



## Earnings Effects of Training Programs

Michael Lechner, Blaise Melly

July 2007 Discussion Paper no. 2007-28

Editor: Prof. Jörg Baumberger  
University of St. Gallen  
Department of Economics  
Bodanstr. 1  
CH-9000 St. Gallen  
Phone +41 71 224 22 41  
Fax +41 71 224 28 85  
Email [joerg.baumberger@unisg.ch](mailto:joerg.baumberger@unisg.ch)

Publisher: Department of Economics  
University of St. Gallen  
Bodanstrasse 8  
CH-9000 St. Gallen  
Phone +41 71 224 23 25  
Fax +41 71 224 22 98

Electronic Publication: <http://www.vwa.unisg.ch>

## Earnings Effects of Training Programs

Michael Lechner and Blaise Melly

Author's address:

Professor Dr. Michael Lechner  
SIAW-HSG  
Bodanstrasse 8  
9000 St. Gallen  
Tel. +41 71 224 2350  
Fax +41 71 224 2298  
Email [michael.lechner@unisg.ch](mailto:michael.lechner@unisg.ch)  
Website [www.siaw.unisg.ch](http://www.siaw.unisg.ch)

Dr. Blaise Melly  
SIAW-HSG  
Bodanstrasse 8  
9000 St. Gallen  
Tel. +41 71 224 2301  
Fax +41 71 224 2298  
Email [blaise.melly@unisg.ch](mailto:blaise.melly@unisg.ch)

## **Abstract**

In an evaluation of a job-training program, the influence of the program on the individual earnings capacity is important, because it reflects the program effect on human capital. Estimating these effects is complicated because earnings are observed for employed individuals only, and employment is itself an outcome of the program. Point identification of these effects can only be achieved by usually implausible assumptions. Therefore, weaker and more credible assumptions are suggested that bound various average and quantile effects. For these bounds, consistent, nonparametric estimators are proposed. In a reevaluation of Germany's training programs of 1993 and 1994, we find that the programs considerably improve the long-run earnings capacity of its participants.

## **Keywords**

Bounds, treatment effects, causal effects, program evaluation

## **JEL Classification**

C21, C31, J30, J68

## 1 Introduction<sup>\*</sup>

For decades, many countries around the world have used active labor market policies to improve the labor market outcomes of the unemployed. Training programs are considered as most important components of this policy. They should increase the employability of the unemployed by adjusting their human capital to the demand in the labor market.

The evaluation of these rather costly programs has been the focus of a large substantive and methodological literature in economics (e.g., see Friedlander, Greenberg, and Robins, 1997, Heckman, LaLonde, and Smith, 1999, Kluve, 2006, and Martin and Grubb, 2001, for overviews). However, this literature could not measure the effects on human capital because it has almost exclusively studied the effects on employment and *realized* earnings or *realized* wages (setting wages or earnings to zero). Analyzing the *realized* wage or earnings distributions with and without training participation reveals only a crude measure of how much productivity the training program added. Expected *realized* earnings are the product of the individual earnings capacity or earnings potential times the probability to take up employment. Therefore, they are influenced by labor demand and labor supply and thus hard to interpret in terms of human capital and earnings capacity that are key policy parameters in relation to such programs. Since such training programs are typically targeted at populations with rather low employment probabilities, it is not surprising that most differences in realized earnings and wages uncovered by evaluation studies are driven by differences in the employment rates and not by changes in potential earnings.

---

<sup>\*</sup> The first author has further affiliations with ZEW, Mannheim, CEPR, London, IZA, Bonn, PSI, London, and IAB, Nuremberg. Financial support from the Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nuremberg, (project 6-531a) is gratefully acknowledged. The data originated from a joint effort with Stefan Bender, Annette Bergemann, Bernd Fitzenberger, Ruth Miquel, Stefan Speckesser, and Conny Wunsch to make the administrative data accessible for research. We thank Josh Angrist for helpful comments on a previous draft of the paper.

Furthermore, the earnings capacity reflects substantive and long-lasting improvements of labor market prospects. In contrast to realized earnings, they are much less dependent on fluctuations in labor demand and supply. Moreover, such gains are not subject to the so-called lock-in effect that is found in many empirical studies.<sup>1</sup> Typically, it takes some time before the employment effects stabilize to the long-term equilibrium (e.g. see Lechner, Miquel, and Wunsch, 2005). Therefore, analyzing earnings capacity instead of realized earnings has the additional advantage of allowing uncovering the long-run earnings effects before the employment effects reach their long-run level. Because programs are changed frequently in the field of active labor market policy, policy advice depends crucially on impact estimates of recent versions of the programs. For many policy questions, it is therefore more interesting to understand the differences in the distribution of earnings, under the presumption that participants and non-participants would have found a job. We analyze this effect in this paper.

Evaluating the effect of a training program on earnings capacity is, however, a complicated econometric problem because of the selective observability of earnings. Participants in training programs are typically low skilled unemployed with 'bad' employment histories and low reemployment rates. Therefore, if we are interested in the earnings effects of such programs we have to deal with the fact that many participants as well as comparable non-participants will not receive any earnings since they did not take-up employment in the first place. To complicate the issue further, we expect that those individuals who take-up jobs are not randomly selected. Instead, on average those unemployed have lower reservation wages given their productivity as observed by their employer. If minimum wage arguments are relevant, then this level of productivity per se plays an additional role.

---

<sup>1</sup> The lock-in effect (see Van Ours, 2004) describes that in the short-run all programs have negative effects as the individual job finding rates are reduced during the program.

Evaluating the earnings capacity effects of training programs is thus not straightforward. A convenient, but generally incorrect, approach is to compare the earnings for both employed groups (participants and non-participants). An alternative popular strategy is to use sample selection models (Heckman, 1979). The identification of such models either requires a distributional assumption or relies on an instrument that determines the employment status but does not affect earnings. Finding such a variable, however, is usually difficult and impossible in our application.

We follow, therefore, another strategy: We derive bounds on the average and quantile program effects on earnings capacity for specific observable populations. After having derived the so-called worst-case bounds that are usually very wide, we consider how these bounds can be tightened by making further economically motivated, but rather weak behavioural assumptions that will be plausible in many applications. It is an advantage of the bounds suggested that they do not depend on the way the selection problem relating to program participation is controlled for: by a randomized experiment, by matching, or by instrumental variables. In our particular application, we use a matching strategy that is reasonable given the informative administrative database available.

We also propose consistent, nonparametric estimators for all bounds and apply them to the evaluation of training programs in West Germany.<sup>2</sup> Active labor market policy is an important (and expensive) tool of the German labor market policy in general. Germany offers several types of training programs, which allows the differentiation of the effects according to program types. Such differentiation is very important for policy advice. Finally, our administrative data contain detailed information usually not available in other studies that allow us to

---

<sup>2</sup> We concentrate on West Germany since East Germany faces unique transition problem, which makes it hard to generalize effects found for East Germany to other OECD countries.

control for selectivity into program participation and to capture credibly important aspects of the effect heterogeneity.

This paper builds on the existing literature on partial identification. Manski (1989, 1990, 1994, and 2003) contributes very prominently to this approach consisting in bounding the effects of interest using only weak assumptions. Blundell, Ichimura, Gosling, and Meghir (2007) introduce a restriction imposing positive selection into work, while Lee (2005) uses an assumption restricting the heterogeneity of the program effects on employment. We consider variants of these assumptions and show that they allow tightening the bounds on the treatment effects. Zhang and Rubin (2003) and Zhang, Rubin, and Mealli (2007) combine these two types of assumption. Angrist, Bettinger, and Kremer (2006) use a similar combination of assumptions to bound the effects of school vouchers on test scores.<sup>3</sup>

This paper contributes to the existing literature in four ways. First, we bound the treatment effects for an observed population consisting of the employed participants. Most of the existing literature bounds the effects for the unobserved population of individuals who would work irrespective whether they participate in a program or not. However, results for an unobserved population are less intuitive and more difficult to communicate. Such a population cannot be characterized, for example, by simple descriptive statistics.

Secondly, we bound not only average but also quantile treatment effects. The effects of a treatment on the distribution of the outcome are of fundamental interest in many areas of empirical research. The policy-maker might be interested in the effects of the program on the dispersion of the outcome, or its effect on the lower tail of the outcome distribution. Interestingly, the distribution is easier to bound than the mean when we do not point identify the ef-

---

<sup>3</sup> They assume directly that the effects of the treatment on the potential test-taking status and on the potential score are positive.



fects. For instance, bounds on the support of the outcome variable are required to bound the mean but not the quantiles of a random variable in presence of missing observations.

