

A NEW APPROACH TO ESTIMATE THE WAGE RETURNS TO WORK-RELATED TRAINING

EDWIN LEUVEN AND HESSEL OOSTERBEEK

ABSTRACT. This paper proposes a new approach to identify the wage effects of 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. The point estimate of the return to training consistently drops when restricting the comparison group this way. While the OLS estimate of the return to training participation is significantly positive, this is no longer the case when we use the new comparison group. This outcome suggests that a large share of what is usually interpreted as returns to training is actually the return to some unobservable characteristic.

JEL Codes: C21, J24, J31

1. Introduction

A large part of the empirical training literature addresses two 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 the effect of training 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 observed and unobserved characteristics as other workers. Workers and firms are likely to (self-)select those workers into training for whom the expected returns are favorable. Workers for whom the expected benefits are low are less likely to get training.

Date: First version August 2001. This version December 2001.

We gratefully acknowledge valuable comments from Jaap Abbring, Joshua Angrist, Guido Imbens, Steve Pischke, Barbara Sianesi, Bas van der Klaauw, and seminar participants in Amsterdam.

The authors are affiliated with University of Amsterdam, Department of Economics, and NWO Priority Program 'Scholar', Roetersstraat 11, 1018 WB Amsterdam, The Netherlands. Oosterbeek is also affiliated with the Tinbergen Institute.

The key issue is how to correct for the potential endogeneity bias. The empirical training literature contains several approaches. The first approach is to augment the wage level 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 that it is hard to find variables which arguably do affect training participation and have no direct effect on wages.

The second approach uses instrumental variables which, like the Heckman-type selection models, requires the availability of a variable which arguably influences training participation but has no direct impact on the wage rate. The only study we are aware of that uses IV to estimated wage returns to training is Leuven and Oosterbeek (2001). The instrumental variable is a dummy equal to one if the worker is 40 years or older and zero otherwise. According to a recently implemented Dutch tax law firms can deduct 40 percent of training costs from their tax payments when the training involves workers of 40 years or older. This tax law gives rise to a discontinuity in training participation around the age of 40, with workers of 41 years having a 17 percent higher training rate than workers of 39 years. Both the OLS and IV estimates presented in this paper are not significantly different from zero. The main limitation of this approach is that the estimate relates to training followed by a fairly specific group of workers, namely those with an age in the vicinity of 40 years. Moreover, since the new tax law only stimulates training that is at least partially funded by the firm, the estimate may relate to training that is at least partly specific.¹

Finally, the wage return to work-related training has been estimated using a fixed effect estimator which is based on taking first differences of the before and after training log wage equations. Examples of studies that follow this approach include Barron et al. (1999), Booth (1993), Greenhalgh and Stewart (1987), Lynch (1992) and Veum (1995). This estimator produces unbiased estimates only when the unobserved individual effects

¹As Stevens (1994) and Acemoglu and Pischke (1999) show, specificity of training need not only relate to the contents of the training but may also be defined in terms of labor market imperfections.

are permanent. As Blundell et al. (1996) argue, in the case of firm training it is quite possible that participation in training is correlated with transitory shocks to productivity. To correct for such effects, Blundell et al. estimate a training participation equation to include Heckman-type selection adjustment terms in the wage growth equation. In contrast to the previous two approaches this method requires at least two data points per observation.

When the effect of training on wages is estimated with a standard wage equation which has a dummy for training participation as one of the regressors, those observations who participated in training are considered as the treatment group and all observations that did not participate in training are considered as the comparison group. The estimate of the wage effect of training thus obtained is biased when assignment to treatment and comparison groups was not random.

In this paper we follow another approach to identify the causal effect of training participation on wages. This approach takes the concept of random assignment literally and gives an estimate of the effect of treatment on the treated. 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 from 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.

