the spillovers is an important prediction that will be tested empirically in Section 4. One last point is that, although both models underline the importance of labour market frictions, in Acemoglu (1996) they are a key driver of the externalities, while in our paper such frictions are not important in the generation of productivity spillovers. Frictions are however crucial in generating or not *wage* spillovers, following any possible productivity spillovers.

3. Empirical Approach

Based on the theoretical model of on-the-job schooling, our empirical work, introduced in this section, is implemented by aggregating individual-level Mincer (1974) equations to the firm level. This follows from the predictions derived from the model as to how average, firm-level wages of specific groups of workers or the entire workforce vary with average, firm-level education.

As mentioned in the introduction, it is critical to draw on data that present enough variability over time in educational attainment. This concern about variability motivated our use of data for a country and a period that document large upgrades in the schooling of its workforce: during the 1990s, Portugal experienced a substantial educational catching-up of its labour force. In our data, presented below, the average years of schooling increased by about 17% over a period of nine years, from 5.9 in 1991 to 6.9 in 1999.[14]

This considerable increase is in part due to the rise of the minimum level of schooling from six to nine years that occurred on 1986. Additionally, the legal constraints that prevented the expansion of private universities were lifted in the same period, allowing for a large increase in enrolment in such institutions, as until then the demand for university education clearly exceeded its supply by public institutions. (Pereira and Martins (2001) describe these and other developments of the Portuguese education system in greater detail.)

[14] These are very low average schooling figures for a European country. They correspond to an average school leaving age of the Portuguese workforce of around 12 or 13. Compulsory schooling was only four years of schooling (schooling leave age of 12) until the early 1960's.

A second concern mentioned in the introduction is that education variability is exogenous with respect to wage determination. This point is further reinforced in the context of our model, in which the average education level of each firm is explicitly regarded as an endogenous variable and it is shown to be affected by parameters such as the factor shares α and β and the spillover parameter γ that may easily vary from firm to firm.

While our consideration of firm fixed effects allows the educational attainment of workers in each firm to be correlated with all time-invariant factors (observed or unobserved) that influence wages at that same firm, our estimates would become inconsistent if any of the variables or parameters mentioned above (including the characteristics of the specific labour market in which the firm is active) varied over time. For instance, a technological shock that affects β will lead to different levels of educated workers and average wages and biased estimates of the impact of education.

This is a possibility that may affect the study of Barth (2002), who also looks at firm-level returns to education. Drawing on the longitudinal dimension of his Norwegian worker-level data, but assuming education variation to be exogenous, he finds a significant effect of the establishment average level of education on workers' wages.[15]

Moreover, there are additional sources of endogeneity bias, not outlined by the model but of practical relevance, that can emerge from the interplay between workforce adjustment and time-varying shocks that also affect earnings. For instance, firms experiencing an increasing demand for their products may hire younger and more educated workers and, simultaneously, due to rent sharing, increase the earnings of both stayers and/or entrants above the market benchmark. This would lead to spurious positive correlation between education and wages and thus bias upward the education coefficient in a firm-level wage equation.

[15] See also Battu et al (2003), who find significantly positive effects in a cross-section study of British establishments, proxying firm average education from the distribution of workers across different occupations.

Alternatively, firms facing negative demand shocks may simultaneously foster the early retirement of their senior workers and demand pay concessions from their remaining employees. In this case, there would be spurious negative correlation between education and wages and the education coefficient would be downward biased. Moreover, measurement error may also bias downward that coefficient, as it typically attenuates the estimate towards zero, also in panel data models (Griliches and Hausman, 1985).

Given these different and conflicting possibilities, we sought to derive consistent estimates using instrumental variables. The first instrument is the firm's lagged education level. As firms keep unchanged a large share of their workforce in each two subsequent periods, we expect there will be a significantly strong correlation between present and lagged education. However, lagged education is unlikely to have a direct role in current wages, as we control for current education and firm fixed effects.

Our second instrument is the lagged share of workers that are of retiring age in each firm-year. The intuition here is as follows: As workers reach their retirement age, they will sooner or later leave the firm (retirement is, in general, not compulsory in Portugal), typically being replaced by younger and, by force of the above-described expansion of the education system, more educated workers. So a larger share, in period t-1, of workers that will qualify for retirement in period t should be positively correlated with firm-average education in period t. Moreover, as for the case of lagged education, we find no reasons for this lagged share of retirement-age workers to directly affect the current level of firm wages.

There is additional exogenous variability related to this instrument, as the retirement thresholds varied differently for different types of cohorts over the 1990's in Portugal. A law issued in 1991 determined that retirement age would be adjusted gradually over the decade for women, as until then it had stayed at 62 (while it was 65 for men). Specifically, it was decided that the women's retirement age should converge gradually to the men's level, increasing by six months every year, starting at 62 years and six months in 1993 until reaching men's retirement age of 65 years in 1998. Our instrument takes this legislative change into account.[16]

Before concluding this subsection, we wish to highlight a possible concern with the method pursued in the paper: it may be that the new workers of higher education hired by firms have different unobservable characteristics than those of workers hired on previous occasions. In this case, differences in earnings over time could be attributable to such differences in unobservable characteristics among stayers, leavers and entrants. Our estimate of the social return could then be biased, as it would capture unobservable factors potentially correlated with education.