The third contribution of this paper is to allow a more general first step selection process. All the existing papers bounding the treatment effects have assumed that the treatment status was randomly determined. While this simplifies the derivation of the results, it does not correspond to the majority of the potential applications and therefore reduces the interest in these methods. In our application, we assume that the treatment status is independent of the outcome variables only conditionally on a set of covariates. To implement our theoretical results we propose new estimators allowing for both continuous and discrete control variables.

Finally, we propose a new, policy relevant application of all our theoretical results. Using our preferred combination of assumptions, we find substantial increases in the earnings capacity for three of the four program groups we consider. In fact, both average treatment effects and most quantile treatment effects significantly exclude a zero potential earnings effect of these three programs. This shows that our bounding strategy is not only credible because it makes weak assumptions, but that this strategy can be very informative for policy makers as well.

The rest of the paper is organized as follows. The next section gives some institutional details about training programs in Germany and discusses data issues. In Section 3, we define the notation and the treatment effects of interest. We also present a unifying framework for analyzing average and quantile treatment effects. Section 4 contains the identification results. Section 5 proposes nonparametric estimators for the bounds derived in Section 4. Section 6 presents the empirical results and Section 7 concludes. The proofs of the various lemmas and theorems are relegated to an appendix that can be downloaded from the web pages of the authors at [www.siaw.unisg.ch/lechner/earnings](http://www.siaw.unisg.ch/lechner/earnings).

## 2 Training programs in Germany

### 2.1 Active and passive labor market policy

Germany belongs to the OECD countries with the highest expenditure on labor market training measured as a percentage of GDP after Denmark and the Netherlands, and it makes up the largest fraction of total expenditure on active labor market policies.<sup>4</sup> Table 1 displays the expenditures for active and passive labor market policies and especially for training programs in West Germany for the years 1991-2003. Training has the objective of updating and increasing the human capital of those workers who became unemployed. It is the most utilized instrument and represents almost 50% of the total expenditure devoted to the active labor market policy.

Table 1: Passive and active labor market policies in West Germany 1991-2003

	1991	1993	1995	1997	1999	2001	2003
Total expenditure in billion EUR	25	35	39	43	42	41	48
Shares of total expenditure in % of							
Passive labor market policy	72	76	80	83	80	77	82
Active labor market policy	28	24	20	17	20	23	18
Training programs	13	10	10	8	10	11	7
Unemployment rate in %	6.2	8.0	9.1	10.8	9.6	8.0	9.3

Source: Lechner, Miquel, and Wunsch (2005).

In Germany, labor market training consists of heterogeneous instruments that differ in the form and in the intensity of the human capital investment, as well as in their duration. We aggregate the different programs into four groups according to their selection of participants, educational content, and organization. *Practice firms* simulate working in a specific field of profession. Their mean duration is 6 months in our sample.<sup>5</sup> *Short training* comprises courses that provide a general adjustment of working skills. Their mean duration is 4 months and does not exceed 6 months. *Long training* is similar to short training but with a duration of more

<sup>4</sup> See Wunsch (2005) for a detailed account of the German labor market policy.

<sup>5</sup> All durations reported in this paper describe the courses, not the behavior of the participants. Thus, these durations are *planned* when the unemployed starts the course.

than 6 months and a mean duration of 11 months. *Re-training* courses enable working in a different profession than the one currently held by awarding new vocational degrees. Their mean duration is 20 months.

## **2.2 Data and definition of the sample**

We use a database obtained by merging administrative data from three different sources: the IAB employment subsample, the benefit payment register, and the training participant data. This is the most comprehensive database in Germany with respect to training conducted prior to 1998. We reconstruct the individual employment histories from 1975 to 1997. It also contains detailed personal, regional, employer, and earnings information. Thus, it allows controlling for many, if not all, important factors that determine selection into programs and labor market outcomes. Moreover, precise measurements of the interesting outcome variables are available up to 2002.

We consider program participation between 1993 and 1994. A person is included in our population of interest if he starts an unemployment spell between 1993 and 1994. The group of participants consists of all persons entering a program between the beginning of this unemployment spell and the end of 1994. We require that all individuals were employed at least once and that they received unemployment benefits or assistance before the start of the program. Finally, we impose an age restriction (25-55 years) and exclude trainees, home workers, apprentices and part-time workers. The resulting sample comprises about 9000 participants and about 270 to 550 participants in the 4 programs.<sup>6</sup>

Our outcome variables are annual employment and earnings during the seventh year after program start. This allows us to concentrate on the long-run effects, which are more interest-

---

<sup>6</sup> We use the same data as Lechner, Miquel, and Wunsch (2005). We also follow their definitions of populations, programs, participation, non-participation and their potential start dates, outcomes, and selection variables. See this paper for much more detailed information on all these topics.

ing policy parameters than the short-term effects, because the former are closer to the permanent effects of the program. Particularly for longer programs, the short-run effects are much influenced by the so-called lock-in effects (Van Ours, 2004), meaning that unemployed reduce their job search activities while being in the program.

### 2.3 Descriptive statistics

Table 2 shows descriptive statistics for selected socio-economic variables in the sub-samples defined by treatment and employment (employed / non-employed) status. This illustrates the 'double selection problem' for the estimation of program effects on earnings.

*Table 2: Descriptive statistics of selected variables by treatment and employment status*

	Non-participation		Practice firm		Short training		Long training		Re-training	
	E	NE	E	NE	E	NE	E	NE	E	NE
Number of observations	3211	5717	127	139	297	264	169	155	254	153
Monthly earnings (EUR)	1561	1462	1636	1548	1757	1656	1942	1669	1637	1519
Age (years)	34	39	35	36	34	36	34	36	30	31
Women (share in %)	38	44	40	29	36	40	40	39	38	37
German	83	81	88	86	93	88	91	94	90	88
Big city	25	27	17	20	22	25	24	34	19	22
Education: no degree	21	27	20	17	13	17	8	10	22	27
University degree	6	5	0	0	6	6	17	10	3	3
Salaried worker	30	28	35	32	40	37	63	52	25	20
Unskilled worker	37	41	34	39	26	36	15	25	51	54

Note: Means for the earnings variable computed 84 months after program start. *E* denotes employed and *NE* denotes non-employed (unemployed or out of labor force) in month 84. "Monthly earnings" are the monthly earnings in the last job prior to current unemployment.

Concerning selection into the programs, the results can be summarized as follows: Participants in re-training are younger compared to other unemployed, which is line with the idea that human capital investments are more beneficial if the productive period of the new human capital is longer. Interestingly the share of foreigners in the programs is only about half the share of foreigners in the group of non-participants. Participants in practice firms and re-train-

ing are less educated and less skilled. Past earnings are somewhat higher for participants in short and more strongly in long training than in practice firms and re-training.

As expected, we observe a positive selection into employment: Employed individuals are better educated, younger, and received higher salaries during their last occupation than non-employed individuals. Interestingly, they reside less frequently in a big city (reflecting the higher unemployment rates in German cities). Thus, there is a clear non-random selection into programs as well as into employment. Understanding and correcting for these two selection processes is the key to recover the 'pure' earnings effects of these training programs.

### **3 Notation, definitions, and effects**

#### **3.1 The standard model of potential outcomes**

To analyze the problem described in the previous section, it is necessary to introduce some notation. Each observation  $i$  in our large sample of size  $N$  is randomly drawn from a large population described by the joint distribution of the random variables  $(Y, S, D, X)$ . The variable  $Y$  and the binary variable  $S$  measure our outcomes of interest, namely earnings and employment. The binary variable  $D$  indicates participation in the training program. Individual characteristics are captured by  $X$  which is defined over a set  $\mathcal{X}$ . We follow the convention that random variables are denoted by capital letters, whereas their realisations are denoted by small letters. Thus, the sample contains the data  $\{y_i, s_i, d_i, x_i\}_{i=1}^N$ . Note that for ease of exposition, we assume there is only one program (and one employment state). This convention also indicates that we are interested in comparing the different participation states with each other. Of course, in the application there are many such binary comparisons that are of interest (see Imbens, 2000, and Lechner, 2001, for a formal multiple treatment framework).

