A review of the wage returns to private sector training$^1$

Edwin Leuven

*Universiteit van Amsterdam* and *CREST-INSEE*

---

$^1$This paper was prepared for the joint EC-OECD Seminar on Human Capital and Labour Market Performance, held in Brussels on Dec 8, 2004. The opinions expressed do not necessarily reflect those of the EC, OECD, or of their member countries. This paper was funded by the EC and commissioned by the OECD.
Contents

1 Introduction 1

2 Theoretical insights 2

3 Measurement of training 3

4 Estimation issues 5
  4.1 Evaluation framework ................................. 5
  4.2 OLS and FE implementations .......................... 6
  4.3 IV and parametric selection models .................. 8

5 Overview of the literature 9
  5.1 Returns to tenure (indirect training measures) ........ 9
  5.2 Direct training measures ............................... 9
    5.2.1 United States & Canada ........................... 9
    5.2.2 UK .............................................. 12
    5.2.3 Other countries .................................. 15
  5.3 High returns? ........................................ 17

6 Discussion and conclusion 18

References 19
Abstract

This paper provides a tentative review of the literature that estimates wage returns to training. It discusses both the measurement and estimation issues. The fundamental problem concerning the recovery of the causal effect of training on earnings lies in the correction for selectivity into training. The discussion of the empirical literature emphasises the size of the estimated returns; something which has been largely neglected in the literature. It is argued that traditional studies that depend on differencing (fixed effect) methods where non-participants are used as a comparison group results in high return estimates.

On the basis of these high returns some have argued that there is substantial underinvestment and therefore scope for public intervention. Such underinvestment could arise because of for example hold-up or liquidity constraints. The paper then shows that studies which exploit arguably exogenous variation in training participation find much smaller wage effects of training.

The evidence collected in this paper calls into question the case for underinvestment based on return studies. There is other evidence which suggests that underinvestment may be less severe than previously believed. First, recent literature in experimental economics shows that individuals are often motivated by reciprocity and fairness considerations which are typically ignored in standard human capital models, while reciprocity can alleviate underinvestment. Second, recent literature emphasising market imperfections also shows that imperfections may give employers more incentives than previously thought to invest in the general training of their employees.

The paper concludes by calling for further study of training mechanisms and subsequent outcomes. Some potential avenues for research in this area are given.
1 Introduction

There is a steadily growing literature that investigates the determinants of individual wage growth over the lifecycle. Understanding wage growth is important for a number of reasons. First, wages are the major determinant of individual welfare. Second, because observed wage growth can have several theoretical explanations and distinguishing between these increases our understanding of the functioning of labour markets. Finally, often mentioned, reason is that it is informative about the extent to which wages are tied to jobs, which in turn is a measure of the cost of worker displacement.

The two basic wage growth patterns that have been studied in this context are (i) how wages increase with experience, and (ii) how wages increase with tenure keeping experience constant. These observed patterns can however have many sources. A first explanation is human capital based and conjectures that wage growth mirrors productivity growth. A second mechanism that generates wage growth is job search and matching Jovanovic (1979a,b); Mortensen (1978). The intuition is simple, good matches are more likely to survive than bad matches, as a consequence “older” matches are on average better which generates a positive relation between experience/tenure and wages. All wage growth comes from mobility among jobs rather than increase in productivity because of human capital investment. Finally, contractual considerations can cause wage profiles to slope upward because postponing rewards can provide an incentive to workers to exert effort early on.

Distinguishing between these alternative explanations of lifecycle wage patterns has proven to be a daunting task. There seems however to be consensus that there are true returns to tenure, although these seem to be much lower for low-skilled workers. At the same time Brown (1989) finds that firm-specific wage growth occurs mainly during periods of on-the-job training. This latter finding emphasises the importance of studying the incidence and effects of more direct measures of training instead of tenure which, in the end, is just a proxy for (firm-specific) human capital.

This chapter therefore reviews the literature on the wage effects of formal private sector training. Many studies find very large returns, several order of mag-
nitudes larger than returns to regular education. Some have therefore argued that this points to underinvestment in training and is evidence for liquidity constraints or hold-up. The review will highlight the endogeneity problems that hamper the identification of causal effects of training. It will discuss at some length recent studies that put more effort in solving for the selectivity into training and consequently find much smaller returns. The chapter will then conclude that the case for underinvestment is weaker than previously thought and will discuss some alternative mechanisms that might explain this result.

The outline of this chapter is as follows. The following section briefly discusses recent insights from the theoretical literature that studies training. Section 3 then turns to the operationalization of training on an empirical level. Section 4 then outlines the estimation issues surrounding the recovery of causal effects of training on wages. Section 5 then briefly summarizes the standard literature based on fixed effects models and then more extensively looks at two recent studies that take an alternative identification approach using Dutch data. Finally section 6 concludes and discusses the implications for the case for underinvestment.