However, there is some indirect evidence that this possibility does not affect our results, particularly for the medium- and large-sized firms considered here. Indeed, these firms (unlike smaller ones) typically set up expensive human resource departments that engage in long and meticulous recruitment processes, targeting and assessing worker characteristics that are unobservables for the labour econometrician. Since only good matches, from the firm and worker points of view, are likely to be stable, new hires should be comparable to their senior colleagues. Moreover, Barth and Dale-Olsen (2003) find corroborative evidence of "assortative matching" in Norwegian firms, in terms of a positive matching along observed and unobserved productivity characteristics between workers of different educational groups in their firms.

3.1. Data

We use a large matched employer-employee panel, "Quadros de Pessoal" [Personnel Records], which covers the universe of Portuguese firms with at least one employee. This data source is based on a compulsory survey administered by Portugal's Department of Employment. A large set of variables, concerning both firm and worker characteristics, is collected, including

[16] There is also some anecdotal evidence supporting the unanticipated nature of this new law. It is argued that the discontentment it created among those female workers forced to work up to three more years than they expected contributed to the downfall of the government that enacted that law in the 1995 general elections.

identifiers for each firm and each worker. These identifiers allow for both firms and workers to be followed over time. Moreover, as the survey is also to be used for inspection purposes, so that the Department of Employment can monitor each firm's compliance with different aspects of Portugal's relatively restrictive labour law, particular care is placed on the reliability of the survey.

In a first step, the analysis in this paper draws on a representative sample of 80% of all firms for each year between 1991 and 1999. We also use information about all workers for each of the firms sampled. Given that we want to focus on firms that are likely to have hiring policies as consistent as possible over time, and that we believe that such policies are positively correlated with firm size, we use in our analysis only those firms that are "large" enough, defined here as a size of at least 50 workers. Moreover, since we need to examine each firm in several periods, we chose to select only those firms that are present in our data in at least four out of the nine years available.

As we also want to minimise measurement error, we dropped firms-year in which more than 20% of workers have missing or incorrect information in the variables required in the wage equation. This procedure leaves us with 4,830 firms and 27,994 firms-year (more than 90% of the original number of firms-year), representing more than 5.9 million workers-year (and, on average, about 213 workers per firm-year).

The descriptive statistics, presented in Table 1, indicate an average schooling attainment across all firms and years of 6.5 years and an average hourly wage of 3.75 euros per hour (1999 prices). Consistently with our assumption about educational expansion, we found in separate calculations that the educational attainment at each firm increases on average by about 2.2% over two contiguous periods.

4. Results

Given the previous discussions, we consider the following wage equation:

(14) $y_{it} = \beta_1 educ_{it} + X_{it}'\beta_2 + \alpha_i + \tau_t + \varepsilon_{it}$

Here y_{it} is the logarithm of average real hourly earnings of firm i in period t. $educ_{it}$ is the average schooling years of the workers of firm i in period t (the empirical counterpart of the variable $\overline{ed}$ of the theoretical model). X_{it} is a set of average characteristics of those workers and their firm in that period: a quadratic of average experience and average tenure, the share of female workers, and size (log number of workers). α_i is the firm fixed effect, τ_t the year dummy, and ε_{it} denotes the error term.

Table 2 presents the first set of results. For the benefit of generality, we also consider pooled OLS and random effects specifications. In these two cases, which assume orthogonality between schooling and the error term, we find large estimates of returns to education, at .18 and .14, respectively.

In the fixed effects specification (3rd column), the estimated return falls considerably, but is still statistically significant and economically relevant, at .05 (5%). Moreover, the Hausman test strongly rejects the null that the difference in the random and fixed effects coefficients is not systematic (the p-value is less than .0005), thus favouring the fixed effects specification. On the other hand, this fixed-effect return is also below most of the equivalent estimates obtained in OLS analysis of private returns to education. Following our approach, these low returns suggest that, at best, there are no spillovers.

However, as discussed in Section 4, there are several reasons for the variability in firm-level education not to be exogenous as assumed in the fixed effects specification. Therefore, we now also instrument education in the fixed-effects model. The results, presented in Table 3, support the validity of the instruments. Firstly, both coefficients for the instruments in the auxiliary regression are statistically significant and positive (the sign expected from our discussion before): the coefficient for lagged schooling in the auxiliary regression is .08 (with a p-value less than .0005) and the coefficient for the lagged share of workers of retirement age is 1.29 (p-value less than .0005). Secondly, the tests of instruments quality (see Bound et al, 1995) are also passed: the partial R^2 is reasonably large, at 0.013, and the F-statistic strongly rejects the null that the instruments are jointly equal to zero.

