An alternative approach to estimate the wage returns to private-sector training

Leuven, E.; Oosterbeek, H.

DOI
10.1002/jae.1005

Publication date
2008

Published in
Journal of Applied Econometrics

An alternative approach to estimate the wage returns to private-sector training[1]

Edwin Leuven Hessel Oosterbeek

[1]We gratefully acknowledge valuable comments from Jaap Abbring, Barbara Sianesi, Steve Pischke, Bas van der Klaauw, and seminar participants in Amsterdam, Copenhagen and Lisboa. The authors are affiliated with the Department of Economics - University of Amsterdam, NWO Priority Program 'Scholar', and with the Tinbergen Institute.

Abstract

This paper follows an alternative approach to identify the wage effects of private-sector training. The idea is to narrow down the comparison group by only taking into consideration the workers who wanted to participate in training but did not do so because of some random event. Narrowing down the participant and comparison group makes them increasingly similar on observed characteristics, supporting the validity of the approach. At the same time the point estimate of the return to training consistently drops from a large and significant return to a point estimate close to zero. This suggests that a large share of what is usually interpreted as returns to private-sector training is actually the return to some unobservable characteristic.
JEL Codes: C21, J24, J31

1 Introduction

The empirical literature on private-sector training addresses two main questions: who gets training, and what is it worth? While the first question has been answered fairly satisfactory, this is not the case for the second question. Even when attention is restricted to the effect of training on wages rather than on productivity, severe problems are posed by the endogeneity of training decisions.

Workers who participate in training, or who participate more often or for longer durations, are unlikely to have the same characteristics as other workers. Like with regular education it seems likely that the selection of workers into training is correlated with workers' unobserved ability. If this is the case, and since ability positively affects wages, a regression of wages on training will not produce a causal effect but suffer from so-called ability bias. A key issue in the estimation of returns to training is thus how to correct for this potential source of bias.

The empirical training literature contains basically two approaches. The first approach is to augment the wage equation with a Heckman-type selection correction term which results from a first stage training participation equation. Results from this approach are reported by Lynch (1992) and Veum (1995) among others. The difficulty with this approach is twofold. First the parametric selection models estimated in the literature are restrictive in the sense that they make assumption about the distribution of the unobservables. Second, it is very hard to find variables which affect training participation and have arguably no direct effect on wages. The problem of finding such credible exclusion restrictions also also hampers the application of an instrumental variable (IV) approach. A second more often used approach is to estimate the wage return to work-related training using fixed effects regressions. This estimator, which is similar in spirit to taking first differences of the before and after training log wage equations, purges permanent individual effects from the estimating equation. Examples of studies that follow this approach include Barron et al. (1993); Booth (1993); Frazis and Loewenstein (2003); Greenhalgh and Stewart (1987); Lynch (1992); Parent (1999); Veum (1995).

The fixed effects estimator produces unbiased estimates only when the unobserved individual effects are permanent. It is conceivable however that, apart from selection based on fixed individual observables and unobservables, selection into training also has dynamic aspects that provide an additional potential source of bias. Consider for example the case where individuals decide to take training because their earnings are temporarily low, faster earnings growth is then expected to occur among the trainees even in the absence of training participation. More in general, fixed effects estimations do not recover causal relationships if wage growth is different for trainees and non-trainees.

Estimates of the wage returns to private-sector training are typically very high. As an illustration take the estimates from Frazis and Loewenstein (2003) who use the NLSY dataset and present a very careful and thorough analysis of these data. They estimate various specifications and their preferred estimate that takes into account heterogeneity in wage growth is a rate of return in the region of 40 to 50 percent for one full-time week of training. These findings are consistent with those of Barron et al. (1993); Loewenstein and Spletzer (1999).[1]

Estimated returns are also high with data from other sources and countries. Bartel (1995) using company data, for example, finds that one day of training increases wages by 2 percent which in her data is equivalent to a rate of return of 60 percent. Blundell et al. (1996) report returns to training incidence (zero-one dummy variable) for men in the UK in the region of 8 percent using OLS estimations, 9 percent for fixed effects estimations, and 7 percent for IV estimations and the returns for women are even higher.