We follow the standard approach in the microeconomic literature to use potential outcomes to define causal effects of interest. This approach was popularized by Rubin (1974), among others. As usual, we define potential values for the employment variable,  $S(d)$ , as well as for the earnings variable,  $Y(d)$ , with respect to program participation. Since potential earnings have a different interpretation when an individual is working compared to not working, we consider potential earnings as depending on two (binary) events, namely participation in a program ( $d=1$ ) and working ( $s=1$ ), i.e.  $Y(d,s)$ . Assuming the validity of the stable unit treatment value assumption (see Rubin, 1980) allows us to relate the different potential outcomes to each other and to the observable outcomes:

$$Y(d) = S(d)Y(d,1) + (1 - S(d))Y(d,0);$$

$$S = DS(1) + (1 - D)S(0);$$

$$\begin{aligned} Y &= DY(1) + (1 - D)Y(0) = \\ &= D[S(1)Y(1,1) + (1 - S(1))Y(1,0)] + (1 - D)[S(0)Y(0,1) + (1 - S(0))Y(0,0)]. \end{aligned}$$

Following the literature, we base our analysis on causal parameters that can be deduced from the differences of the marginal distributions of potential outcomes.<sup>7</sup> First, consider average and quantile treatment effects *on Y caused by D* for a population defined by a specific treatment status  $d$ . To define the quantile effects, let  $F_{V|W}(v; w)$  be the distribution function of  $V$  conditional on  $W$  evaluated at  $v$  and  $w$ .  $V$  and  $W$  may be vectors of random variables. The corresponding  $\theta^{\text{th}}$  ( $0 \leq \theta \leq 1$ ) quantile of  $F_{V|W}(v; w)$  is denoted by  $F_{V|W}^{-1}(\theta; w)$ . Using this definition, we obtain the following earnings effects of participating in a program:

---

<sup>7</sup> We do not investigate issues related to the joint distribution of potential outcomes, e.g.  $F_{Y(1)-Y(0)}(y)$ , since the latter is very hard to pin down with reasonable assumptions. For a thorough discussion of these issues, see

$$ATE^D(d) = E(Y(1)|D=d) - E(Y(0)|D=d);$$

$$QTE_\theta^D(d) = F_{Y(1)|D}^{-1}(\theta; d) - F_{Y(0)|D}^{-1}(\theta; d).$$

For  $d=1$ , we obtain the so-called treatment effects on the treated, whereas for  $d=0$  we obtain the treatment effects on the non-treated. The average effects unconditional on treatment status are thus a weighted average of those two effects. To minimize redundancies, we do not consider the latter effects explicitly.

These parameters defined for various outcome variables are the usual objects of investigation in empirical evaluation studies. However, depending whether individuals work ( $S=1$ ) or not ( $S=0$ ),  $Y$  measures very different objects. For working individuals, the data usually contains some earnings measure, whereas for non-working individuals it is either zero, or contains some non-wage income like unemployment or retirement benefits. In the former case, the causal effect would measure some productivity gain due to the program, whereas in the latter case we would estimate the impact of the program on a measure of disposable income. These parameters are interesting in their own right and are frequently estimated in empirical studies (e.g. Lechner, Miquel, and Wunsch, 2005). However, they fail to answer the important question whether the program would lead to earnings increases if employment had been found. The failure of answering this important policy question comes from the fact that the potential outcomes,  $Y(1)$  and  $Y(0)$ , are not defined conditional on the employment state, and thus mix employment and earnings effects.

Therefore, to answer questions about the potential earnings effects for individuals had they taken up a job after the program, we compare potential outcomes for different participation states in a (potential) world in which all individuals had found a job, which is not observable

---

Heckman, Smith, and Clemens (1997). Of course, this distinction does not matter for linear operators like the expectation, for example, since the expectation of the difference equals the difference of the expectations.

for non-working individuals. In particular, we investigate the (pure) earnings effects for those individuals who found a job under the treatment:

$$ATE^{D,1}(d,1) = E(Y(1,1)|D=d, S(d)=1) - E(Y(0,1)|D=d, S(d)=1);$$

$$QTE_{\theta}^{D,1}(d,1) = F_{Y(1,1)|D,S(d)}^{-1}(\theta; d,1) - F_{Y(0,1)|D,S(d)}^{-1}(\theta; d,1).$$

Since the problem is symmetric in  $d$ , we consider only the “doubly treated” population and concentrate on  $ATE^{D,1}(1,1)$  and  $QTE_{\theta}^{D,1}(1,1)$ . By doing so, we also refrain from explicitly investigating, for example, effects on benefits receipts that would be captured by  $ATE^{D,0}(d,1)$  and  $QTE^{D,0}(d,1)$ . Again, the technical arguments would be almost identical.

We could also consider the treatment effects for the whole population (irrespective of whether individuals have found a job or not). However, such effects may be of less policy interest than the effects for the effectively treated population, particularly in the context of narrowly targeted programs.

The effects for other populations have been considered in the literature as well. Recently Card, Michalopoulos, and Robins (2001) considered earnings effects for those workers who were induced to work by program participation. Similarly, Zhang and Rubin (2003) and Lee (2005) consider earnings effects for individuals who would work irrespective whether they participate in a program or not. Of course, both such populations are unobserved and, thus, difficult to describe. They cannot be characterized, for example, by simple descriptive statistics. Furthermore, Card, Michalopoulos, and Robins (2001) and Lee (2005) severely restrict the heterogeneity of the treatment effect. While Card, Michalopoulos, and Robins (2001) assume that the treatment effect on employment is positive for all observations, Lee (2005) assumes that this treatment effect is either positive for everybody or negative for everybody.



However, heterogeneous effects are a typical finding in program evaluation studies, as confirmed by our application.

Note that all subpopulations considered so far are defined by variables whose values are not caused by the treatment (note that although  $S$  is caused by  $D$ ,  $S(d)$  is by construction not caused by  $D$ ). For example, if we consider effects conditional on  $S$ , the causal interpretation of such effects is unclear, because part of the effect of  $D$  on  $S$ , and thus on  $Y$ , is already 'taken away' by the conditioning variable  $S$  (see Lechner, 2008). Therefore, we will *not* consider the effect of  $D$  on those participants and nonparticipants who actually found a job.

### 3.2 Unified notation for average and quantile effects

In this paper, we consider explicitly the identification and estimation of average *and* quantile treatment effects. To do so, we introduce a notation that encompasses both types of effects to avoid redundancies in our formal arguments.

Let  $g(\cdot)$  be a function mapping  $Y$  into the real line. We will show below that we only need to consider identification of  $E[g(Y(d,s))|X=x, D=d', S=s']$  for  $d, d', s, s' \in \{0,1\}$  and  $x \in \mathcal{X}$  to examine the identification of the average and quantile treatment effects. Letting  $g(Y)=Y$ , we obtain all *ATEs* defined above. Letting  $g(Y)=\mathbb{1}(Y \leq \tilde{y})$ , we identify the distribution function of  $Y$  evaluated at  $\tilde{y}$ .<sup>8</sup> The distribution function can then be inverted to get the quantiles of interest and to obtain all *QTEs* defined above.

Define  $\underline{K}_g \equiv \inf_y g(y)$  as lower bound of  $g(\cdot)$  and  $\bar{K}_g \equiv \sup_y g(y)$  as its upper bound. These bounds may or may not be finite depending on  $g(\cdot)$  and the support of  $Y$ . If we estimate the distribution function,  $g(\cdot)$  is an indicator function, which is naturally bounded between 0 and

---

<sup>8</sup> The indicator function  $\mathbb{1}(\cdot)$  equals one if its argument is true.

1. If we estimate the expected value of  $Y$ ,  $g(\cdot)$  is the identity function and  $\underline{K}_Y$  and  $\bar{K}_Y$  are the bounds of the support of  $Y$ . If we estimate the variance of  $Y$ ,  $g(Y) = (Y - E(Y))^2$ . In this case, and in the absence of further information on  $E(Y)$ , the lower bound on  $g(\cdot)$  is 0 and the upper bound is  $0.25(\bar{K}_Y - \underline{K}_Y)^2$ .<sup>9</sup>

Lemma 1 shows that tight bounds on the conditional expectations can be integrated to get tight bounds unconditionally on  $X$ .<sup>10</sup>

*Lemma 1 (bounds on the unconditional expected value of  $g(\cdot)$ )*

Let  $\underline{b}_g(x)$  and  $\bar{b}_g(x)$  be tight lower and upper bounds on  $E(g(Y)|X=x)$ . Then  $E(\underline{b}_g(X))$  and  $E(\bar{b}_g(X))$  are tight lower and upper bounds on  $E(g(Y))$ . This result holds in the population and all subpopulations defined by values of  $D$  and  $S$ .