2 Theoretical insights

Standard competitive theory, as put forward by Becker (1962), distinguishes between general training and specific training. General training is of equal value in many firms whereas specific training is only useful in one firm. In this competitive world workers reap all the returns to general training and consequently finance it, either directly or through lower wages.

Underinvestment in general training occurs therefore only if workers are liquidity constrained. Since firms will not finance general training, the negative poaching externality in which firms underinvest in general training because of the poaching of trained workers by other firms disappears. Finally, firms finance specific on-the-job training but might let workers share in the returns to reduce inefficient turnover.

The recent literature demonstrates how market imperfections may render training that is technologically general de facto specific because wages will be less than marginal product (Stevens 1994; Acemoglu and Pischke 1999). This restores
investment incentives for technologically general training for the employer and may alleviate underinvestment in such training. The poaching externality however reappears since any source of imperfect competition leading to wages below marginal product, combined with any source of uncertainty about labour turnover makes that the worker and the firm do not internalise positive externalities and underinvest (Stevens, 1994).

3 Measurement of training

Training and, more in general, the stock of human capital is difficult to measure. Initially, studies of training used labor market experience as a proxy for general training and job tenure as a proxy for specific training. In the 1980s datasets containing self-reported measures of training became available. These datasets were typically based on household surveys such as the Current Population Survey (CPS) and the National Longitudinal Survey of Youth (NLSY) in which respondents were asked whether they participated in some form of training in a specific reference period. In the CPS this is the period since the start of the respondents current job, while in the NLSY this is the period since the last interview. Apart from these household surveys, employer based (see Barron et al. (1987)) and matched employer-employee surveys (Lynch and Black (1998) for example) are also slowly becoming available. Finally there are a few studies that use administrative data from a single firm (e.g. Bartel (1995)).

First there is the issue whether to collect stock or flow measures. Most surveys collect flow measures, namely: the amount of training over a particular period. In the CPS on the other hand the reference period covers the period since the start of the job. This implies that if all reported training is specific this question would measure the stock of training.

The responses to these training questions are sensitive to the period covered by the training questions (e.g. Loewenstein and Spletzer, 1997), the longer the reference period the more training will be reported. A factor that is likely to in-

---

1An early attempt at measuring training more directly is found in the Michigan Panel Study of Income Dynamics (PSID). Respondents were asked: “on a job like yours, how much time would it take the average new person to become fully trained and qualified?” This seems to measure a characteristic of the job instead of the amount of training the respondent participated in.
roduce measurement error is the retrospectiveness of these self-reported training measures. Recollection problems are expected to increase both with the span of time between the training spell and the interview, and the detail of the training questions. Training questions that measure flows are therefore probably more accurate than training questions that attempt to measure stocks.

Surveys often ask about training incidence, but increasingly try to measure the length of training spells in an attempt to more accurately measure training effort. The true costs of training are, however, difficult to measure. First, respondents are unlikely to be fully aware of the opportunity cost of training. To the extent that the employer paid for training, the respondents are also unlikely to have information on the direct (monetary) cost of training.

Training is almost inherently heterogeneous and some aggregation is therefore inevitable. The aggregation implicit in the training questions varies between surveys. The types of training measured by these surveys therefore vary and are typically derived from the institutional setting, and often combine mode of delivery and provider. The NLSY, for example, asks about: training followed at business schools, apprenticeship programs, vocational or technical institutes, but also about correspondence courses, formal company training run by the employer, military training, seminars or training programs run at work by someone other than employer, seminars or training programs outside of work, and training given by vocational rehabilitation centers. It is not immediately clear to what extent this is an economically sensible classification.\(^2\)

The above illustrates the conceptual and practical complexity of collecting information on training. Little is known about the extent to which these conceptual measurement problems lead to actual measurement error. The only study to date hinting at this is Barron et al. (1997). These authors use data from a matched employer-employee survey dataset to see to what extent employer and employee responses are consistent. This approach has of course the drawback that both the worker and the establishment are likely to report training with error. The ‘truth’ is therefore not observed and the extent of measurement error cannot be determined. Barron et al. (1997) find that correlations between worker and establishment mea-

\(^2\)Apart from formal training there is a growing interest in measuring informal training (e.g. Barron et al., 1997; Frazis et al., 1996; Loewenstein and Spletzer, 1999b).
sures are less than 0.5 and that establishments report 25 percent more hours of training on average than do workers. On average, incidence rates are similar between worker and establishment reports, although 30 percent disagree on whether on-site formal training occurred.