These results 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 education. 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.[2]

In this paper we follow a slightly different approach to estimate the returns to private-sector training. We will use OLS as a benchmark result that does not correct for selectivity. We will then compare these results with estimates based on an approach that takes the concept of random assignment literally. The idea 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 specially 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

[1] Lynch (1992) and, to some extent Parent (1999), find lower returns using the NLSY data. This is most likely due to to the fact that they use pre 1988 data, when training spells lasting less than one month were not reported. It also shows that especially short training spells correlate with high wage growth.

[2] There are of course expections. Some studies find smaller returns, typically for continental European countries, (e.g. Pischke 2001; Goux and Maurin 2000), although other studies find larger returns for the same countries (e.g. Fougére et al. 2001; Kuckulenz and Zwick 2003).

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 us an estimate that is similar of 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. A second important thing to note is that our point estimate is well below the lower bounds typically estimated in literature cited above. The results of the analysis therefore suggest that the high returns to private-sector training previously found in the literature are most likely explained by the spurious correlation of training with confounding factors that affect wages and that fixed effect estimations are not sufficient to take this into account.

The next section discusses at more length the questions that were used to create the new comparison group and the identifying assumptions underlying this approach. Section 3 presents the data and compares participant and comparison groups in terms of observed characteristics. Section 4 presents the empirical results. Section 5 concludes.

2 Method

2.1 Construction of participant and comparison groups

Studies that estimate causal wage returns to private sector training compare wages of training participants with the wages of an appropriate comparison group. The training measure used in this paper to define the group of participants is a conventional one.[3] The exact phrasing of the question that is used to determine this reads:

> ”Did you spend time following a course/training for purposes of your work or career opportunities during the past 12 months?”

All respondents who followed private sector training during the 12 months prior to the interview are assigned to what we will refer to as Participant group I. Without any correction for selectivity the comparison group consists of all respondents who did not follow a course or training during the 12 months prior to the interview. We refer to this group as Comparison group I.

If training participation is randomly assigned to workers, the difference between the average wage in Participant group I and Comparison group I gives the causal effect of training on wages. It is unlikely, however, that training is assigned on a random basis. Selection into training requires that: (1) the worker is willing to undertake training, and (2) the employer is prepared to provide it. The factors underlying these selection mechanisms are likely to be related, directly or indirectly, to future outcomes and may lead to differences between the training participants and potential comparison groups in terms of characteristics that are not observed by the analyst. As a result comparing wages of Participant group I and Comparison group I will give a biased estimate of the causal effect of training on wages.

The identification strategy that is proposed in this paper reduces the comparison group to those workers who are willing to undertake training and whose employers are prepared to provide it, but who did not follow the training they wanted to follow due to some random event. Narrowing down the comparison group proceeds in two steps. The first step reduces the comparison group to the group of untrained workers who wanted to follow training but did not do so. This is done on the basis of information from a question which asks respondents the following question.

> ”Was there any course/training related to work or career you wanted to follow but did not during the past 12 months?”

Persons who respond that there was such a course and who did not follow any training at all during the past 12 months, are assigned to Comparison group II. Notice that Comparison group II is a subsample of Comparison group I. Comparison

[3]To illustrate, the training question in the NLSY reads “Since [the date of the last interview], did you attend any training program or any on-the-job training designed to help people find a job, improve skills, or learn a new job?”

group II is arguably a more suitable comparison group than Comparison group I as it singles out all workers who were motivated to participate in training. Hence, it takes care of the first of the two selection mechanisms.

The second step is to further reduce the comparison group to untrained workers whose non-participation is due to some random event. Respondents who indicated their intention to be trained (i.e. wanted to follow some course/training but did not do so), were subsequentially asked the reason for not following the course/training. To answer this question respondents had to choose one out of five alternatives:

1. A random event (N=77)
2. Lack of time (N=93)
3. Own financial contribution too high (N=13)
4. Lack of support from the employer (N=21)
5. Other reasons (N=45)

Among all 249 respondents who indicated that there was a training course they would have wanted to follow there are in total 77 respondents who say that they did not do so due to some random event. The respondents were given the following examples of such events: family circumstances, transient illness, or sudden absence of a colleague. These persons constitute the final comparison group referred to as Comparison group III.

Comparison group III consists of respondents who did not follow any training course at all during the 12 months prior to the interview due to some random event. Participant group I, however, consists of respondents who received at least one course. Comparison group III therefore seems a more appropriate comparison group for the group that received exactly one training/course than for the group who received two or more courses. For this reason Participant group II is constructed which consists of respondents who followed exactly one training/course. Table 1 summarizes the definition of the participant and comparison groups.