In the main equation, we find that the education coefficient almost triples with respect to the previous results, increasing from .05 to .133 (14.2%), while it is still precisely determined (p-value of .019). Moreover, the over-identification test is not rejected, with a test statistic of 1.3 (p-value of .25). This is a reassuring result although it has to be taken into account bearing in mind that over-identification tests typically have low power.[17]

These results are also encouraging in that our estimate of a firm-level social return of 14.2% comfortably exceeds most international OLS estimates of private returns, even those for Portugal, a country which typically ranks at the top of the international distribution of those returns: Pereira and Martins (2001) estimate an OLS private return of between 8% and 11% over the 1991-98 period. Other studies include Vieira (1999), who follows the strategy of Harmon and Walker (1995) and documents IV estimates lower than the OLS ones, at around 5%, and Modesto (2003), who examines the self-selection involved in progressing or not from compulsory education and finds marginal returns at that stage not greater than 10%. (See also Martins and Pereira (2004), who present OLS and quantile regression results for comparable micro datasets covering sixteen Western countries. Portugal tops the international distribution, with a return at the mean of about 11%.)

This relatively large gap between firm- and individual-level returns supports the idea that private returns are not irrelevant from the social point of view, as in signalling models, and that there is a considerable additional spillover effect on top of the private return. Moreover, since we have reasons to believe that rent sharing is an important feature of the Portuguese labour market (Martins, 2004), this higher firm-level estimate is precisely the result expected given the model's implications in this type of labour markets if productivity spillovers are relevant in

[17] We have also considered different retirement-age thresholds (more than 60 or 63 years old) since early retirement applies in some cases. Our results remained largely unchanged. However, as expected, the strength of the instrument becomes weaker as we move farther from the 65 level (for men). These findings are available upon request.

practice. (Further work will examine in greater detail the link between labour market structure and wage spillovers.)

On the other hand, as we mentioned in the introduction, evidence of a stronger relationship between average schooling and average earnings than between individual schooling and individual earnings does not necessarily, in general, imply positive spillovers. At least in the context of more aggregate units of analysis, such as regions, imperfect substitution between educated and uneducated workers may also induce such result. In the case of such larger units of analysis, educational expansion may increase the earnings of uneducated workers not because they become more productive but just because they become scarcer.

Nonetheless, as we explained before, we consider that it is unlikely that this general-equilibrium effect will be relevant in firms, unlike in cities or regions, for instance. Indeed, the model assumes that wages for workers of different skills will be constant from the firms' point of view, since individual firms, as small units, cannot affect prices – wages may only change for the less skilled to the extent they benefit from on-the-job schooling spillovers and become more productive.

We now test these assumptions by contrasting how does pay for the more and less educated workers evolve as a function of average education in their firm. Our empirical approach is as follows: First, in order to make our estimates of the firm-level return to education less affected by the impact of entrants, we consider only workers that have been in the firm for at least 36 months. This period is, in general, the time threshold at which employment contracts have either to become permanent or be terminated. Workers with levels of tenure of 36 months or more will thus have a stronger degree of bargaining power, allowing them to benefit from any spillover that may occur, unlike those workers with temporary contracts.[18]

For these workers-stayers, we then consider two alternative thresholds between "educated" and "uneducated" workers, which are, for each firm-year, the respective mean and median levels of

[18] Using a subsample of the present data set and the same tenure threshold, Martins (2004) finds in his study of rent sharing that high-tenure workers benefit almost twice as much from firm rents than their low-tenure colleagues.

schooling.[19] After that, we separate workers whose education exceeds or is below each threshold and aggregate their characteristics (schooling, experience, etc) for each firm-year, after which we run similar regressions as before. This approach should result in clearer estimates of the spillovers, as we focus on the impact of average education (determined by both stayers and entrants) on educated and less educated stayers separately.

Our empirical model is an extended version of (14), now including the characteristics of each subset of workers (educated and less educated) plus the previously-used control for the average schooling across all workers:

(15) $y_{ijt} = \beta_1 educ_{it} + \beta_2 educ_{ijt} + X_{ijt}'\beta_3 + \alpha_i + \tau_t + \varepsilon_{it}$

y_{ijt} is the logarithm of the average earnings of workers of type j (educated or uneducated) in firm i in period t (that have been in the firm for at least 36 months). As before, $educ_{it}$ is the average level of schooling years of the workers of firm i in period t, regardless of whether they are stayers or new hires. X_{ijt} refer to the same set of average characteristics of the workers of type j in firm i in period t. The remaining variables have the same interpretation as before.

The results are presented in Table 4. With respect to the first-stage equations, we find little differences in the role of the instruments across the two sub-groups (educated and uneducated workers) and across the two education thresholds (mean and median).[20] More interestingly, we find for the main equation that the impact of firm average schooling is much greater for the uneducated stayers than for their educated counterparts. For instance, taking the mean-education threshold, an increase in firm average education of one year significantly increases uneducated workers wages by 0.024. The equivalent increase for educated workers is only 0.008 and not significant. Moreover, the same pattern is obtained for the median threshold, with a wage increases for the uneducated workers of 0.033 and an insignificant wage increase for the educated workers.

[19] It should be recalled that the subgroup of educated stayers has no immediate theoretical counterpart in the model of Section 2.