These results suggest that training is measured with substantial error. If measurement error is ’classical’ (is additive and independent of the true value) then this will result in downward biased estimates. But since training has a mixed discrete continuous character the effect is less clear cut. Frazis and Loewenstein (2003) show that in specifications that include both a participation indicator and training duration, the coefficient on participation will be upward biased and the coefficient on duration downward biased as in the classical case.

4 Estimation issues

This section discusses the estimation issues involved in return studies. First the fundamental evaluation problem is briefly explained, after which the methods that are employed in the literature are discussed.

4.1 Evaluation framework

There is by now a large literature that estimates returns to private sector training. The fundamental problem in estimating such returns is treated extensively in the evaluation literature. The basic framework is as follows. Assume training is mere participation and denoted by \( d_i \) a binary indicator variable. The outcome is the wage. Now denote the wage with training by \( w_i(1) \) and outcome without training by \( w_i(0) \). We observe

\[
w_i = d_iw_i(1) + (1 - d_i)w_i(0)
\]

We can now define several parameters of interest. A common one is the average effect of training on the trained (ATT):

\[
E[w_i(1) - w_i(0)|d_i = 1] = E[w_i(1)|d_i = 1] - E[w_i(0)|d_i = 1]
\]

(1)
If training is exogenous in the sense that $d_i$ is uncorrelated with $w_i(0)$ then we can estimate the ATT simply by contrasting mean outcomes of the trained and non-trained as follows.

$$E[w_i(1) - w_i(0)|d_i = 1] = E[w_i|d_i = 1] - E[w_i|d_i = 0]$$

If $d_i$ is also uncorrelated with $w_i(1)$ then this is also the average treatment effect $E[w_i(1) - w_i(0)]$, otherwise we would need to estimate $E[w_i(1)|d_i = 0]$. In practice training is unlikely to be exogenous. Estimating causal effects then becomes a challenge. Wherein lies the problem? The first term on the right hand side of (1) is readily observed since it is merely the average wage among the training participants. The second term $E[w_i(0)|d_i = 1]$ is however more evasive since it is the wage that participants would have earned if they would not have participated in training. This is however never observed. Recovering this counterfactual is the fundamental evaluation problem.

### 4.2 OLS and FE implementations

The prototypical equation that the literature estimates takes the following form

$$\ln w_{it} = x_{it}'\beta + \gamma d_{it} + \epsilon_{it}$$

(2)

Where $w_{it}$ is the wage of individual $i$ at time $t$, $x_{it}$ is a vector of control variables, $d_{it}$ a measure of the stock of training and $\epsilon_{it}$ the residual/error term.

Note that this implicitly assumes that

$$w_i(0) = x_{it}'\beta, w_i(1) = x_{it}'\beta + \gamma$$

and therefore that training is exogenous conditional on $x_i$.

Early studies estimated (2) by ordinary least squares (OLS). The important drawback of OLS is that it only gives an unbiased estimate of $\gamma$ if training is uncorrelated with the error term: $E[d_{it}\epsilon_{it}] = 0$. OLS therefore ignores the possibility that there are unobserved individual characteristics such as ability that affect wages and correlate with training.
Without covariates \( x_{it} \) and \( d_{it} \) mere participation, the OLS estimate of the effect of training is simply the difference in mean earnings between those that participated in training and those that did not.

The current state of affairs in the literature is to estimate fixed effect versions of (2) where it is assumed that \( \epsilon_{it} = c_i + u_{it} \) so that the estimation equation now becomes

\[
\ln w_{it} = x_{it}' \beta + \gamma d_{it} + c_i + u_{it}
\]  

(3)

This effectively estimates (2) in deviation of individual means and is comparable to estimating (2) in first differences.\(^3\) The fixed effects estimator (3) takes into account any confounding influence of unobserved individual characteristics that correlate both with wages and training as long as these are fixed over time (they are picked up by \( c_i \)). The resulting estimate of \( \gamma \) is now the difference in wage growth between those that trained and those that did not. The crucial identifying assumption here is thus that participants would have experienced the same wage growth as non-participants in the absence of training. Some studies (f.e. OECD (2004); Loewenstein and Spletzer (1998, 1999a)) have estimated equations that control for match-specific effects. These fixed effects estimates are within-job estimates and therefore do not capture returns to training in the form of mobility to better jobs. Return estimates from these type of studies are therefore expected to be lower than returns estimates from standard fixed effects models.

A few studies (Pischke, 2001; Frazis and Loewenstein, 2003) have recognized that training participants may experience higher wage growth in the absence of training than non-participants. If this is the case, then standard fixed effect estimates will be biased. To take this into account Pischke (2001); Frazis and Loewenstein (2003) estimated fixed effect growth equations by adding individual specific growth rates of earnings \( \delta_{it} \) to equation (3). For estimation purposes we now need three periods on each individual and sufficient variability in training receipt.