Given that the assignment of respondents to Comparison group III is crucial for the approach of this paper, some further discussion is warranted. A first thing to note is that respondents are not assigned to the final comparison group when they mention one of the other categories as reason for not having followed training. These other categories include the more obvious ones such as lack of time, too expensive and lack of employer support. They also include the category ”other

Table 1: Definition of the participant and comparison groups

	Definition
Participant I	At least one training course
Participant II	Exactly one training course
Comparison I	No training
Comparison II	No training, but wanted to follow training course
Comparison III	No training, but wanted to follow training course and did not do so because of a random event

reasons”. This is an open category which interviewees had to respond to when they mentioned ”other reason”. In the category ”other reason” the following reasons for not participating were mentioned: language problems, merger of current employer, no available transportation, change of job, moving house, stay abroad, pregnancy. The respondents that mentioned these reasons did therefore not considered these events as random. Secondly, in the absence of the random event which withheld them from training, the respondents in Comparison group III would have participated in training. As such Comparison group III serves to identify the effect of the treatment on the treated.

Of the three training measures; participation, number of courses and number of hours, the last one is the probably the most accurate measure of the investment in human capital. The analysis is nevertheless based on the other two measures. With Participant group I training is measured as mere participation, whereas with Participant group II training is measured as one course versus no training at all. The number of hours of training could not be used as the unit of measurement of training investment since we did not ask respondents whether they missed an additional hour of training due to some random event.

2.2 Identifying assumptions

This subsection highlights the identifying assumptions implicit in our empirical strategy outlined in the previous subsection by writing down the well known potential outcome model. It not only makes precise the identifying assumption, but also shows how the analysis relates to standard (local) IV methods. This is particularly useful when we discuss how the estimates that we obtain compare to local IV estimates or average treatment effects if the identifying assumptions are violated.

Let Y_{1i} and Y_{0i} represent potential outcomes (wages) for individual i with and without training participation (D_i). The observed outcome Y_i is related to potential outcomes and training participation in the following way:

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i$$

Let instrument Z_i take value one if a random event occurred that may withhold individual i from following training, while it takes value zero if such an event did not occur. Training participation depends on Z_i in the following way.

$$D_i = D_{0i} + (D_{1i} - D_{0i})Z_i$$

where D_{0i} indicates whether individual i would participate in training when the random event does not occur. D_{1i} indicates whether individual i would participate in training when the random event does occur.

Following Imbens and Angrist (1994) the next condition defines an instrument

CONDITION 1 (existence of instrument): Let Z_i be a binary random variable such that (i) Z_i is jointly independent of $\{Y_{0i}, Y_{1i}, D_{0i}, D_{1i}\}$, and (ii) $\Pr(D_i = 1|Z_i)$ is a nontrivial function of Z_i.

Angrist and Imbens show that under the monotonicity condition $D_{0i} \geq D_{1i}$ what they coin a local average treatment effect (LATE) can be identified:

$$\begin{aligned} LATE &= E[Y_{1i} - Y_{0i}|D_{0i} - D_{1i} = 1] \\ &= \frac{E[Y_i|Z_i = 0] - E[Y_i|Z_i = 1]}{\Pr(D_i = 1|Z_i = 0) - \Pr(D_i = 1|Z_i = 1)} \end{aligned} \tag{1}$$

They also note (p. 469, see also Heckman 1990) that access to an instrument which assures that the probability of participation equals zero allows the identification of the average treatment effect of the treated. If individuals never participate if the random event $Z_i = 1$ occurs, we have that

CONDITION 2 (homogeneity):

$$\Pr(D_i = 1|Z_i = 1) = 0 \tag{2}$$

Under this condition the LATE is equivalent to the average treatment effect on the treated (ATT)

$$ATT = E[Y_{1i} - Y_{0i}|D_i = 1] = \frac{E[Y_i|Z_i = 0] - E[Y_i|Z_i = 1]}{\Pr(D_i = 1|Z_i = 0)} \tag{3}$$

The analysis in this paper differs however from the standard IV or LATE approach because we do not observe Z_i for every individual. As a consequence we cannot implement the estimator (3) in the analysis. Instead we pursue the following approach. The parameter of interest is the effect of training on the participants:

$$E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1] \tag{4}$$

The first term on the right-hand side equals the observed average wage for participants which is readily observed since $E[Y_{1i}|D_i = 1] = E[Y_i|D_i = 1]$. The second term on the right-hand side of (4) remains to be identified.