The proof of this lemma (as well as all other proofs) can be found in the Technical Appendix. Naturally, if  $\underline{b}_g(x) = \bar{b}_g(x)$  for  $\forall x \in \mathcal{X}$ , then  $E(g(Y))$  is identified. Letting  $g(\cdot)$  be the identity function, we obtain sharp bounds on the average treatment effect on the treated:

$$\begin{aligned} E(Y|D=1, S=1) - E(\bar{b}_{Y(0,1)}(X)|D=1, S=1) \\ \leq ATE^{D,1}(1,1) \leq E(Y|D=1, S=1) - E(\underline{b}_{Y(0,1)}(X)|D=1, S=1). \end{aligned}$$

Similarly, by letting  $g(Y) = 1(Y \leq \tilde{y})$  and using the same principles, we obtain bounds on the unconditional distribution function. Lemma 2 shows how the bounds on the unconditional distribution function can be inverted to get bounds on the unconditional quantile function.

---

<sup>9</sup> The highest possible variance is obtained if  $y = \bar{K}_Y$  with probability 0.5 and  $y = \underline{K}_Y$  with probability 0.5.

<sup>10</sup> We define *tight* (or *sharp*) bounds as finite bounds that cannot be improved upon without further information.

*Lemma 2 (bounds on the quantile function)*

Let  $\underline{r}_Y(\tilde{y})$  and  $\bar{r}_Y(\tilde{y})$  be tight lower and upper bounds on the distribution function of  $Y$  evaluated at  $\tilde{y}$ . Let  $0 < \theta < 1$  and define  $\underline{r}_{QY}(\theta)$  and  $\bar{r}_{QY}(\theta)$  as follows:

$$\underline{r}_{QY}(\theta) \equiv \inf_{\tilde{y}} \{\bar{r}_Y(\tilde{y}) \geq \theta\} \quad \text{if } \lim_{\tilde{y} \rightarrow -\infty} \bar{r}_Y(\tilde{y}) > \theta,$$

$$\equiv -\infty \quad \text{otherwise;}$$

$$\bar{r}_{QY}(\theta) \equiv \sup_{\tilde{y}} \{\underline{r}_Y(\tilde{y}) \leq \theta\} \quad \text{if } \lim_{\tilde{y} \rightarrow \infty} \underline{r}_Y(\tilde{y}) < \theta,$$

$$\equiv \infty \quad \text{otherwise.}$$

The tight lower and upper bounds for the  $\theta^{\text{th}}$  quantile of  $Y$  are  $\underline{r}_{QY}(\theta)$  and  $\bar{r}_{QY}(\theta)$ .

If  $Y$  has a bounded support,  $-\infty$  and  $\infty$  are replaced by the bounds on that support. Note that the upper bound on the distribution function determines the lower bound on the quantiles (et vice versa). Furthermore, note that low quantiles are bounded by below and high quantile by above only if  $Y$  has a bounded support. The implication of Lemmas 1 and 2 is that we only need to determine tight bounds of the conditional expected value of  $g(Y(d, s))$ , in particular  $g(Y(0, 1))$ , to bound sharply the *ATEs* and *QTEs* of interest. This is done in the next section.

## 4 Identification

### 4.1 First step assumptions

To concentrate on the special problems coming from the 'double selection problem' into programs and employment, we assume that the data are rich enough to identify the distributions of the marginal potential outcomes,  $Y(d)$ , for all values of the treatment. Here, to keep the notation tractable and because we use this assumption in the application, we assume independence of treatment,  $D$ , and potential outcomes,  $Y(d)$ ,  $S(d)$ , conditional on confounders,  $X$ , as in the standard matching literature. There are other ways to identify  $Y(d)$  and  $S(d)$ , for ex-

ample using a continuous instrument as in Heckman and Vytlacil (2005). Our results concerning the identification of the effects on potential earnings do not depend on the assumption used to identify  $Y(d)$ .

It will be notationally convenient for the derivation of the technical properties in the next section to use a slightly stronger condition than required for the identification of the effect of  $D$  on  $S(0)$  and  $Y(0)$  alone. For the latter it would suffice that  $Y(0) [= SY(0,1) + (1-S)Y(0,0)]$  and  $S(0)$  are mutually independent of  $D$  conditional on  $X$ . Instead we assume that  $Y(0,1)$ ,  $Y(0,0)$ ,  $S(1)$  and  $S(0)$  are jointly independent of  $D$  conditional on  $X$ . In terms of our application, this additional restriction does not entail further substantive behavioral restrictions concerning the assignment process to the training program. Furthermore, to be able to recover the necessary information from the data, common support assumptions are added in part b). Note that the second part of the common support assumption is, again, not necessary for the identification of the distributions of  $Y(0)$  and  $S(0)$ . It is added to be used below when interest is in the identification of the distribution of  $Y(0,s)$ . Finally, in part c) of Assumption 1 we add standard regularity conditions guaranteeing that the objects of interest exist.

*Assumption 1 (conditional independence assumption for first stage)*

a) Conditional independences:  $\{Y(0,0), Y(0,1), S(0), S(1)\} \perp D | X = x$  for  $\forall x \in \mathcal{X}$ ;<sup>11</sup>

b) Common support:  $P(D = 1 | X = x) < 1$   $\forall x \in \mathcal{X}$ ;

$$P(S(d) = 1 | X = x) > 0 \text{ for } \forall d \in \{0,1\} \text{ and } \forall x \in \mathcal{X};$$

c)  $E[g(Y(d,s)) | X = x, D = 1, S = s]$  is finite for  $\forall s, d \in \{0,1\}$  and  $\forall x \in \mathcal{X}$ .

---

<sup>11</sup> This notation means that the joint distribution of  $Y(0,0)$ ,  $Y(0,1)$ ,  $S(0)$ , and  $S(1)$  conditional on  $X$  is independent of the distribution of  $D$  conditional on  $X$ . Conditional independence of the potential outcomes is sufficient for conditional independence for all functions  $g(\cdot)$  of the potential outcomes. Weaker conditions are sufficient for important special cases. For instance, mean independence is sufficient for ATE.

Lemma 3 states that these conditions are sufficient to identify the causal effects of  $D$  on earnings and employment outcomes.

*Lemma 3 (Assumption 1 identifies effects of  $D$  on  $S(d)$  and  $Y(d)$ )*

If Assumption 1 holds then  $E(S(1) - S(0) | D = 1)$  is identified. If  $g(Y) = Y$ , then  $ATE^D(1)$  is identified. If  $g(Y) = 1(Y \leq \tilde{y})$  for  $\forall \tilde{y}$  in the support of  $Y$ , then  $QTE_\theta^D(1)$  is identified for  $\forall \theta \in (0, 1)$ .

Imbens (2004) provides an excellent survey of estimators for ATEs consistent under Assumption 1. Efficient estimation of such average treatment effects is discussed for example in Hahn (1998), Heckman, Ichimura, and Todd (1998), Hirano, Imbens, and Ridder (2003), and Imbens, Newey, and Ridder (2005). Firpo (2007) and Melly (2006) discuss efficient estimation of quantile treatment effects. Note that by restricting  $X$  to be a constant we obtain the special case of a random experiment, which is analyzed by Lee (2005).

#### **4.2 Point identification of the effects on potential earnings**

The conditions necessary to identify the program effects on employment and on earnings, as discussed in the previous section, are not sufficient to identify the effects of  $D$  on the potential earnings given the employment status,  $Y(d, s)$ . One possible set of restrictions that lead to point identification of distributions of these potential outcomes is given in Assumption 2:

*Assumption 2 (conditional independence of potential earnings)*

a) Conditional independence:  $Y(d, 1) \perp S(d') | X = x, D = d'$  for  $\forall d, d' \in \{0, 1\}$  and  $\forall x \in \mathcal{X}$ ;

b) Common support:  $0 < P((1 - D)S(0) = 1 | X = x)$ , for  $\forall x \in \mathcal{X}$ .

Condition a) states that selection into employment is independent from potential earnings. Therefore, it is appropriate to compare working participants to non-working participants with the same characteristics  $X$ . Of course, depending on how informative  $X$  is, this assumption may contradict standard economic models designed to analyze individual employment decisions (e.g. Roy, 1951). Lemma 4 shows that Assumptions 1 and 2 are sufficient to identify the distribution of  $Y(d,1)$ :

*Lemma 4 (Assumptions 1 and 2 identify treatment effects on potential earnings)*

If Assumptions 1 and 2 are satisfied with  $g(Y) = Y$ , then  $ATE^{D,1}(1,1)$  is identified. If these assumptions hold with  $g(Y) = 1(Y \leq \tilde{y})$  for  $\forall \tilde{y}$  in the support of  $Y$ , then  $QTE_{\theta}^{D,1}(1,1)$  is identified for  $\forall \theta \in (0,1)$ ,  $d=0,1$ .