[20] One exception is that the average schooling of the uneducated workers plays a greater role in explaining total average schooling than the average schooling of the educated workers. This is due to the large positive skewness of the distribution of schooling within firms.

On the other hand, this pattern is reversed if we look at the impact of each groups' own schooling. While this impact is not significant for the uneducated workers, the return is significant and ranges between 0.075 and 0.077 for educated stayers.[21] Overall, these results are consistent with the model and, in particular, with the existence of spillovers for the less educated, as these uneducated workers benefit from the schooling of their co-workers, while the educated workers do not.

These results are also important in that they go against an alternative explanation, unrelated to spillovers, for the higher returns to education uncovered at the firm level in this paper. This alternative explanation is based on non-linear and, in particular, convex returns to education, which indeed have been documented for the Portuguese case (see Pereira and Martins, 2001). Under such non-linearities, returns at the firm level could exceed those at the individual level, as the former returns at the firm level, in within-firm estimations, can be more than proportionately driven by the inflow of more educated workers – who benefit from higher individual returns to education. However, the existence of spillovers to the less educated workers, documented in this subsection, indicates that, at the very least, the higher firm-level returns are not only a result explainable by non-linearities.

Further support for our findings can be found in the results of Silva (2003). Silva's paper draws on the same data used here to study *county-level* social returns to education, adopting an empirical approach based on displaced workers that move to different counties. Unlike in the our paper, he generally finds small or insignificant externalities. However, his estimates may be affected by measurement error, as county-level education variables obtainable from the "Quadros de Pessoal" data set cannot include important categories of workers, such as the self-employed or public servants, not to mention individuals outside the labour force.

In any case, in one specification which also controls for differences in education across counties, Silva (2003) documents positive and significant results for average firm education

[21] It should be mentioned that, in this specific approach, the instruments for retirement shares are generally not significant (and in some cases have negative signs). The over-identification test is, however, passed in all specifications.

(3rd row, Table 5, page 45). This is an important finding for our own results as it is consistent with our model of on-the-job schooling and, in particular, with the view that most of the education spillovers arise inside firms and not at more aggregate levels of analysis.

However, even for his coefficients on differences between firms average education levels (before and after worker displacement), Silva (2003) generally finds smaller coefficients than those presented in the next Section. This is also consistent with the model, which predicts asymmetric wage increases related to on-the-job schooling for job movers. Specifically, only if the new firm of a job mover has a higher level of average education will wages increase on account of education spillovers. If the new firm does not have a higher level of average education, then there is no scope for spillovers but the new workers will not be negatively affected either. Wages will then stay unchanged in this case.

Finally, as further evidence of robustness, we have replicated the analysis above for groups of firms of different sizes (results not shown but available upon request). We found returns always above 10% and some evidence that larger firms exhibit larger returns. This may suggest that the "spilloverability" of education is positively affected by firm size.

4.1. Extension

In this sub-section, we replicate our previous analysis of equation (14) but considering now information aggregated at different job levels within each firm, rather than at the firm level, as before. Our motivation for this exercise is that the model would predict a stronger spillover effect in this case, as there is greater scope for spillovers between educated and less educated workers within a job level, rather than across all job levels: on-the-job schooling is likely to be less non-excludable across than within job levels.

The "Quadros de Pessoal" data include information on eight types of job levels, ranging between apprentices and low-skilled blue-collar occupations to high- and top-level management (see a brief description in Appendix 2). This specific range of job levels, unchanged over the period covered, has to be adopted by all firms that submit their information to the Department of Employment and is thus generally comparable both across and within firms. The descriptive statistics for the resulting new data are presented in Table 5. (We considered only the seven job levels above apprenticeships, as the latter level presents considerable measurement error.) Notice the large increase in the number of observations, from 27,994 firms-year (Table 1) to 177,662 job-levels-firm-year. Notice also the increase in (unweighted) average education, as the thinner job levels (with fewer workers) typically include more educated individuals.

We then regress log average wages in each job-level/firm/year cell on the mean characteristics of that cell, considering also cell fixed effects and instrumenting education in a similar way than before. The results are presented in Table 6 and indicate a significant and precisely estimated return to average education of 0.209 (23%). This finding is consistent with our expectations under the framework of the on-the-job schooling model since it is considerably larger than our estimate for the firm-level analysis.

Other reassuring results are that the over-identification test is passed (p-value of 0.45) and the F-statistic of the instruments is very large. However, the indicators of instrument quality are not as good as before: the coefficient of retirement shares is not significant and the partial R^2 statistic is relatively low.[22]

4.2 Implications

Before concluding, we discuss in this subsection some implications of our findings. First, we provide a more directly interpretable measure of the economic impact of on-the-job schooling and its spillovers, as derived in this paper, by computing some simple, back-of-the-envelope estimates of how much that type of schooling affects wages. For this exercise we consider a

[22] One explanation for these latter findings is that average education at each job-level cell is subject to job upgrading processes which are not much affected by retirement-related forces. Moreover, measurement error is likely to be more acute within job-levels taken separately than together in firms, for instance because firms may occasionally change their coding practices (as to how to allocate each given worker to a job level, for instance). Promotions may also negatively affect the strength of the retirement instrument.

spillover effect of 7%, conservatively halfway between the 14% derived in our first estimation and the 23% obtained for job-level cells, after subtracting 10%, the latter figure corresponding approximately to the modal estimate of the OLS private returns to education for Portugal (see the references above).