Condition 2 implies that an individual will not participate in training if the random event happens. This condition also implies that if an individual participated in training he did not experience a random event: $\Pr(Z_i = 1|D_i = 1) = 0$. As a consequence the second term in the right hand side of (4) reduces to the following expression:

$$\begin{aligned} E[Y_{0i}|D_i = 1] &= E[Y_{0i}|D_{0i} = 1,\ Z_i = 0] \cdot \Pr(Z_i = 0|D_i = 1) \\ &= E[Y_{0i}|D_{0i} = 1,\ D_{1i} = 0,\ Z_i = 0] \end{aligned}$$

For Comparison group III, those without training due to a random event, we not only know that they did not participate in training ($D_i = 0$) and experienced a random shock ($Z_i = 1$), but since they report that they would have participated in training without the random shock we also know that for them $D_{0i} = 1$. We therefore observe the following expression.

$$E[Y_i|i \in \{\text{Comparison group III}\}] = E[Y_{0i}|D_{0i} = 1,\ D_{1i} = 0,\ Z_i = 1]$$

We are now in the position to identify the second term in the right-hand side of equation (4) as follows.

$$E[Y_{0i}|D_i = 1] = E[Y_i|i \in \{\text{Comparison group III}\}]$$

if the following condition holds

$$E[Y_{0i}|D_{0i}=1, D_{1i}=0, Z_i=0] = E[Y_{0i}|D_{0i}=1, D_{1i}=0, Z_i=1] \qquad (5)$$

which is implied by the independence assumption in condition 1.

Note that condition 2 is necessary since Z_i is only partially observed. It is a stronger condition than both the monotonicity condition that identifies a local average treatment effect and the usual IV condition that the instrument affects participation. It allows us, however, to identify something more interesting, namely the effect of the treatment on the treated instead of a local average treatment effect.

To investigate how weakening the identifying assumptions will change the interpretation of our estimator, first consider what happens if we replace the homogeneity condition with the local average treatment effect monotonicity assumption $D_{0i} \geq D_{1i}$ (see f.e. Imbens and Angrist 1994; Angrist et al. 1996). In this case our estimator will provide an upper bound to the local average treatment effect if $E[Y_{1i}|D_{0i}=1, D_{1i}=1] \geq E[Y_{1i}|D_{0i}=1, D_{1i}=0]$, that is: the expected outcome for "always-takers" is not smaller than the expected outcome for "compliers". This is a condition that is likely to hold if individuals self-select into training on the basis of returns, since one would expect that returns are higher for always-takers than for compliers.

This result can be seen as follows. First note that the LATE is defined by the following expression

$$\begin{aligned} LATE = E[Y_{1i} - Y_{0i}|D_{0i} - D_{1i} = 1] &= E[Y_{1i}|D_{0i}=1, D_{1i}=0] \\ &\quad -E[Y_{0i}|D_{0i}=1, D_{1i}=0] \\ &= E[Y_i|D_{0i}=1, D_{1i}=0, Z_i=0] \\ &\quad -E[Y_i|D_{0i}=1, D_{1i}=0, Z_i=1] \end{aligned}$$

where the second step follows by condition 1. The estimator implemented in the analysis is (the sample analog of)

$$ATT = E[Y_i|D_i=1] - E[Y_i|D_{0i}=1, D_{1i}=0, Z_i=1] \qquad (6)$$

which recovers an upper bound to LATE if

$$E[Y_i|D_i=1] \geq E[Y_i|D_{0i}=1, D_{1i}=0, Z_i=0]$$

which happens if

$$E[Y_{1i}|D_{0i}=1,D_{1i}=1] \geq E[Y_{1i}|D_{0i}=1,D_{1i}=0]$$

since the average outcome of the treated is a weighted average of “always takers” and the “compliers” that participated in training.