An alternative to identify the treatment effects on potential earnings is the presence of a continuous instrument for the participation decision  $S$ . The nonparametric identification of the resulting sample selection models is discussed in Das, Newey, and Vella (2003). It would be straightforward to use their results in our context. In our data set, as often in applications, there is no plausible continuous instrument. Discrete instruments (i.e. exclusion restrictions for discrete variables) do generally not allow identifying the above defined treatment effects. However, since they identify effects for some complier population, they do necessarily not lead to point identification of the causal effects defined above, but reduce the uncertainty about the true effects. Therefore, we discuss the case of discrete instruments further below.

### **4.3 Worst case bounds**

Since Assumption 2 is not plausible in our and probably the majority of applications and no continuous instruments are available for the second stage selection process, we give up on trying to achieve plausible point identification. Instead, we bound the treatment effects using

weaker assumptions that appear to be more reasonable in our empirical study (and many other applications).

Theorem 1 shows that knowing the effects of  $D$  on  $Y(d)$  and  $S(d)$  reduces the uncertainty. To state this theorem concisely, we denote the expected value of  $Y$  over its upper part of its distribution up to  $p$ -% largest values conditionally on  $X = x$  by  $E_{\max|p}(Y|X = x)$ . Similarly,

$E_{\min|p}(Y|X = x)$  denotes the same expected value but over the  $p$  fraction of the lower part of the distribution of  $Y$ .<sup>12</sup>

*Theorem 1 (worst-case bounds)*

Assumption 1 holds. If  $p_{S|X,D}(x, 0) + p_{S|X,D}(x, 1) > 1$ ,<sup>13</sup> then the lower and upper bounds on  $E[g(Y(0,1))|X = x, D = 1, S = 1]$  are given by

$$\underline{b}_{g(Y)}(x) = \frac{E_{\min|p_{S|X,D}(x,0)+p_{S|X,D}(x,1)-1}}{p_{S|X,D}(x,0)} (g(Y)|X = x, D = 0, S = 1) \frac{p_{S|X,D}(x, 0) + p_{S|X,D}(x, 1) - 1}{p_{S|X,D}(x, 1)} + \underline{K}_g \frac{1 - p_{S|X,D}(x, 0)}{p_{S|X,D}(x, 1)}, \quad \text{and}$$

$$\bar{b}_{g(Y)}(x) = \frac{E_{\max|p_{S|X,D}(x,0)+p_{S|X,D}(x,1)-1}}{p_{S|X,D}(x,0)} (g(Y)|X = x, D = 0, S = 1) \frac{p_{S|X,D}(x, 0) + p_{S|X,D}(x, 1) - 1}{p_{S|X,D}(x, 1)} + \bar{K}_g \frac{1 - p_{S|X,D}(x, 0)}{p_{S|X,D}(x, 1)}.$$

If  $p_{S|X,D}(x, 0) + p_{S|X,D}(x, 1) \leq 1$ , then the bounds are  $\underline{K}_g$  and  $\bar{K}_g$ .

<sup>12</sup> This type of notation can also be found in Zhang and Rubin (2003).

<sup>13</sup> Generally,  $p_{V|W}(w)$  denotes the probability that all elements of the vector of binary variables  $V$  are jointly equal to one, conditional on  $W = w$ .

Note that the bounds are observable. Theorem 1 shows that we can learn part of the nonparticipation-employment outcome of employed participants from the employment outcomes of non-participants. However, since Assumption 1 is silent about the selection of participants and non-participants into employment, there remains uncertainty coming from the employment outcomes of those who would not work if participating or who would not work if non-participating. Clearly, without further assumptions nothing can be learned about the average counterfactual outcome of working from those who would not work either as participants or as non-participants. The importance of this uncertainty decreases as the probability of non-working for participants and non-participants decreases.

It is clear that these bounds will be very wide if the employment probabilities are not high. In our application, the employment probabilities are low because we consider a sample of persons who are unemployed when treatment starts. The employment probability at the end of our sample period is never higher than 60%. Therefore, we cannot expect to obtain informative bounds without further restricting the selection process into employment. Such restrictions will be imposed below.

#### **4.4 Exclusion restrictions**

As already noted in Section 4.2, a continuous instrument for employment would allow identification of the effects of interest. Here, we analyze the more realistic case of discrete instruments. Although such instruments reduce uncertainty, they do not identify the effects.

Following Manski (1994, Section 3.1), we assume that  $Y$  is independent of  $Z$  given  $X$ :

*Assumption 3 (exclusion restriction)*

a) There is a random variable  $Z$  with support  $\mathcal{Z}$  such that:

$$Y(0,1) \perp Z \mid X = x, D = 1, S = 1, \forall x \in \mathcal{X}.$$

b) Assumption 1 holds with  $Z$  included in the list of control variables  $X$ .



Assumption 3-b) implies that the bounds derived in Theorem 1 are valid if we condition on  $X$  and  $Z$ . Assumption 3-a) implies that the bounds must be the same for all values of  $Z$ . Theorem 2 formalizes these intuitions:

*Theorem 2 (exclusion restriction)*

Assumptions 1 and 3 hold. For the case  $p_{S|X,Z,D}(x, z, 0) + p_{S|X,Z,D}(x, z, 1) > 1$  the lower and upper bounds are given by:

$$\begin{aligned} \underline{b}_{g(Y)}(x, z) &= \frac{E}{\min \left| \frac{p_{S|X,Z,D}(x, z, 0) + p_{S|X,Z,D}(x, z, 1) - 1}{p_{S|X,Z,D}(x, z, 0)} \right|} (g(Y)|X = x, Z = z, D = 0, S = 1) \times \\ &\quad \times \frac{p_{S|X,Z,D}(x, z, 0) + p_{S|X,Z,D}(x, z, 1) - 1}{p_{S|X,Z,D}(x, z, 1)} + \underline{K}_g \frac{1 - p_{S|X,Z,D}(x, z, 0)}{p_{S|X,Z,D}(x, z, 1)} \quad \text{and} \\ \bar{b}_{g(Y)}(x, z) &= \frac{E}{\max \left| \frac{p_{S|X,Z,D}(x, z, 0) + p_{S|X,Z,D}(x, z, 1) - 1}{p_{S|X,Z,D}(x, z, 0)} \right|} (g(Y)|X = x, Z = z, D = 0, S = 1) \times \\ &\quad \times \frac{p_{S|X,Z,D}(x, z, 0) + p_{S|X,Z,D}(x, z, 1) - 1}{p_{S|X,Z,D}(x, z, 1)} + \bar{K}_g \frac{1 - p_{S|X,Z,D}(x, z, 0)}{p_{S|X,Z,D}(x, z, 1)}. \end{aligned}$$

If  $p_{S|X,Z,D}(x, z, 0) + p_{S|X,Z,D}(x, z, 1) \leq 1$ , we get  $\underline{b}_{g(Y)}(x, z) = \underline{K}_g$  and  $\bar{b}_{g(Y)}(x, z) = \bar{K}_g$ . The lower bound on  $E[g(Y(0,1))|X = x, D = 1, S(1) = 1]$  is given by  $\sup_{z \in Z} \underline{b}_{g(Y)}(x, z)$  and the upper bound by  $\inf_{z \in Z} \bar{b}_{g(Y)}(x, z)$ .