We then borrow from the Lester range methodology, as discussed in the rent-sharing literature, and work out the percentage wage gain of an hypothetical worker that moved from a firm at the bottom of the distribution of average firm-level education (as proxied by the 10th percentile, which corresponds to 4.5 years of schooling) to a different firm at the of the top of the same distribution (90th percentile: 9.3 years of schooling). The two firms would have precisely the same average characteristics, except for the education of their workforce. For that spillover of 7%, the resulting wage gain would be 34%, a figure that can be regarded as considerably large.

What further implications can a figure of this size have? It may be relevant for research that seeks to understand the increasing levels of within wage inequality observed in some countries, including the U.S. and the U.K., particularly during the 1980's (Katz and Autor, 1999). For instance, a process of increasing education dispersion within firms (which may or not have corresponded to the case of those countries) would, according to the model and assuming that productivity spillovers lead to wage spillovers, increase within wage inequality. This is because workers in firms that bring in more educated workers would see their wages increase unlike workers in firms whose workforce's average educational attainment stays unchanged.

On-the-job schooling can also be interpreted in the context of the emerging literature on education and wage risk (see Pereira and Martins, 2002, Carneiro et al., 2003, and Hogan and Walker, 2003, for some recent contributions). Our model and findings suggest that education can lead to wage uncertainy not only in the sense that, for instance, the returns to investments in one's education cannot be fully predicted but also because one's wages will be affected by the education of the colleagues one is paired with.

As to policy implications, our findings support the case for the public funding of (higher) education. Its benefits fall not only on the individuals that acquire those skills directly at schools but also on those persons that are "spilled over" at work. With respect to the evaluation of different labour market types, our findings are however less standard. Indeed, our results indicate that there will be less on-the-job schooling under competitive markets, as competition prevents employers from fully benefiting from their hiring of educated workers.

Finally, we also derive some results about the scope for education to generate endogenous growth, which are however less straightforward. On the one hand, the external benefit of education increases with the levels of education. On the other hand, increasing levels of education imply that the relative share of individuals that benefit from those external effects is increasingly smaller. External effects would then disappear in a possible long-run scenario in which all individuals have similarly high levels of education.

5. Conclusions

We contribute to the literature on social returns to education by putting forward a model of on-the-job schooling and testing empirically some of its implications. This model is shown to lead to a stronger relationship between wages and education at the firm level than at the individual level, at least in non-monopsonistic labour markets. The gap between the individual and firm level results is also shown to depend positively on the size of the education spillover. The empirical results are then based on the estimation of Mincer firm-level wage equations applied to a large Portuguese matched employer-employee panel. Endogeneity concerns, clarified by the model, led us to control for firm fixed effects and to instrument firm average education.

Consistently with the predictions of the model, we find firm-level returns to education significantly above their individual-level counterparts. We also find evidence of significant wage spillovers to less-educated workers: their pay increases by 2% to 3% per extra year of education of workers in their firm. However, the subset of educated incumbent workers does not seem to benefit from such spillovers, a result again predicted by the model. Finally, the education spillover is stronger when considering a more disaggregated level of analysis (job levels within firms), within which one expects stronger interactions among workers.

Taken as a whole, our evidence indicates that education has a significant external effect on productivity and wages within firms, implying social returns to education greater than private returns. More specifically, the results suggest that there is a multiplier effect in the provision of education, as its benefits are not only circumscribed to the individuals that invest in their own human capital but also on the workers that have not made that investment at school but then go on to interact with educated colleagues at their place of work.

References

Acemoglu, Daron (1996) "A Microfoundation for Social Increasing Returns in Human Capital Accumulation", Quarterly Journal of Economics, 111, 779-804.

Acemoglu, Daron, and Jörn-Steffen Pischke (1999a) "The Structure of Wages and Investment in General Training", Journal of Political Economy, 107, 539-572.

Acemoglu, Daron, and Jörn-Steffen Pischke (1999b) "Beyond Becker: Training in Imperfect Labour Markets", Economic Journal, 109, F112-F142.

Acemoglu, Daron and Joshua Angrist (2000) "How Large are Human Capital Externalities? Evidence from Compulsory Schooling Laws", NBER Macroeconomics Annual, 9-59.

Arai, Mahmood (2003) "Wages, Profits and Capital Intensity: Evidence from Matched Worker-Firm Data", Journal of Labor Economics, 21, 593-618.

Ashenfelter, Orley, and Alan Krueger (1994) "Estimates of the Economic Returns to Schooling from a New Sample of Twins", American Economic Review, 84, 1157-73.

Barth, Erling (2002) "Spillover Effects of Education on Co-Worker Productivity. Evidence from the Wage Structure", paper presented at the European Society of Population Economics Annual Conference, Bilbao.