Now consider condition 1. The fact that the survey uses the phrase ”random event” is very suggestive that there is indeed no relation with the (non-systematic) component of the potential outcomes and that condition 1 therefore holds. It should be noted, however, that this cannot be completely ruled out. If the occurrence of such a random event correlates with non-observed characteristics that influence wages then condition 1 will be violated. An example in place are respondents with children. Children are sometimes sick and their sickness might prevent their parents from participating in training. If the number of children is also related to wages then condition 1 will no longer hold if we fail to condition on presence of children. However, if condition 1 is violated then Z_i is likely to correlate negatively with Y_{0i} since the random events (such as sickness, family circumstances, etc.) are more likely to be detrimental than beneficial to productivity. This implies that $E[Y_{0i}|Z_i=1] \leq E[Y_{0i}|Z_i=0]$ and our estimate will therefore be an upper bound of the ATT. We want to stress however that the questionnaire emphasized the transient and sudden nature that these events should have.

The approach followed here differs from the use of “no-shows” as a comparison group as has been done for the evaluation of active labor market programs (Bell et al., 1995; see also Heckman et al., 1999, p. 1940). No-shows are applicants to the program who have been accepted but nevertheless fail to participate in the program. Because the reasons for this non-participation are unknown, it may be related to systematic but unobserved characteristics which may thus lead to biased estimates. Translated to our application, workers belonging to Comparison group II with the exception of those who mention lack of support from their employer would constitute the group of no-shows. Going from Comparison group II to our preferred Comparison group III attempts to delete those cases from the comparison group for whom non-participation is likely to be related to non-random factors.

3 Data

The data were collected in January and February 2001. Interviews were held by telephone using computer-aided techniques. The data are a representative sample

Table 2: Sample means per participant and comparison group

	Participant		Comparison		
	I	II	I	II	III
	(1)	(2)	(3)	(4)	(5)
Female	0.48	0.49	0.54	0.52	0.58
Age	38.25	37.70	39.80	38.17	37.70
Children	0.97	0.99	1.05	1.18	1.16
Non-Dutch	0.06	0.06	0.05	0.06	0.03
Single	0.15	0.15	0.13	0.13	0.17
Temporary job	0.13	0.15	0.15	0.15	0.17
Firm tenure (months)	112	109	101	95	97
Education					
- Low	0.11	0.13	0.20	0.15	0.12
- Intermediate	0.48	0.49	0.53	0.55	0.56
- High	0.41	0.38	0.27	0.30	0.32
Firm size					
- up to 50	0.31	0.35	0.41	0.41	0.39
- 50 to 200	0.25	0.23	0.22	0.24	0.22
- more than 200	0.43	0.42	0.37	0.35	0.39
N	1021	582	1145	249	77

of the Dutch population aged 16-64. The employed persons were asked questions concerning their employment characteristics, and wages. They also responded to an extensive set of questions about the training activities they undertook in the 12 months prior to the interview.

Table 2 presents sample means for the two participant groups and three comparison groups. These means relate to gender, age, education, firm size, number of children, being non-Dutch, being single, temporary job status and firm tenure. The first four variables are often included in wage equation as controls. The empirical analysis in the next section presents results from wage equations with and without controls for these variables.

The means reported in Table 2 already hint at the fact that Comparison group III is more comparable to Participants groups I and II than Comparison groups I

Table 3: Tests of equality between participant (P) and comparison (C) groups, p-values

	PI vs. CI (1)	PI vs. CII (2)	PI vs. CIII (3)	PII vs. CI (4)	PII vs. CII (5)	PII vs. CIII (6)
Female	0.004	0.212	0.072	0.036	0.344	0.106
Age	0.000	0.908	0.630	0.000	0.520	0.999
Children	0.099	0.009	0.175	0.318	0.035	0.264
Non-Dutch	0.713	0.698	0.240	0.714	0.746	0.239
Single	0.345	0.588	0.585	0.418	0.609	0.598
Temporary job	0.345	0.465	0.361	0.809	0.978	0.676
Firm tenure	0.324	0.363	0.977	0.193	0.246	0.870
Education	0.000	0.002	0.230	0.000	0.054	0.567
Firm size	0.000	0.003	0.253	0.006	0.035	0.481

Note: The p-values are based on t-tests for the continuous variables age, number of children, firm tenure and log wage and on rank-sum tests for the categorical variables female, education, firm-size, non-Dutch, single and temporary job.

and II are. By and large the means of Comparison group III are closer to those of Participants groups I and II than the means Comparison groups I and II. This is most notably the case for the variables age, education and firm size. With the exception of female and children, the means of the other variables in Table 2 are not very different across all five groups. Formal test statistics about this are reported in Table 3.