#### 4.5 Positive selection into employment

In a standard labor supply models individuals accept a job offer if the offered wage is higher than same reservation wage, denoted by  $Y^R$ , i.e.  $S(0) = \mathbb{1}(Y(0,1) \geq Y^R)$ . This relation motivates the assumption that the employment probability conditional on  $X$  should be smaller for smaller potential earnings than for higher potential earnings. Therefore, we get  $\Pr(S(0) = 1|X = x, Y(0,1) \leq \tilde{y}) \leq \Pr(S(0) = 1|X = x, Y(0,1) > \tilde{y})$ , if  $Y(0,1)$  and  $Y^R$  are not too

strongly correlated (see Blundell, Gosling, Ichimura, and Meghir, 2007). Such a condition is equivalent to assuming that the distribution of  $Y(0,1)$  given  $S(0) = 1$  stochastically dominates the distribution of  $Y(0,1)$  given  $S(0) = 0$  and is stated formally as follows:<sup>14</sup>

*Assumption 4 (positive selection into employment of nonparticipants)*

$$F_{Y(0,1)|X,D,S(0)}(\tilde{y}; x, 0, 0) \geq F_{Y(0,1)|X,D,S(0)}(\tilde{y}; x, 0, 1).$$

Note that the positive selection condition is only imposed on those individuals not participating in a program. Assumption 4 tightens the bounds derived in Theorem 1:

*Theorem 3 (positive selection into employment)<sup>15</sup>*

a) If Assumptions 1 and 4 hold, and  $g(\cdot)$  is an monotone increasing function, then:

$$E[g(Y(0,1))|X = x, D = 1, S = 1] \leq \max_{p_{S|X,D}(x,1)} E[g(Y)|X = x, D = 0, S = 1].$$

b) If Assumptions 1 and 4 hold and  $g(\cdot)$  is a monotone decreasing function, then:

$$E[g(Y(0,1))|X = x, D = 1, S = 1] \geq \min_{p_{S|X,D}(x,1)} E[g(Y)|X = x, D = 0, S = 1].$$

Note that the positive selection assumptions tighten only one of the two bounds of the treatment effects.

#### **4.6 Conditional uniformity of the treatment effect on employment**

Lee (2005) restricts the individual treatment effect on the employment probability to have the same sign for all of the population. He calls this a *monotonicity* assumption.<sup>16</sup> Although, Lee's assumption is similar to the *monotonicity* assumption of Imbens and Angrist (1994), they re-

<sup>14</sup> See Blundell, Gosling, Ichimura, and Meghir (2007) for a proof.

<sup>15</sup> If  $g(\cdot)$  is not monotonic, Assumption 4 can be replaced by  $E[g(Y(0,1))|X = x, D = 0, S(1) = 1, S(0) = 1] \geq E[g(Y(0,1))|X = x, D = 0, S(1) = 1, S(0) = 0]$  for part a) or with a  $\leq$  sign for part b).

strict the effect of the instrument on the treatment status, while Lee restricts the effect of the treatment on sample selection. Since *monotonicity* may be considered a strange name for these assumptions (the effect on a binary variable is necessarily monotonous), we call this assumption *uniformity*.<sup>17</sup>

Lee's (2005) assumption appears to be overly restrictive for the type of application we consider. For instance, it excludes the possibility that a training program has positive effects on long-term unemployed but negative effects on short-term unemployed. However, this type of heterogeneity is typically found in the literature. Thus, we impose the weaker assumption that the direction of the effect on employment is the same for all individuals with the same characteristics  $X$ . This assumption is satisfied if the vector of characteristics is rich enough to capture the program effect heterogeneity on employment.

A second difference with Lee (2005) is that we bound the effect for an observable population. Lee bounds the effect on earnings for the population who would work with or without the program. Therefore, if the program has a positive effect on employment, then he bounds the effects for the non-treated population, while if the program has a negative effect on employment, he bounds the effects on the treated. When the employment effect is heterogeneous with respect to  $X$ , the population for which the effect is estimated is a mixture of treated and non-treated, which is unobservable and difficult to interpret, and thus of limited use as a policy parameter.

The third difference with Lee (2005) is that we consider a broader range of identifying assumptions for the first step of the selection process, thus making the approach applicable outside the setting of random experiments.

---

<sup>16</sup> The same assumption is also made by Zhang and Rubin (2003).

The formal definition of *uniformity* is given in Assumption 5:

*Assumption 5 (conditional uniformity of the treatment effect on employment)*

For each  $x \in \mathcal{X}$ , either a)  $P(S(1) \geq S(0) | X = x, D = 0) = 1$ ,  
or b)  $P(S(1) \leq S(0) | X = x, D = 0) = 1$ .

Theorem 4 shows that Assumption 5 allows tightening the bounds considerably:

*Theorem 4 (conditional uniformity of the treatment effect on employment)*

a) Assumptions 1 and 5-a) hold. The bounds are given by the following expressions:

$$\begin{aligned} E[g(Y)|X = x, D = 0, S = 1] &\frac{p_{S|X,D}(x, 0)}{p_{S|X,D}(x, 1)} + \underline{K}_g \frac{p_{S|X,D}(x, 1) - p_{S|X,D}(x, 0)}{p_{S|X,D}(x, 1)} \leq \\ &\leq E[g(Y(0,1))|X = x, D = 1, S = 1] \leq \\ &\leq E[g(Y)|X = x, D = 0, S = 1] \frac{p_{S|X,D}(x, 0)}{p_{S|X,D}(x, 1)} + \bar{K}_g \frac{p_{S|X,D}(x, 1) - p_{S|X,D}(x, 0)}{p_{S|X,D}(x, 1)}. \end{aligned}$$

b) Assumptions 1 and 5-b) hold. The bounds are given by the following expressions:

$$\begin{aligned} \min \left| \frac{p_{S|X,D}(x,1)}{p_{S|X,D}(x,0)} \right| E[g(Y)|X = x, D = 0, S = 1] &\leq E[g(Y(0,1))|X = x, D = 1, S = 1] \leq \\ &\leq \max \left| \frac{p_{S|X,D}(x,1)}{p_{S|X,D}(x,0)} \right| E[g(Y)|X = x, D = 0, S = 1]. \end{aligned}$$

Interestingly, we obtain point identification if  $p_{S(1)|X}(x) = p_{S(0)|X}(x)$  (i.e.  $p_{S|X,D}(x, 1) = p_{S|X,D}(x, 0)$ ). The reason is that under the uniformity assumption, both treatment and control groups are comprised of individuals whose sample selection was unaffected by the assignment to treatment, and therefore the two groups are comparable. Sample selection correction

---

<sup>17</sup> These assumptions are fundamentally different from the monotone treatment response assumption of Manski (1997) and from the monotone instrumental variables assumption of Manski and Pepper (2000), because those authors assume certain functions to be monotone.

procedures are similar in this respect because they condition on the participation probability. However, they require continuous exclusion restrictions to achieve nonparametric identification. In the absence of such exclusion restrictions, there is only identification if the employment probabilities are, by chance, the same.

Theorem 4-b) comprises the result of Proposition 4 in Lee (2005) as a special case. This result has the appealing feature that the bounds do not depend on the support of  $g(\cdot)$ . Thus, the bounds are finite even when the support of  $Y$  is infinite. Obviously, this is irrelevant for the distribution function or if the support of  $Y$  is naturally bounded.

#### **4.7 Combination of assumptions**

Combining Assumptions 3, 4, and 5 leads to tighter bounds. The exclusion restriction is particularly easy to combine with any other assumption. The lower (upper) bound is given by the maximum (minimum) of the lower (upper) bound evaluated at each value the instrument can take. Combined with the uniformity assumption, an exclusion restriction is powerful if there is a value of the instrument such that the employment probabilities are (almost) the same for the participants and the non-participants. As discussed in Angrist (1997), sample selection models are working this way. The difference is that a discrete exclusion restriction identifies intervals and not points, because we will generally not find a value of the instrument that attains the equality exactly.

Next, we examine the combination of the positive selection assumption and the conditional uniformity assumption. Adding positive selection as defined in Assumption 4 to the conditional uniformity assumption tightens the lower bound on  $E[g(Y(0,1))|X = x, D = 1, S = 1]$ :

*Theorem 5 (positive selection into employment and uniformity)*

a) Assumptions 1, 4, and 5-a) hold. If  $g(\cdot)$  is an monotone increasing function, then, the upper bound given in Theorem 4-a) tightens to:

$$E\left[g(Y)|X=x, D=0, S=1\right] \frac{p_{S|X,D}(x,0)}{p_{S|X,D}(x,1)} + \\ + \frac{E}{\max\left|\frac{p_{S|X,D}(x,1)-p_{S|X,D}(x,0)}{p_{1-S|X,D}(x,0)}\right|} \left(g(Y)|X=x, D=0, S=1\right) \frac{p_{S|X,D}(x,1)-p_{S|X,D}(x,0)}{p_{S|X,D}(x,1)} .$$

b) Assumptions 1, 4, and 5-a) hold. If  $g(\cdot)$  is a monotone decreasing function, then the lower bound given in Theorem 4-a) tightens to:

$$E\left(g(Y)|X=x, D=0, S=1\right) \frac{p_{S|X,D}(x,0)}{p_{S|X,D}(x,1)} + \\ + \frac{E}{\min\left|\frac{p_{S|X,D}(x,1)-p_{S|X,D}(x,0)}{p_{1-S|X,D}(x,0)}\right|} \left(g(Y)|X=x, D=0, S=1\right) \frac{p_{S|X,D}(x,1)-p_{S|X,D}(x,0)}{p_{S|X,D}(x,1)} .$$