\(^3\)With \( T = 2 \) first differencing and fixed estimation are equivalent. With \( T > 2 \) they are not. Which of the two is more efficient depends on the properties of \( u_{it} \). If \( u_{it} \) is i.i.d. FE is more efficient and if \( u_{it} \) follows a random walk first differencing is more efficient.
4.3 **IV and parametric selection models**

As an alternative to fixed effect based approaches a number of papers have estimated $\gamma$ using selection models or instrumental variables (IV). The outcome equation (2) is now augmented with a participation equation

$$d^*_it = w^*_i \eta + \nu_{it}$$

Selection models specify a joint parametric distribution for $(\epsilon, \upsilon)$ and can be estimated using maximum likelihood. An alternative is a 2-step method where in the first step a control function $\hat{\lambda}_{it} = \lambda (w^*_it \hat{\eta})$ is estimated such that $E[d_{it}(\rho \hat{\lambda}_{it} + \epsilon_{it})] = 0$. In the second step equation (2) is augmented with the control function, and OLS then gives a consistent estimate of the effect of training on wages $\gamma$.

It has been pointed out that selection models and control function methods can be very sensitive to misspecification of the joint distribution of $(\epsilon, \upsilon)$ and are identified exclusively on functional form and distributional assumptions unless $w_{it}$ includes variables that are not included in $x_{it}$. Unfortunately not any variable will do. What is needed is a variable $z_{it}$ that affects participation but is orthogonal to the error term in (2): $E[z_{it}\epsilon_{it}] = 0$. This is commonly referred to as an exclusion restriction, and to $z_{it}$ as an instrumental variable. In addition $z_{it}$ will need to have a significant effect on participation. With an instrument a common approach is to estimate $\gamma$ using two-stage least squares (2SLS). This avoids the parametric restrictions of selection models but does not allow extrapolation.\(^4\) A final thing to note is that when treatment effects are heterogeneous, the estimated effect is a so called local average treatment effect (Imbens and Angrist, 1994). This is the average effect for the implicitly defined subpopulation that is induced to participate by the exogenous variation generated by the instrument $z_{it}$.

\(^4\)Extrapolation in the selection model is based purely on distributional and functional form restrictions.
5 Overview of the literature

5.1 Returns to tenure (indirect training measures)

The initial literature that estimated wage returns to training was based on indirect training measures. These studies regress wages on labor market experience and job tenure (seniority). The coefficient on labor market experience is then interpreted as the return to general training, whereas the coefficient of job tenure is interpreted as the return to specific human capital. Abraham and Farber (1987), Altonji and Shakotko (1987) and Topel (1986) are early attempts to estimate the return to seniority up and above the return to experience. They find only small effects of seniority on wage growth. Topel (1991) reexamined the data and concluded that the findings in these studies are biased because of measurement error and selectivity issues. He finds that 10 years of current job seniority raises the wage of a typical male worker in the U.S. by 25 percent.

Human capital theory predicts upward sloping productivity profiles. Wage profiles are assumed to proxy these productivity profiles. There are several other theories (e.g. deferred compensation, self-selection, and matching theories) besides human capital theory that predict upward sloping wage profiles, and as such it is hard to argue that this is a definitive test. One would like to know to what extent wage growth correlates with productivity growth. Medoff and Abraham (1981) and Medoff and Abraham (1980) use performance ratings among professional and managerial employees in three U.S. corporations. Medoff and Abraham do not find any statistically significant correlation between these ratings and wage growth. They conclude that the on-the-job training model explains only a small part of the observed return to labor market experience. This result rests on the assumption that these ordinal performance ratings are unbiased measures of productivity.

5.2 Direct training measures

5.2.1 United States & Canada  For the US training returns have been estimated using a number of datasets. Some of these, such as the Panel Study of Income Dynamics (PSID) or the Current Population Survey (CPS), have very limited infor-
information on training. The most widely used datasets are the Employment Opportunity Pilot Project Survey (EOPP) and especially the National Longitudinal Survey of Youth Cohort (NLSY) which has arguably the most precise and comprehensive information on training.

Lynch (1992) is one of the first studies that uses the early waves of the NLSY (1981 and 1983) to estimate wage returns. The subsample she considers are those who did not graduate from college and finished schooling by the 1980 interview date. She presents return estimates both using Heckman two-step selectivity corrections and fixed effects regressions. In the estimations she controls for tenure, experience and personal and job characteristics. The two-step estimates show that a week of company training (completed or uncompleted) is associated with a 0.2 per cent higher wage. This estimate is significant for uncompleted training. The fixed effect estimates do not show a significant impact on wages. One drawback of these early NLSY data is that training is only reported if it lasted longer than one month. It seems likely that many training is left unreported. Of the 12,686 individuals in the NLSY only 3,064 are used in the analysis. Of these 128 report on-the-job training which is 4.2 per cent of the sample.