Table 3 reports test statistics for significant differences between the participant groups and comparison groups. The first of these columns shows that Participant group I and Comparison group I are significantly different with respect to each of the variables gender, age, number of children, education and firm size. Replacing Comparison group I by Comparison group II removes the significant differences with regard to gender and age, but the differences for number of children, education and firm size remain significant. When we compare Participant group I with Comparison group III, there appear to be no significant differences with respect to age, education, firm size and number of children. Only for gender do we observe a significant difference at the 10 percent level. This finding is in line with the aforementioned potential problem; a random event refraining someone from attending a training course is more likely to occur for women than for men. As examples of such random event the questionnaire refers to family circumstances and illness. While both events are arguably random, it is not surprising that they affect women slightly more than men: women are more often ill than men, and (at least in the Netherlands) there is still a tendency for women to bear a larger share of family responsibilities than men do.

The last three columns repeat the same exercise but now with Participant group I (all trained workers) replaced by Participant group II (workers who attended exactly one training/course). The results are very similar to those in the previous three columns. The most important difference is that now Participant group II and Comparison group III are no longer different with respect to their gender composition. This indicates that men and women have the same probability that a random event allocates them to Comparison group III rather than to Participant group II. In part the results on the tests are driven by sample size, but it is important to note that participants and comparisons do actually become increasingly similar. This is especially the case for the important dimension on which they differ most, namely education.

The questionnaire also asked the respondents who followed a training course about the characteristics of this course. For a number of these training characteristics these questions were also asked to the respondents who wanted to follow training but did not do so. For instance, respondents who attended a course were asked who provided the course, while respondents who wanted to follow training but did not do so were asked who would have provided training. Such questions were asked with respect to the type of training, the provider of training, who paid the direct costs of training and whether training (would have) resulted in a certificate. Table 4 presents the descriptive statistics of these characteristics for Participant group I, Comparison group II and Comparison group III. For respondents in Participant group I who followed more than one course, the answers relate to the first course they mention.

There are significant differences between the characteristics of the first training attended by respondents in Participant group I and the characteristics of the training which respondents in Comparison group II wanted to follow. These differences are in terms of the training provider, the party that pays the direct costs, and whether the course leads to a certificate. Such differences are not present when Participant group I and Comparison group III are compared. Again sample sizes partially drive the results of the tests, yet it is remarkable how similar comparisons and participants become on finance. This is important given that the focus of our analysis is on wage returns which in turn depends on cost sharing.

The results in this section suggest that, with respect to respondents' observable characteristics, Comparison group III is more comparable to the two participant groups than Comparison groups I and II. Moreover, the courses actually followed by Participant group I and the courses which respondents in Comparison group III wanted to follow are not significantly different with regard to observable characteristics. This is no longer true when Participant group I is compared with Comparison group II. This does not prove that Comparison group III is identical to a real randomly generated comparison group, but it is an indication that Comparison group III is more appropriate than Comparison groups I and II.

4 Estimation results

Table 5 shows the coefficients of training in log wages equations for different combinations of participant and comparison groups and for different sets of control variables. As sets of control variables we distinguish between: none, a female dummy only since the results in Table 2 point to some difference between Participant group I and our preferred Comparison group III with respect to this variable,

Table 4: Characteristics of training per participant and comparison group

	Participant	Comparison	
	I	II	III
	(1)	(2)	(3)
Type			
- Foreign language	0.03	0.07	0.09
- Safety	0.12	0.05	0.08
- IT	0.19	0.21	0.19
- Management	0.13	0.12	0.12
- Communication	0.06	0.03	0.03
- Marketing	0.02	0.02	0.01
- Finance and administrative	0.05	0.06	0.03
- Other occupation-related	0.27	0.25	0.27
- Other	0.14	0.19	0.18
Provider			
- Commercial organization	0.36	0.40	0.35
- Employer	0.17	0.21	0.32
- Sector/branch	0.08	0.08	0.08
- Higher education Institute	0.09	0.04	0.01
- Vocational school	0.05	0.01	0.01
- Supplier	0.00	0.12	0.12
- Other	0.24	0.12	0.11
Finance			
- Employer	0.81	0.67	0.84
- Employee	0.11	0.26	0.14
- Both	0.03	0.07	0.01
- Other	0.04	0.00	0.00
Certificate	0.72	0.78	0.79
Test of equality (Pr>|t|)		PI vs. CII	PI vs. CIII
Type		0.3667	0.9301
Provider		0.0138	0.1209
Finance		<.0001	0.3732
Certificate		0.0653	0.2128

and a full set of control variables with a female dummy, age and age squared, education dummies and firm size dummies.