Barth, Erling and Harald Dale-Olsen (2003) "Assortative Matching in the Labour Market? Stylised Facts About Workers and Plants", paper presented at the Comparative Analysis of Enterprise Data Conference, London.

Battu, Harminder, Clive Belfield, and Peter Sloane (2003) "Human Capital Spillovers within the Workplace: Evidence for Great Britain", Oxford Bulletin of Economics and Statistics, 65, 575-594.

Becker, Gary (1975) "Human Capital: A Theoretical and Empirical Analysis, With Special Reference to Education", University of Chicago Press, Chicago.

Behrman, Jere and Mark Rosenzweig (2002) "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?", American Economic Review, 92, 323-334.

Blanchflower, David, Andrew Oswald, and Peter Sanfey (1996) "Wages, Profits and Rent-Sharing", Quarterly Journal of Economics, 111, 227-252.

Bound, John, David Jaeger and Regina Baker (1995) "Problems with Instrumental Variables Estimation when the Correlation between the Instrument and the Endogenous Explanatory Variable Is Weak", Journal of the American Statistical Association, 90, 443-450.

Carneiro, Pedro and James Heckman (2003) "Human Capital Policy", IZA Discussion Paper 821.

Carneiro, Pedro, Karsten Hansen, and James Heckman (2003) "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice", International Economic Review, 44, 361-422.

Ciccone, Antonio and Giovanni Peri (2002) "Identifying Human Capital Externalities: Theory with an Application to US Cities", IZA Discussion Paper 488.

Currie, Janet and Enrico Moretti (2003) "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings", Quarterly Journal of Economics, 118, 1495-1532.

Gemmell, Norman (1997) "Externalities to Higher Education", Report to the National Committee of Inquiry into Higher Education, St Clements House, Norwich.

Griliches, Zvi and Jerry A. Hausman (1985). "Errors in Variables in Panel Data: A Note with an Example", Journal of Econometrics, 31, 93-118.

Harmon, Colm and Ian Walker (1995) "Estimates of the Economic Return to Schooling for the United Kingdom", American Economic Review, 85 (1995), 1278-1286

Hildreth, Andrew and Andrew Oswald (1997) "Rent Sharing and Wages: Evidence from Company and Establishment Panels", Journal of Labor Economics, 15, 318-337.

Hogan, Vincent and Ian Walker (2003) "Education Choice Under Uncertainty and Public Policy", University College Dublin, mimeo.

Katz, Lawrence and David Autor (1999) "Changes in the Wage Structure and Earnings Inequality," in Handbook of Labor Economics, O. Ashenfelter eds., volume 3A, 1463-1555, North-Holland, Amsterdam.

Kramarz, Francis (2003) "Wages and International Trade", CREST, mimeo.

Krueger, Alan and Mikael Lindahl (2001) "Education for Growth: Why and for Whom?", Journal of Economic Literature, 39, 1101-1136.

Lochner, Lance and Enrico Moretti (2004) "The Effect of Education on Criminal Activity: Evidence from Prison Inmates, Arrests and Self-Reports", American Economic Review, 94, 155-189.

Lucas, Robert (1988) "On the Mechanics of Economic Development", Journal of Monetary Economics, 22, 3-42.

Manning, Alan (2003) "Monopsony in Motion. Imperfect Competition in Labor Markets", Princeton University Press, Princeton.

Martins, Pedro (2004) "Rent Sharing Before and After the Wage Bill", Queen Mary, University of London, mimeo.

Martins, Pedro and Pedro Pereira (2004) "Does Education Reduce Wage Inequality? Quantile Regression Evidence from 16 Countries", Labour Economics, 11, 355-371.

Mincer, Jacob (1974) "Schooling, Experience and Earnings", National Bureau of Economic Research, New York.

Milligan, Kevin, Enrico Moretti and Philip Oreopoulos (2004) "Does Education Improve Citizenship? Evidence from the U.S. and the U.K.", Journal of Public Economics, 88, 1667-1695.

Modesto, Leonor (2003) "Should I stay or should I go? Educational choices and earnings: An empirical study for Portugal", Journal of Population Economics, 16, 307–322.

Moretti, Enrico (2004a) "Estimating the Social Return to Higher Education: Evidence from Longitudinal and Repeated Cross-Sectional Data", Journal of Econometrics, 121, 175-212.

Moretti, Enrico (2004b) "Workers' Education, Spillovers and Productivity: Evidence from Plant-Level Production Functions", American Economic Review, 94, 565-690.

Pereira, Pedro and Pedro Martins (2001) "Portugal", in: C. Harmon, I. Walker and N. Westergaard-Nielsen, eds, *Education and Earnings in Europe – a Cross Country Analysis of Returns to Education*. Edward Elgar, Cheltenham.

Pereira, Pedro and Pedro Martins (2002) “Is There A Return-Risk Link in Education?”, Economics Letters, 75, 31-37, 2002.

Rauch, James (1993) “Productivity Gains from Geographic Concentration of Human Capital: Evidence from the Cities”, Journal of Urban Economics, 34, 380-400.

Rudd, Jeremy (2000) “Empirical Evidence on Human Capital Spillovers”, Federal Reserve Board Working Paper.