Veum (1995) uses the NLSY for the years 1986 to 1990. After 1986 the training questions changed and also covered training lasting less than a month. Respondents could report information for up to 4 training programs. Veum considers those who had completed formal schooling by the 1986 interview. About 18% report having participated in company training, while the average time spent on this training was 135 hours. This does indeed suggest that many company training programs are of short duration. Veum finds that one hour of company training increases wages by 0.7 to 0.9 per cent.

Parent (1999) uses the NLSY for the longer period of 1979 to 1991. About 16 per cent of the individuals report having participated in on-the-job training. Parent (1999) estimates both simple OLS regressions and IV regressions with Hausman-Taylor type instruments that are orthogonal to the individual fixed effects.\(^5\) The

\(^5\)Hausman and Taylor (1981) observed that for time-varying variables the deviation from their time mean can be used as instruments. They also showed that the time means of the exogenous time-invariant variables can be used as instruments for the time-invariant endogenous variables. This requires that there are at least as many exogenous time-invariant variables as endogenous time-invariant variables.
<table>
<thead>
<tr>
<th>Study</th>
<th>Data</th>
<th>Period</th>
<th>Sample</th>
<th>Training</th>
<th>Method</th>
<th>Estimate</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>United States</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lynch (1992)</td>
<td>NLSY</td>
<td>1980-1983</td>
<td>No college graduates</td>
<td>Weeks of OJT</td>
<td>OLS+SC</td>
<td>0.0020</td>
<td>-0.0002 (0.0012)</td>
</tr>
<tr>
<td>Veum (1995)</td>
<td>NLSY</td>
<td>1986-1990</td>
<td>Hours of OJT</td>
<td>OLS</td>
<td>FE</td>
<td>0.0073</td>
<td>0.0090 (0.0392)</td>
</tr>
<tr>
<td>Parent (1999)</td>
<td>NLSY</td>
<td>1979-1991</td>
<td>Years of OJT</td>
<td>OLS</td>
<td>HT</td>
<td>0.1692</td>
<td>0.1216 (0.0372)</td>
</tr>
<tr>
<td>Loewenstein and Spletzer (1998)</td>
<td>NLSY</td>
<td>1988-1991</td>
<td>Formal Company Training</td>
<td>FE</td>
<td></td>
<td>0.0346</td>
<td>0.0193 (0.0193)</td>
</tr>
<tr>
<td><em>Canada</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent (2003)</td>
<td>FSLS</td>
<td>1995</td>
<td>Men, aged 22-24</td>
<td>Career of job-related training</td>
<td>FE</td>
<td>0.1034</td>
<td>0.0311 (0.0311)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Women, aged 22-24</td>
<td></td>
<td>FE</td>
<td>0.0168</td>
<td>0.0292 (0.0292)</td>
</tr>
</tbody>
</table>
OLS estimate of the return to one full-time year of training is 18 percent. This estimate drops slightly in his partial fixed effect estimation to 12 percent.

Finally, Frazis and Loewenstein (2003) estimate various specifications using the NLSY data for the years 1979 to 2000. They estimate rates of return (instead of wage returns), and find that at median training of 60 hours the rate of return is in the 150-175 percent range, while their preferred estimate that takes into account heterogeneity in wage growth (fixed effects wage growth regressions) is a rate of return in the region of 40 to 50 percent for one full-time week of training.

Parent (2003) uses data from a Follow-Up to the School Leavers Survey (FSLS). He estimates fixed effects models for men and women separately. For men participation in employer-supported training increases hourly wages by more than 10 percent. For women the effects is much more modest, about 2 percent, and not statistically significant. Parent also reports returns to weekly earnings. These returns are higher, both for men (0.1364) and for women (0.0564), suggesting that there might be employment effects.

With the exception of Lynch (1992) (and to some extent Parent, 1999), return estimates are high for the US. There are various possible explanations for Lynch’s results. First, her sample is made up of less educated individuals. Second the data are for the early 1980s, while it is widely documented that returns to skill increased substantially over the 1980s. A probably more likely explanation for her relatively low return estimates, is the fact that training spells that lasted less than a month are not reported in her data (the same holds for most of the data used in Parent, 1999). This not only suggests that there are quickly decreasing returns to training, but is also suggestive that the high returns that subsequent studies found are largely due to relatively short training spells.