Without controls we find a log wage difference of 0.172 between Participant group I and Comparison group I. Adding a control for female or a full set of controls reduces this difference to 0.159 and 0.107 respectively. Repeating this for Participant group II instead of Participant group I produces somewhat lower point estimates. But in all cases the wage differential between trained and untrained workers remains very substantial and is highly significant. This is in accordance with OLS estimates of the effects of training incidence reported in other studies (cf. section 1).

When we replace Comparison group I by Comparison group II the point estimates become somewhat smaller, but for both participant groups and all three specifications of the wage equation, the training premium is very substantial. As the number of observations in the comparison group reduces greatly (from 1145 to 249), the estimate is less precise but in all cases the coefficient differs significantly from zero.

This picture changes dramatically when Comparison group III serves as the comparison group; see column (8). In all cases the point estimate is reduced by about a factor five or more, and in none of the cases do we find a training premium significantly different from zero. One might argue that our research design does not permit identification of these effects because of the limited sample size in Comparison group III. It should be noted however, that increasing sample sizes to conventional numbers (as for example with Comparison group I) would still not give us enough precision to identify effects of this size since the standard errors would only go down from 0.05 to 0.03. If the intervention is modest its effects are difficult to identify.[4] As already pointed out in the introduction, it is important to note that the point estimate that we obtain is well below the typical lower bound estimated in the literature.

5 Conclusion

Estimating returns to private sector training has turned out to be a very challenging research program. It is very difficult to come up with instruments, and no one

[4]Simulations show that when randomly drawing 77 observations from Comparison group I in over 90 percent of the cases the estimated effect is larger than the point estimates reported in column (8) of Table 5. In less than 11 percent of the cases the simulations return a *p*-value exceeding the *p*-value reported for Comparison group III.

Table 5: Effect of training on wages for different combinations of participant and comparison groups and control variables

	Control	Comparison group I			Comparison group II			Comparison group III		
	variables (1)	coef. (2)	s.e. (3)	p-value (4)	coef. (5)	s.e. (6)	p-value (7)	coef. (8)	s.e. (9)	p-value (10)
Participant I	None	0.172	(0.030)	[0.000]	0.134	(0.042)	[0.001]	0.053	(0.049)	[0.281]
	Female	0.159	(0.030)	[0.000]	0.124	(0.042)	[0.003]	0.026	(0.050)	[0.603]
	All	0.107	(0.025)	[0.000]	0.089	(0.034)	[0.010]	0.014	(0.056)	[0.808]
N		2166			1270			1098		
Participant II	None	0.125	(0.035)	[0.000]	0.087	(0.046)	[0.058]	0.006	(0.052)	[0.913]
	Female	0.114	(0.035)	[0.001]	0.079	(0.046)	[0.084]	-0.019	(0.053)	[0.721]
	All	0.098	(0.029)	[0.001]	0.074	(0.037)	[0.043]	-0.005	(0.056)	[0.929]
N		1727			831			659		

Note: controls are dummy for female, age, age squared, dummies for level of education and firm size. Standard errors in round brackets; p-values in square brackets. Estimations use sample weights.

has found a natural experiment. Consequently the literature has been relying on fixed effects methods for the last two decades and estimated returns tend to be high, often several orders of magnitude higher than returns to schooling. The main contribution of this paper lies in proposing an alternative approach to estimate the wage returns to private sector training. The idea is to restrict the group of untreated individuals to those who were willing to receive training but who did not do so due to some random event. Restricting the comparison group to those who were willing to participate eliminates biases due to self-selection of workers. Restricting the group of non-participating "applicants" to those who did not participate due to some random event subsequently eliminates biases due to the selection process of firms.

The appropriateness of this newly created comparison group is corroborated by the similarity of this comparison group and the participant group in terms of a number of observed characteristics. Moreover, the courses that members of the comparison group wanted to follow and the characteristics of the courses actually followed by members of the participant group are not different in terms observed training characteristics.