Silva, João Cerejeira (2003) “Local Human Capital Externalities or Sorting? Evidence from a Displaced Workers Sample”, Universidade do Minho Working Paper

Vieira, José Cabral (1999) “Returns to Education in Portugal”, Labour Economics, 6, 535-541.

Tables

Table 1 - Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Hourly earnings	27,994	3.75	2.47	0.92	87.16
Log Hourly Earnings	27,994	1.08	0.47	-0.11	4.41
Education	27,994	6.45	2.00	0.00	16.65
Experience	27,994	23.38	6.05	3.72	43.65
Experience^2	27,994	712.95	310.96	23.25	1942.44
Tenure	27,994	108.18	63.65	0.00	357.16
Tenure^2	27,994	231.36	218.91	0.00	1402.50
Female	27,994	0.42	0.30	0.00	1.00
Firm Size	27,994	213.41	615.13	50	29433
Age	27,994	35.83	5.72	18.51	53.70
Share Retirement Age	27,994	0.01	0.02	0.00	0.21
1991	27,994	0.10	0.29		
1992	27,994	0.10	0.30		
1993	27,994	0.11	0.31		
1994	27,994	0.11	0.32		
1995	27,994	0.12	0.33		
1996	27,994	0.12	0.32		
1997	27,994	0.12	0.33		
1998	27,994	0.12	0.32		
1999	27,994	0.10	0.31		
Lagged Education	23,164	6.38	1.98	0.00	16.65
Lagged Share Retirement	23,164	0.01	0.02	0.00	0.21

Table 2 - Results

	Pooled OLS		Random Effects		Fixed Effects	
	Coeff.	**St. Error**	**Coeff.**	**St. Error**	**Coeff.**	**St. Error**
Schooling	0.178**	0.002	0.135**	0.003	0.050**	0.003
Experience	0,072**	0.004	0.066**	0.002	0.035**	0.003
Experience^2	-0.001**	0.000	-0.001**	0.000	-0.001**	0.000
Female	-0.348**	0.009	-0.318**	0.009	-0.159**	0.021
Log Size	0.054**	0.004	0.047**	0.003	-0.003	0.005
Adj. R^2	0.7836					
Firms-year	27,994		27,994		27,994	

Notes:

All regressions include a quadratic on tenure and year dummies.

The Hausman test about the difference between the random and fixed effects models is strongly rejected.

* - significant at the 5% level

** - significant at the 1% level

Table 3 - Results, Fixed Effects and Instruments

	Coeff.	St. Error
First Stage		
Lagged Schooling	0.082**	0.005
Share of 65 and over	1.289**	0.246
Adjusted R^2	0.5186	
Partial R^2	0.0131	
F-statistic	153.14	(P-value= 0,000)
Main Equation		
Schooling	0.133**	0.019
Experience	0.066**	0.008
Experience^2	-0.001**	0.000
Female	-0.128**	0.017
Log Size	0.021**	0.007
Within R^2	0.5302	
Between R^2	0.2502	
Overall R^2	0.2529	
Overid. Test Statistic	1.307	(P-value= 0.253)
Observations	23,164	

Notes:
* - significant at the 5% level
** - significant at the 1% level

Table 4 - Results, Different Sub-Samples

	Mean Education		Median Education	
	Coeff.	St. Error	Coeff.	St. Error
Uneducated workers (stayers)				
First Stage				
Group Average Schooling	0.544**	0.011	0.596**	0.010
Lagged Total Average Schooling	0.168**	0.006	0.152**	0.006
Share of 65 and over	-0.170	0.279	-0.144	0.270
Main Equation				
Total Average Schooling	0.024**	0.011	0.033**	0.012
Group Average Schooling	-0.008	0.007	0.006	0.008
Overall R^2	0.3928		0.5062	
Observations	22,841		22,883	
Groups	4,824		4,828	
Educated workers (stayers)				
First Stage				
Group Average Schooling	0.021**	0.005	0.013**	0.005
Lagged Total Average Schooling	0.179**	0.006	0.188**	0.006
Share of 65 and over	-0.020	0.296	0.004	0.294
Main Equation				
Total Average Schooling	0.008	0.012	-0.003	0.012
Group Average Schooling	0.075**	0.002	0.077**	0.002
Overall R^2	0.7315		0.6504	
Observations	22,771		22,578	
Groups	4,820		4,804	

Notes:
Both equations consider the same additional variables as in Table 2, although now they refer to each specific subset of workers (educated and uneducated), and not to the entire firm.
For each period, only workers present in the firm in the current and previous period ("stayers") are considered, except in the total average schooling variable.