5.2.2 UK There are numerous studies that estimate wage returns for the United Kingdom. An early study is Greenhalgh and Stewart (1987) that uses 1975 data from the British National Training Survey. The outcome variable they consider is occupational status. It is found that training, defined as anything that may have helped an individual to learn/do his work, has a significant effect on occupational status, but the marginal benefit is zero after four weeks. Booth (1991) uses data from the 1987 British Social Attitudes Survey (BSAS). The outcome variable that
Booth considers is annual earnings. She finds high returns; 11.2 per cent for men, and 18.1 percent for women. Potential selectivity into training is not taken into account. Booth (1993) improves on this using the 1980 British National Survey of Graduates and Diplomats (BNSG). This study reports estimates using both selectivity corrected OLS and fixed effect estimates. She finds that 1 week of training in the first year on the job increases earnings by 1 per cent both for men and women. Unfortunately her exclusion restriction in the training probit lack explanatory power so that her Heckman two-step procedure is basically identified on functional form only. Turning to her fixed effect estimates it is seen that there are no longer returns for men, while the point estimate for women remains the same.

A number of studies have used the National Child Development Survey (NCDS) to estimate the wage returns to training. These studies look at training incidence and wage growth over the period 1981 to 1991. Blundell et al. (1996) use a quasi-differencing approach that allows unobserved heterogeneity to affect wages differentially over time. In addition the remaining transitory shocks are instrumented using individuals’ first job characteristics and wages, observed ability, family background, pre-1981 training and post school qualification variables other than a degree. They find that participation in employer provided on-the-job training increases wages by 3.6 per cent for men. For women there is no significant effect on wages. Participation in off-the-job training has higher returns, about 7 percent for men and, 5 per cent for women.

Blundell et al. (1999) report OLS, fixed effect and instrumental variable estimates using the same data but for a somewhat larger sample. The instrument set used is similar to the one in Blundell et al. (1996). The analysis considers employer provided training courses. For men, OLS gives a wage return of 8.3 per cent, the fixed effect estimate is lower at 5 per cent while the IV estimate is 6.5 per cent. Returns to training courses that do lead to a qualification are of similar magnitude. For women the returns are somewhat less clear cut, but still considerable. The OLS estimate of the return to an employer provided training course that does not lead to a qualification is about 15 per cent estimated return from the fixed effect specification is about 12 percent, while the IV estimate drops to about 3 percent. Returns are higher for training courses that do lead to a qualification. Here the fixed effect estimate is over 17 per cent and the IV estimate about 8 per
<table>
<thead>
<tr>
<th>Study</th>
<th>Data</th>
<th>Period</th>
<th>Sample</th>
<th>Training</th>
<th>Method</th>
<th>Estimate</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Booth (1991)</td>
<td>BSAS</td>
<td>1987</td>
<td>Males Females</td>
<td>incidence, formal job-related</td>
<td>OLS</td>
<td>0.106</td>
<td>(0.038)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.166</td>
<td>(0.040)</td>
</tr>
<tr>
<td>Booth (1993)</td>
<td>BNSG</td>
<td>1986/87</td>
<td>Males, Graduates Females, Graduates</td>
<td>weeks in 1st year, employer provided</td>
<td>OLS+SC</td>
<td>0.010</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.010</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Males, Graduates Females, Graduates</td>
<td>weeks in 1st year, employer provided</td>
<td>FE</td>
<td>-0.002</td>
<td>(0.003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.010</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Blundell et al. (1996)</td>
<td>NCDS</td>
<td>1981-1991</td>
<td>Males, 33 year-olds</td>
<td>incidence, on-the-job empl provided</td>
<td>quasi-difference</td>
<td>0.036</td>
<td>(0.018)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>incidence, off-the-job empl provided</td>
<td></td>
<td>0.066</td>
<td>(0.017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Females, 33 year-olds</td>
<td>incidence, on-the-job empl provided</td>
<td>quasi-difference</td>
<td>0.003</td>
<td>(0.032)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>incidence, off-the-job empl provided</td>
<td></td>
<td>0.046</td>
<td>(0.032)</td>
</tr>
<tr>
<td>Blundell et al. (1999)</td>
<td>NCDS</td>
<td>1981-1991</td>
<td>Males, 33 year-olds</td>
<td>empl prov course without qual</td>
<td>OLS</td>
<td>0.083</td>
<td>(0.017)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FE</td>
<td>0.050</td>
<td>(0.019)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IV</td>
<td>0.065</td>
<td>(0.044)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Females, 33 year-olds</td>
<td>empl prov course without qual</td>
<td>OLS</td>
<td>0.142</td>
<td>(0.022)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FE</td>
<td>0.110</td>
<td>(0.026)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IV</td>
<td>0.027</td>
<td>(0.056)</td>
</tr>
<tr>
<td>Arulampalam and Booth (2001)</td>
<td>NCDS</td>
<td>1981-1991</td>
<td>33-year-olds</td>
<td>incidence, work related</td>
<td>Selection model</td>
<td>0.342</td>
<td>(0.174)</td>
</tr>
</tbody>
</table>
cent.