Applying this approach leads to a reduction of the wage return to training from 7 to 17 percent (depending on covariates included and the exact participant group) to 0.6 percent for one training course with a median duration of 50 hours.

Without data from a real field experiment with a randomly assigned comparison group, it is not possible to prove the usefulness of our approach. Just as with instrumental variable estimates, it ultimately depends on the plausibility of the identifying assumptions whether an estimate is convincing or not. The key assumptions here are: (1) that a random event blocks training participation, and (2) the usual exclusion restriction that the non-systematic component of wages are independent of the random event. As argued above, loosening assumption 1 and violation of assumption 2 both lead to overestimation of the wage effects of training. This implies that if these identifying assumptions are not fulfilled, our estimate of 0.6 percent is an upper bound of the true wage return to private sector training, which is well below the typical lower bound estimated in the literature.

Recent theoretical contributions (Acemoglu and Pischke, 1999; Stevens, 1994) have emphasized the importance of imperfections in training markets. This literature shows that market imperfections give employers incentives to contribute more to the general training of their employees. The driving force behind this result is that market imperfections allow firms to capture part of the returns of training that is (at least partly) general. This result has implications for the literature that studies wage returns, namely that these will tend to be smaller (ceteris paribus) when market imperfections are more important. Market imperfections can not only explain differences in training practices between countries but also (i) differences in the level of wage returns between countries, and (ii) wage returns that are small. Exploration of these links are an important area for future research that studies wage returns.

References

Acemoglu, D. and Pischke, J.-S. (1999). The structure of wages and investment in general training. *Journal of Political Economy*, 107(3):539–572.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

Barron, J. M., Black, D. A., and Loewenstein, M. A. (1993). Gender differences in training, capital, and wages. *Journal of Human Resources*, 28(2):343–364.

Bartel, A. P. (1995). Training, wage growth, and job performance: Evidence from a company database. *Journal of Labor Economics*, 13:401–425.

Bell, S., Orr, L., Blomquist, J., and G.G.Cain (1995). *Program Applicants as a Comparison Group in Evaluating Training Programs*. Kalamazoo MI: W.E. Upjohn Institute for Employment Research.

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

Booth, A. (1993). Private sector training and graduate earnings. *Review of Economics and Statistics*, 75(1):164–170.

Fougére, D., Goux, D., and Maurin, E. (2001). Formation continue et carrires salariales. une valuation sur donnes individuelles. *Annales d'conomie et de Statistique*, 62:49–69.

Frazis, H. and Loewenstein, M. (2003). Reexamining the returns to training: Functional form, magnitude, and interpretation. Working Paper 367, Bureau of Labor Statistics.

Goux, D. and Maurin, E. (2000). Returns to firm-provided training: Evidence from French worker-firm matched data. *Labour Economics*, 7:1–19.

Greenhalgh, C. and Stewart, M. (1987). The effects and determinants of training. *Oxford Bulletin of Economics and Statistics*, 49:171–189.

Heckman, J., LaLonde, R., and Smith, J. (1999). The economics and econometrics of active labor market programs. In Ashenfelter, O. and D.Card, editors, *Handbook of Labor Economics*, volume 3A, chapter 31. Elsevier Science Publishers B.V., Amsterdam.

Heckman, J. J. (1990). Varieties of selection bias. *American Economic Review, Paper and Proceedings*, pages 313–318.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.

Kuckulenz, A. and Zwick, T. (2003). The impact of training on earnings - differences between participant groups and training forms. Mimeo, ZEW Mannheim.

Loewenstein, M. and Spletzer, J. (1999). General and specific training: Evidence and implications. *Journal of Human Resources*, 34(4):710–733.

Lynch, L. (1992). Private sector training and the earnings of young workers. *American Economic Review*, 82(1):299–312.

Parent, D. (1999). Wages and mobility: The impact of employer-provided training. *Journal of Labor Economics*, 17(2):298–317.

Pischke, J.-S. (2001). Continuous training in Germany. *Journal of Population Economics*, 14:523–548.

Stevens, M. (1994). A theoretical model of on-the-job training with imperfect competition. *Oxford Economic Papers*, 46:537–562.

Veum, J. R. (1995). Sources of training and their impact on wages. *Industrial and Labor Relations Review*, 48(4):812–826.