* - significant at the 5% level
** - significant at the 1% level

Table 5 - Descriptive Statistics (Firm/Job-levels)

Variable	Cells	Mean	Std. Dev.	Min	Max
Hourly earnings	177,662	4.49	3.86	0.51	198.08
Log Hourly Earnings	177,662	1.24	-0.61	0.67	5.29
Education	177,662	7.43	3.36	0.00	17.00
Experience	177,662	23.82	10.30	0.00	76.00
Experience2	177,662	751.84	542.17	0.00	5776.00
Tenure	177,662	112.62	89.58	0.00	758.00
Tenure2	177,662	254.57	340.69	0.00	5745.64
Female	177,662	0.37	0.37	0.00	1.00
Firm Size	177,662	231.76	647.31	50	29433
Age	177,662	37.25	9.59	14.00	87.00
Share Retirement Age	177,662	0.02	0.08	0.00	1.00
1991	177,662	0.09	0.29		
1992	177,662	0.10	0.30		
1993	177,662	0.10	0.31		
1994	177,662	0.11	0.31		
1995	177,662	0.12	0.33		
1996	177,662	0.12	0.33		
1997	177,662	0.12	0.33		
1998	177,662	0.12	0.32		
1999	177,662	0.11	0.31		
Lagged Education	142,176	7.30	3.31	0.00	17.00
Lagged Share Retirement	142,176	0.02	0.08	0.00	1.00

Table 6 - Results, Fixed Effects and Instruments (Job Levels)

	Coeff.	St. Error
First Stage		
Lagged Schooling	0.015**	0.002
Share of 65 and over	-0.062	0.052
Adjusted R^2	0.3503	
Partial R^2	0.0003	
F-statistic	23.94	(P-value= 0,000)
Main Equation		
Schooling	0.209**	0.037
Experience	0.069**	0.010
Experience^2	-0.001**	0.000
Female	-0.180**	0.010
Log Size	0.041**	0.005
Within R^2		
Between R^2	0.6853	
Overall R^2	0.6515	
Overid. Test Statistic	0.58	(P-value= 0.446)
Observations	142,176	

Notes:
Both equations consider the same additional variables, as in Table 2.
* - significant at the 5% level
** - significant at the 1% level

Appendix 1 – Proofs of some results.

A. Proof of Result 3

It was derived before that $\overline{ed}_P < \overline{ed}_M$. This implies that $\overline{ed}_P W_S + (1-\overline{ed}_P)W_U < \overline{ed}_M W_S + (1-\overline{ed}_M)W_U$ and that $W_1 = \frac{W_S + W_U}{\overline{ed}_P W_S + (1-\overline{ed}_P)W_S} > \frac{W_S + W_U}{\overline{ed}_M W_S + (1-\overline{ed}_M)W_S} = W_0$. Taking $W_2 = \frac{W_S + W_U \gamma L_S}{\overline{ed}_P W_S + (1-\overline{ed}_P)W_S \gamma L_S}$, then $\frac{dW_2}{d\gamma L_S} = \frac{W_U W_S (2\overline{ed}_P - 1)}{(\overline{ed}_P W_S + (1-\overline{ed}_P)W_U \gamma L_S)^2}$. This derivative is positive if $(2\overline{ed}_P - 1) > 0$, which is the case if $(\overline{ed}_P - 1) > 0.5$. The last inequality will hold only if $L_S > 1$, which is true by assumption, implying that $W_2 > W_1$. Since $W_1 > W_0 >$(8) and since $W_2 =$(12), then it must be the case that (12) > (8). Q.E.D.

B. Proof of Result 4

The sign of the partial derivative of (12), $\frac{\partial \ln \overline{W}^P}{\partial \overline{ed}^P}$, with respect to γ is $sign\left[\frac{\frac{\partial L^P}{\partial \gamma} L^P - \frac{\partial L^P}{\partial \gamma}(1+L^P)}{{L^P}^2}\right] = -sign\left(\frac{\partial L^P}{\partial \gamma}\right)$. This is then equal to

$$-sign\left[\frac{W_U (W_S + \gamma W_U)^{\frac{1}{\alpha+\beta-1}-1} \frac{1}{\alpha+\beta-1}\left[\lambda^{\alpha}(\alpha+\beta)\right]^{\frac{1}{1-(\alpha+\beta)}} - \frac{\alpha}{1-(\alpha+\beta)} \gamma^{\frac{\alpha}{1-(\alpha+\beta)}-1} (W_S + \gamma W_U)^{\frac{1}{\alpha+\beta-1}} (\alpha+\beta)^{\frac{1}{1-(\alpha+\beta)}}}{\left[\gamma^{\alpha}(\alpha+\beta)\right]^{\frac{2}{\alpha+\beta-1}}}\right]$$

which is positive. Q.E.D.

Appendix 2 – Hierarchical Levels

Table A.1 - Hierarchical Levels

Level	Tasks
8 - Top Managers	Definition of the firm's general policy. Strategic planning Creation or adaptation of technical and administrative methods.
7 - Intermediary Managers	Organisation and adaptation of the guidelines established by the superiors.
6 - Supervisors, team leaders	Orientation of teams, requiring the knowledge of action processes.
5 - Higher-skilled professionals	Tasks requiring a high technical value and defined in general terms by the superiors.
4 - Skilled professionals	Complex or delicate tasks, usually not repetitive.
3 - Semi-skilled professionals	Well defined tasks, mainly manual or mechanical with low complexity.
2 - Non-skilled professionals	Simple tasks and totally determined.
1 - Apprentices, interns, trainees	

Source: Grade levels as defined by law.