Finally, Arulampalam and Booth (2001) estimate a hurdle model on the NCDS data, where the number of training occurrences is instrumented with the local unemployment rate in 1981, marital status, the presence of children, early ability measures and pre-1981 training courses. It is found that participation is associated with a 41% higher wage growth between 1981 and 1991. Arulampalam and Booth (2001) only find significant returns to incidence, the number of training courses is insignificant.

5.2.3 Other countries For countries other than the US or the UK, evidence on training returns is more scant. Goux and Maurin (2000) estimate wage returns to firm provided for France. While their OLS estimate is 7.1 per cent and significant, it drops to -5.7 with a large standard error after correcting for selectivity. This finding is somewhat at odds with the results of Fougère et al. (2001). They find high returns to training participation for job-switchers of about 30 percent. For non-switchers the point estimate is still a sizeable 13 percent, but no longer statistically significant. Pischke (2001) is a careful study using the German Socio Economic Panel (GSOEP). He presents both fixed effect estimates and is the first one to estimate fixed effect wage growth regressions. He finds that one year of full-time work-related training increases wages by 2.6 to 3.8 percent. These estimates are however not significant. Kuckulenz and Zwick (2003) have access to the 1999 “Qualification and Career Survey” (BIBB/IAB) a 0.1 per cent sample of all employed Germans. They find that participation in work-related training is associated with more than 15 per cent higher wages after correcting for the endogeneity of training. Their exclusion restriction includes self perceived training needs and dummy variables indicating whether the employer went through a period of downsizing or workplace restructuring. Schöne (2002) finds that training participation is associated with 1 per cent higher wages in Norway. Gerfin (2003) finds effects twice that size for Switzerland using matching methods, where it is worthwhile to note that the average training course lasted 17 hours (median 8). Finally, Leuven and Oosterbeek (2002) find cross sectional returns of about 10 per cent to training, but this return drops to almost zero after narrowing down the comparison group of non-participants.
Table 3: Return Studies: Other countries

<table>
<thead>
<tr>
<th>Study</th>
<th>Data</th>
<th>Period</th>
<th>Sample</th>
<th>Training</th>
<th>Method</th>
<th>Estimate</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>France</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goux and Maurin (2000)</td>
<td>FQP</td>
<td>1988-93</td>
<td>participation, firm provided</td>
<td>OLS+SC</td>
<td>-0.057</td>
<td>(0.068)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OLS</td>
<td>0.071</td>
<td>(0.020)</td>
</tr>
<tr>
<td>Fougère et al. (2001)</td>
<td>FQP</td>
<td>1993</td>
<td>Job-switchers, Non-switchers</td>
<td>participation, firm provided</td>
<td>Switching Regr.</td>
<td>0.293</td>
<td>(0.121)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.128</td>
<td>(0.091)</td>
</tr>
<tr>
<td><strong>Germany</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pischke (2001)</td>
<td>GSOEP</td>
<td>1986-89</td>
<td>years, work related</td>
<td>FE, growth</td>
<td>0.038</td>
<td>(0.027)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FE</td>
<td>0.026</td>
<td>(0.019)</td>
<td></td>
</tr>
<tr>
<td>Kuckulenz and Zwick (2003)</td>
<td>BIBB/IAB</td>
<td>1998/99</td>
<td>incidence, work-related</td>
<td>Selection Model</td>
<td>0.15</td>
<td>(0.04)</td>
<td></td>
</tr>
<tr>
<td><strong>Norway</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schone (2002)</td>
<td>NSOE</td>
<td>1989/1993</td>
<td>incidence, employer provided</td>
<td>OLS</td>
<td>0.053</td>
<td>(0.007)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>FE</td>
<td>0.011</td>
<td>(0.006)</td>
<td></td>
</tr>
<tr>
<td><strong>Netherlands</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leuven and Oosterbeek (2002)</td>
<td>EPIO</td>
<td>2000</td>
<td>incidence, work related</td>
<td>OLS</td>
<td>0.098</td>
<td>(0.029)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SC</td>
<td>-0.005</td>
<td>(0.056)</td>
<td></td>
</tr>
<tr>
<td><strong>Switzerland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gerfin (2003)</td>
<td>SLFS</td>
<td>1998-2000</td>
<td>Males</td>
<td>incidence, work-related</td>
<td>Matching</td>
<td>0.02</td>
<td></td>
</tr>
</tbody>
</table>
5.3 High returns?

The results discussed above illustrate the fact that for a variety of datasets and countries the estimated returns to private-sector training are substantial. Moreover, the returns to private-sector training are very high compared to, for example, the returns to schooling. The return to a year of full-time education is around 10 percent, where in contrast the literature finds returns at least as high for a week of private-sector training. This raises the question whether these estimates are indeed causal effects.