The next section discusses at more length the questions that were used to create the new comparison group (or to cleanse the original 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 work-related 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. 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 work-related training during the 12 months prior to the interview are assigned to 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 econometrician. 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 due to some random event did not follow the training they wanted to follow. This proceeds in two steps. The first step reduces the comparison group to the group of untrained workers who did want to follow training. This is done on the basis

of information from a question which asks respondents whether there was training they wanted to follow, but did not follow. Literally the question reads:

”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 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 wanted to follow some course/training but did not do so, were 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, excess demand for training places, 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

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

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 one or more courses. For this reason Participant group II is constructed which consists of respondents who followed exactly one course/training. 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 reasons". This is an open category which interviewees had to respond to when they mentioned "other reason". The category "other reason" includes the following: language problems, merger, no transportation available, change of job, moving house, stay abroad, pregnancy. The respondents that mentioned these reasons have thus not considered these as random events. 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 have been used as the unit of measurement of training investment if we had known whether respondents missed an additional hour of training due to some random event.

There are 54 respondents who did follow at least one training course and also wanted to attend another course but did not do so due to some random event. This suggests an additional source of information to identify the wage effect of training. The respondents who followed exactly one course and did not attend a second course due to some random event can be compared with the respondents who followed exactly two courses to estimate the effect of a second training course. And respondents who followed exactly two course and not a third course because of some random event can be compared with the respondents who followed exactly three courses to estimate the effect of a third training course. Unfortunately the 54 respondents who followed some training but wanted more are spread out thinly over different numbers of actual training courses: 29 of them followed one course, 12 followed two courses, and the remaining 13 followed at least three courses. This makes the estimates of the associated wage effects very imprecise.

2.2 *Identifying assumptions*

In this subsection we formalize the identifying assumptions. 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. The parameter of interest is the effect of training on the participants

$$(1) \quad E[Y_{1i} - Y_{0i}|D_i = 1] = E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]$$

The first term on the right-hand side equals the observed average wage for participants:

$$E[Y_{1i}|D_i = 1] = E[Y_i|D_i = 1]$$

The second term remains to be identified. By definition

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

In our application for all i for which $D_i = 0$ it is known, by definition of the survey instrument, that $Z_i = 1$ if $D_{0i} = 1$ and $D_{1i} = 0$, and $Z_i = 0$ otherwise. Unlike in a standard IV approach, Z_i is not observed when $D_i = 1$. It is therefore necessary to assume a common effect of the instrument on participation. This allows us to infer that individuals in the participant group have not been exposed to the random event that would have distracted them from training. This is captured by the following assumption.

Assumption 1 (Homogeneity). $\Pr(D_{1i} = 1) = 0$

This assumption implies that $\Pr(Z_i = 1|D_i = 1) = 0$. It then follows that

$$E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_{0i} = 1, D_{1i} = 0, Z_i = 0]$$

We observe the following expression.

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

where the equality follows by the definition of Z_i . The second term in the right-hand side of equation (1) is therefore identified as follows:

$$E[Y_{0i}|D_i = 1] = E[Y_i|D_i = 0, Z_i = 1]$$

if the following condition holds:

$$(2) \quad E[Y_{0i}|D_{0i} = 1, D_{1i} = 0, Z_i = 0] = E[Y_{0i}|D_{0i} = 1, D_{1i} = 0, Z_i = 1]$$

Condition (2) is implied by the following exclusion restriction.

Assumption 2 (Exclusion restriction). $E[Y_{0i}|Z_i] = E[Y_{0i}]$

Assumption 1 is necessary since Z_i is not observed when $D_i = 1$. This is a stronger assumption than the monotonicity assumption made to identify a local average treatment effect and the usual IV assumption that the instrument affects participation.² It allows us, however, to identify the effect of treatment on the treated.

Note that under the local average treatment effect assumptions (see f.e. Imbens and Angrist, 1994; Angrist et al., 1996) our estimator will provide an upper bound to the local average treatment effect if $E[Y_{1i}|D_{0i} = 1, D_{1i} = 1] \geq E[Y_{1i}|D_{0i} = 1, D_{1i} = 0, Z_i = 1]$. 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.