Theorem 5 has two limitations that we will remedy by changing slightly the formulation, but not the substance, of the positive selection assumption. First, the positive selection assumption compares  $F_{Y(0,1)|X,S(0)}(\tilde{y}; x, 0)$  and  $F_{Y(0,1)|X,S(0)}(\tilde{y}; x, 1)$ , but not  $F_{Y(0,1)|X,S(0),S(1)}(\tilde{y}; x, 0, 1)$  and  $F_{Y(0,1)|X,S(0),S(1)}(\tilde{y}; x, 1, 1)$ . Therefore, it is possible that  $F_{Y(0,1)|X,S(0),S(1)}(\tilde{y}; x, 0, 1)$  dominates  $F_{Y(0,1)|X,S(0),S(1)}(\tilde{y}; x, 1, 1)$ , although at the same time  $F_{Y(0,1)|X,S(0)}(\tilde{y}; x, 1)$  dominates  $F_{Y(0,1)|X,S(0)}(\tilde{y}; x, 0)$ . This implausible scenario is ruled out in Assumption 6:

*Assumption 6 (positive selection into employment conditionally on  $S(1)=1$ )*

$$F_{Y(0,1)|X,S(0),S(1)}(\tilde{y}; x, 0, 1) \geq F_{Y(0,1)|X,S(0),S(1)}(\tilde{y}; x, 1, 1) .$$

Combining Assumption 6 with the conditional uniformity assumption leads to simple and intuitive bounds

*Theorem 6 (positive selection into employment conditionally on  $S(1)=1$  and uniformity)*

a) Assumptions 1, 5-a), and 6 hold. If  $g(\cdot)$  is a monotone increasing function, then:

$$E[g(Y)|X=x, D=0, S=1] \geq E[g(Y(0,1))|X=x, D=1, S=1].$$

b) Assumptions 1, 5-a), and 6 hold. If  $g(\cdot)$  is a monotone decreasing function, then:

$$E[g(Y)|X=x, D=0, S=1] \leq E[g(Y(0,1))|X=x, D=1, S=1].$$

The intuition for this result is better understood with an example. Suppose that a program has a positive effect on employment. This means that the "quality" of the employed participants is lower than that of the employed non-participants.<sup>18</sup> If, despite this lower quality, the program effect on the observed earnings is positive, this must imply that the program has a positive effect on potential earnings for the treated.

Note that neither Theorem 5 nor Theorem 6 allows to tighten the bounds if  $P(S(1) \leq S(0) | X=x) = 1$ . The intuition for this results is that in this case, all observations with  $S(1)=1$  also have  $S(0)=1$ . Thus, the problem for identifying the counterfactual mean is not that we do not know the value for the population with  $S(0)=0$  (this is irrelevant for the estimation of the effects on the doubly treated population), but that we do not know which of the observations with  $S(0)=1$  have  $S(1)=1$  as well. To tighten the bounds in this particular case, we suggest Assumption 7:

*Assumption 7 (positive selection into employment for  $Y(0,1)$  with respect to  $S(1)$ )*

$$F_{Y(0,1)|X,D,S(0),S(1)}(\tilde{y}; x, 0, 1, 0) \geq F_{Y(0,1)|X,D,S(0),S(1)}(\tilde{y}; x, 0, 1, 1).$$

---

<sup>18</sup> The positive selection assumption implies that the higher the employment probability the lower the "quality" of the workers.

Note that this assumption is conceptually different from Assumptions 4 and 6 because it relates the control outcome to the treated employment status and is therefore more restrictive. Similar assumptions have been made by Angrist, Bettinger, Bloom, King, and Kremer (2002, especially footnote 20), Zhang and Rubin (2004, Assumption 2) and Angrist, Bettinger, and Kremer (2006, especially proposition 1). To motivate this assumption, suppose that  $Y(1,1) = Y(0,1) + \alpha$ , with  $\alpha \geq 0$  and suppose further that unemployed individuals accept a job if their potential earnings exceeds a certain threshold,  $\mu: S(d) = 1(Y(d,1) \geq \mu)$ , for  $d \in \{0,1\}$ .

This implies the following inequalities:

$$\begin{aligned} E(Y(0,1)|S(0)=1, S(1)=1) &= E(Y(0,1)|S(1)=1) = E(Y(0,1)|Y(1,1) \geq \mu) \geq \\ &\geq E(Y(0,1)|Y(1,1) \geq \mu - \alpha) = E(Y(0,1)|Y(0,1) \geq \mu) = E(Y(0,1)|S(0)=1). \end{aligned}$$

Since  $E(Y(0,1)|S(0)=1)$  is a weighted average of  $E(Y(0,1)|S(0)=1, S(1)=1)$  and  $E(Y(0,1)|S(0)=1, S(1)=0)$ , the inequality implies that Assumption 7a) is satisfied for  $Y(0,1)$ .

*Theorem 7 (positive selection into employment with respect to  $S(1)$  and uniformity)*

a) Assumptions 1, 5-b), and 7 hold. If  $g(\cdot)$  is a monotone increasing function, then:

$$E[g(Y)|X=x, D=0, S=1] \leq E[g(Y(0,1))|X=x, D=1, S=1].$$

b) Assumptions 1, 5-b), and 7 hold. If  $g(\cdot)$  is a monotone decreasing function, then:

$$E[g(Y)|X=x, D=0, S=1] \geq E[g(Y(0,1))|X=x, D=1, S=1].$$

The intuition for this result is the same than for the result of Theorem 6.



## 5 Estimation

This paper focuses on the identification issues as well as on the empirical study that motivated the methodological innovation. Naturally, we bridge the gap between the identification results and the empirical study by proposing some estimators as well. However, due to space constraints we keep this part of the paper brief. We start by proposing consistent, nonparametric estimators. However, the combination of the dimension of the control variables and the sample sizes in this application are such that a fully nonparametric estimation strategy would lead to very imprecise estimators. Therefore, in Section 5.2 we suggest to use a (parametric) propensity score to reduce the dimension of the estimation problem and so to gain precision.

### 5.1 Nonparametric estimators

Here, we provide consistent, nonparametric estimators for all elements appearing in the different bounds of Theorems 1 to 7. Since we are interested in average as well as quantile effects, we consider two special cases of the  $g$ -function, namely  $g(Y) = Y$  and  $g(Y) = \mathbb{1}(Y \leq \tilde{y})$ .

The conditional employment probabilities  $p_{s|x,d}(x,d)$  for  $d \in \{0,1\}$  could be estimated nonparametrically using Nadaraya-Watson or local linear regression. However, a local nonlinear estimator (Fan, Heckman, and Wand, 1995), like a local probit for instance, should be more suited for binary dependent variables.<sup>19</sup>

$E[\mathbb{1}(Y \leq \tilde{y})|X = x, D = 0, S = 1]$  could be estimated by a local probit as well. However, for the  $QTEs$  we need to estimate the conditional distribution function evaluated at a large number of  $\tilde{y}$ , which is computationally very intensive.<sup>20</sup> Moreover, since we need to estimate the complete conditional distribution anyway, it is natural and faster to estimate the whole distri-

---

<sup>19</sup> Moreover, in Frölich (2006) the local parametric estimator appears to have better small sample properties.

bution by using locally weighted quantile regressions (Chaudhuri, 1991). By exploiting the linear programming representation of the quantile regression problem, it is possible to estimate all quantile regression coefficients efficiently (see Koenker, 2005, Section 6.3). The estimated conditional quantiles, though not necessarily monotonous in finite samples, may be inverted using the strategy proposed by Melly (2006) to get the estimated conditional distribution function.

The conditional expectations of earnings,  $E(Y|X = x, D = 0, S = 1)$ , is estimated by a local linear least squares regression.