The results can be explained by heterogeneity in returns, selectivity and measurement error. Selectivity for example may bias the estimated return to training. If the researcher has information that arguably correlates with training participation but not with wages, then instrumental variable methods potentially offer a way to obtain consistent estimates of the return to training.

To illustrate the issue, consider Leuven and Oosterbeek (2004), who exploit a provision in the Dutch tax system that allows employers to deduct an extra 40 percent of the training cost of employees that are 40 years or older from their taxable profits. The structure of the age-dependent tax deduction is therefore discontinuous at age 40. All workers younger than 40 are excluded from this additional deduction, while all workers aged 40 or older are included. This structure constitutes a perfect example of a so-called regression discontinuity (RD) data design (cf. Thistlewaite and Campbell, 1960). This implies a credible exclusion restriction. The paper then shows that the tax provision significantly shifts participation in a discontinuous way around age 40. This is the second requirement of a valid instrument. The authors then proceed by using the discontinuity as an instrument for training participation. The IV point estimates give no support for substantial returns to employer provided training. Unfortunately the estimates are too imprecise to warrant firm conclusions. Although IV seems promising, and a few studies discussed above have followed this approach, it is very difficult to come up with credible exclusion restrictions: variables that affect wages but only through training are difficult to find.

An alternative approach is followed in Leuven and Oosterbeek (2002). The idea forwarded in this paper is to narrow down the comparison group to those non-
participants who did not participate due to some random event. This is achieved by using the information obtained through two especially designed survey questions. The first is whether there was any training related to work or career that the respondent wanted to follow but did not do so. The second asks whether this non-participation was due to some random event such as family circumstances, excess demand for training places, transient illness, or sudden absence of a colleague. Respondents who give an affirmative answer to both questions are arguably a more appropriate comparison group. Under two assumptions this approach gives an estimate of the effect of treatment on the treated.

OLS gives an estimate that is similar in magnitude to those found for the studies cited above, and is 12.5 percent for participating in one training course (with median duration of 40 hours) during the past 12 months. Restricting the comparison group to workers who wanted to participate in training but did not do so, reduces the estimated return to 8.7 percent. When the comparison group is further restricted to those workers who wanted to participate in training but did not do so due to some random event, the point estimate of the return to training is 0.6 percent.

Although sample sizes do not allow precise estimation of the latter effect, the credibility of the proposed strategy is supported by the fact that on each subsequent narrowing down of the comparison group, the participants and comparison individuals are increasingly similar on observed characteristics. In line with this increased similarity of trainees and non-trainees the point estimate of the return to training consistently drops. This suggests that the high returns to private-sector training previously found in the literature could be explained by the spurious correlation of training with confounding factors that affect wages, and that existing methods are not sufficient to take this into account.

6 Discussion and conclusion

This paper has provided a tentative review of the literature that estimates wage returns to training. The measurement and estimation issues have been extensively discussed. As argued above, the fundamental problem concerning the recovery of the causal effect of training on earnings lies in the correction for selectivity.
into training. The emphasis has been methodological on the one hand, while emphasising the size of the estimated returns. Something that has been largely neglected in the literature.

Section 5 shows that traditional studies that depend on differencing (fixed effect) methods where non-participants are used as a comparison group results in high return estimates. On the basis of these high returns some have argued that there is substantial underinvestment and therefore scope for public intervention. Such underinvestment could arise because of for example hold-up or liquidity constraints.

Studies that exploit arguably exogenous variation in training participation find much smaller wage effects of training. This calls into question the case for underinvestment based on return studies. There is other evidence which suggests that underinvestment may be less severe than previously believed. First, recent literature in experimental economics shows that individuals are often motivated by reciprocity and fairness considerations which are typically ignored in standard human capital models. In a recent study Leuven et al. (2005) show how reciprocity can alleviate underinvestment, and they also provide supporting empirical evidence for this mechanism. Second, recent literature emphasising market imperfections also shows that imperfections may give employers more incentives than previously thought to invest in the general training of their employees.

To conclude there seems to be a case for further study of training mechanisms and subsequent outcomes. These studies will need to address recent developments in the theoretical literature emphasizing market imperfections and behavioral considerations such as reciprocity. The literature studying outcomes should incorporate recent developments in the econometrical literature and address selectivity issues in a systematic way. First, more effort should be made to find identifying exogenous variation in training participation. Second, the literature has ignored the presence of heterogenous returns, something which seems particularly relevant in the case of on-the-job training which is heterogenous by nature. A first step would be to further refine traditional fixed effects models by combining them with matching methods that account for heterogenous returns.
References