²Note that assumption 1 implies the following assumptions necessary to identify a local average treatment effect (see f.e. Imbens and Angrist, 1994; Angrist et al., 1996). Monotonicity: $\Pr(D_{0i} \geq D_{1i}) = 1$, random assignment of the instrument: $(D_{0i}, D_{1i}) \perp Z_i$ and a non-zero average causal effect of the instrument on participation: $E[D_{1i} - D_{0i}] \neq 0$.

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 assumption 2 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 assumption 2 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 assumption 2 will no longer hold if we fail to condition on presence of children.

If assumption 2 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, again, be an upper bound. We want to stress however that the questionnaire emphasized the transient and sudden nature 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.

Table 2. Sample means per participant and comparison group

	Participant		Comparison		
	I (1)	II (2)	I (3)	II (4)	III (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
<i>Education</i>					
- 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
<i>Firm size</i>					
-	0.31	0.35	0.41	0.41	0.39
- 50-200	0.25	0.23	0.22	0.24	0.22
-	0.43	0.42	0.37	0.35	0.39
N	1021	582	1145	249	77

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 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

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.

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 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.

The questionnaire asks 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.

Table 4. Characteristics of training per participant and comparison group

	Participant	Comparison		Test of equality	
	I (1)	II (2)	III (3)	PI vs. CII (4)	PI vs. CIII (5)
<i>Type</i>				0.3667	0.9301
- 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		
<i>Provider</i>				0.0138	0.1209
- 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		
<i>Finance</i>				0.000	0.3732
- 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	0.0653	0.2128

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.

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, 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.

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

N	2166	1270	1098
Participant II	None	0.087 (0.046)	0.006 (0.052)
	Female	0.079 (0.046)	-0.019 (0.053)
	All	0.074 (0.037)	-0.005 (0.056)
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.

more, and in none of the cases do we find a training premium significantly different from zero.³

The point estimate of the wage return to training obtained when using Comparison group III may seem excessively low. Given what we (claim to) measure, however, we do not think this to be the case. When interpreted as a causal effect, the training coefficients in the bottom panel of Table 5 measure by how much the wage rate increases as a result of participating in exactly one training course during the past 12 months. The results in the first and second columns suggest that participating in one course raises wages by 7 percent or more. This return is about equal to the (presumably causal) wage return of an extra year of initial schooling. But initial schooling refers to full-time acquisition of human capital while the median duration of a training spell in our dataset equals only 50 hours. If this were true, this would point to an excessive overinvestment in initial schooling relative to training. Moreover, the low return to training reported in the final column of Table 5 is consistent with the findings we reported earlier using an entirely different - instrumental variable - approach (Leuven and Oosterbeek 2001, cf. the introduction).

5. Conclusion

This paper proposes an alternative approach to estimate the wage returns to work-related 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.

³An immediate concern is that this result is driven by the small number of only 77 observations in Comparison group III. Simulations show that when using 77 observations of 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.

The appropriateness of this newly created comparison group is revealed by the similarity of this comparison group and the participant group in terms of a large 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 zero percent. While these first returns are unbelievably high given the amount of time involved in the investment, the zero return is in line with other findings for the Netherlands.

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 approximately zero is an upper bound of the true wage return to work-related training.

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., Berger, M., and Black, D. (1999). Do workers pay for on-the-job training? *Journal of Human Resources*, 34(2):235–252.

- 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 of Work Related Training*. London: Institute of Fiscal Studies.
- Booth, A. (1993). Private sector training and graduate earnings. *Review of Economics and Statistics*, 75(1):164–170.
- 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.
- Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475.
- Leuven, E. and Oosterbeek, H. (2001). Evaluating the effects of a tax deduction on training. Mimeo, Department of Economics, University of Amsterdam.
- Lynch, L. (1992). Private sector training and the earnings of young workers. *American Economic Review*, 82:299–312.
- 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.