The majority of the bounds for the mean contain conditional, asymmetrically trimmed means like  $E_{\max|p}(Y|X = x, D = 0, S = 1)$  and  $E_{\min|p}(Y|X = x, D = 0, S = 1)$ .<sup>21</sup> Lee (2005) proposes an estimator for the case with discrete  $X$ . We propose a new estimator allowing for discrete and continuous  $X$ . Koenker and Portnoy (1987) suggest an estimator based on linear quantile regression that allows estimating conditional trimmed means. They consider estimators of the form  $\int_0^1 J(\theta)\hat{\beta}(\theta)d\theta$ , where  $\hat{\beta}(\theta)$  is the  $\theta^{\text{th}}$  quantile regression coefficient vector. We apply

their estimator with a particular weight function,  $J(\theta)$ , and use nonparametric quantile regression. We estimate  $E_{\max|p}(Y|X = x, D = 0, S = 1)$  by  $\int_{1-p}^1 x\hat{\beta}(x, \theta)d\theta$  and

$E_{\min|p}(Y|X = x, D = 0, S = 1)$  by  $\int_0^p x\hat{\beta}(x, \theta)d\theta$  where  $\hat{\beta}(x, \theta)$  is the  $\theta^{\text{th}}$  local linear quantile regression evaluated at  $x$ .

---

<sup>20</sup> This is particularly problematic, because we rely on bootstrap based inference.

Lemma 1 shows that the bounds of the unconditional expected values equal the expected values of the conditional bounds. Thus, we estimate the bounds on the *ATE* by the mean of the conditional bounds evaluated at the treated observations. Similarly, for the *QTE*, we estimate the unconditional distribution by integrating the bounds on the conditional distribution. These bounds, which are monotone, are inverted to obtain the bounds of the *QTE* as shown in Lemma 2.

## 5.2 Using the propensity score to reduce the dimensionality of the problem

In our application, the number of control variables  $X$  necessary to make Assumption 1 plausible is too high to attempt a fully nonparametric estimation strategy, even with very large samples. Rosenbaum and Rubin (1983) show that the propensity score represents a useful dimension reduction device, because conditional independence of assignment and treatment (Assumption 1a) holds conditional on the (one-dimensional) propensity score as well:

$$\{Y(0,0), Y(0,1), S(0), S(1)\} \perp D | X = x \Rightarrow \{Y(0,0), Y(0,1), S(0), S(1)\} \perp D | p_{D|X}(x).$$

We estimate the propensity scores (for each program compared to nonparticipation) with parametric binary probits.<sup>22</sup> In a second step, we estimate the response functions and bounds nonparametrically conditional on the propensity score, allowing for arbitrary effect heterogeneity.

---

<sup>21</sup> For the distribution function,  $E_{\max|p} (1(Y \leq \tilde{y}) | X = x, D = 0, S = 1) = 1$  if  $1 - p < E(1(Y \leq \tilde{y}) | X = x, D = 0, S = 1)$

and  $E_{\max|p} (1(Y \leq \tilde{y}) | X = x, D = 0, S = 1) = [1 - E(1(Y \leq \tilde{y}) | X = x, D = 0, S = 1)] / p$  otherwise, such that we do

not need to estimate a trimmed mean. A similar result holds for the lower bound.

<sup>22</sup> Drake (1993) and Zhao (2005) find that estimators based on misspecified propensity scores were only slightly biased and much less biased than estimators based on incorrect response models.

Similarly, if Assumption 4, 6, and 7 (positive selection into employment) are valid conditionally on  $X$ , they are also valid conditionally on the propensity score. In fact, these assumptions are less restrictive conditional on the propensity score, as the score is less fine than  $X$ .

In contrast, conditioning only on the propensity score instead of  $X$  would considerably strengthen Assumption 5 (uniformity). The uniformity assumption states that the sign of the program effect on employment is the same for all observations with the same value of  $X$ . Therefore, the conditioning set must capture the heterogeneity of the employment effects. Since the conditioning set must also satisfy Assumption 1, we condition on the propensity score as well as on variables suspected to be related to employment effect heterogeneity.

## **6 The earnings effects of training programs in West Germany**

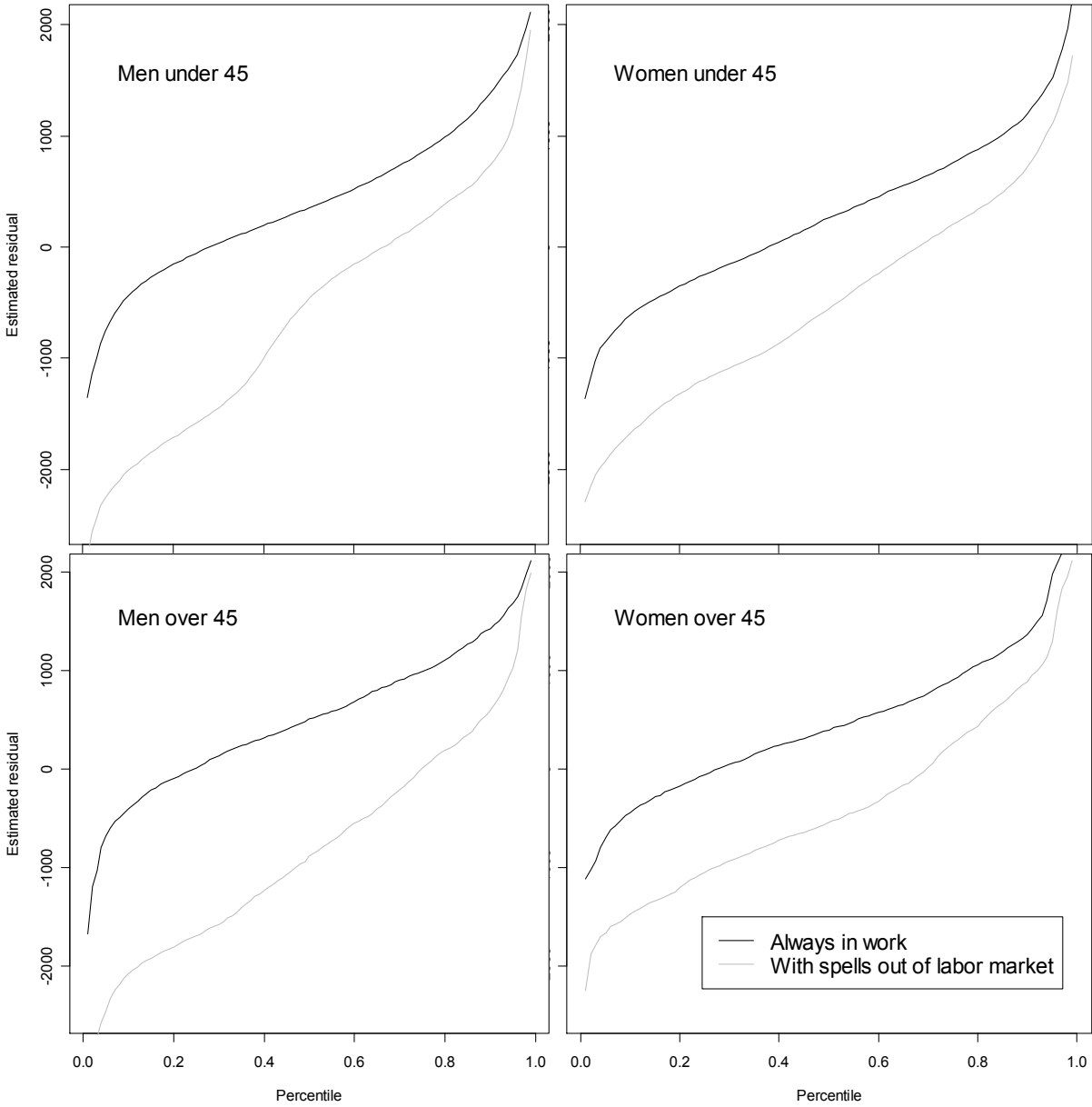
### **6.1 Validity of the identifying assumptions**

As explained in Section 2, we use the same data and variables as Lechner, Miquel, and Wunsch (2005). They discuss extensively why Assumption 1 is plausible in this setting. Briefly, their argument is that the data, which was specifically compiled to evaluate these programs, contains the major variables that jointly influence (marginal) outcomes and participation in the different training programs. For example, we control for education, age, family status, detailed regional differences, as well as previous employment histories including earnings, and position in job, specific occupation, and industry. Some potentially important factors are still missing such as ability, motivation, and jail and detailed health histories, but we are confident to capture indirectly these factors with almost 20 years of employment histories and the other covariates.

Assumptions 4, 6, and 7 (positive selection into employment) could be violated by a strong enough positive relationship between actual earnings and reservation earnings. This could be particularly the case for women, if high earnings women are matched with high earnings men,

or for older unemployed with high productivity, if they have saved enough to retire. Figure 1 uses the panel structure of our data to show that such effects may not be important. Following Blundell, Gosling, Ichimura, and Meghir (2007) we estimate cross sectional earnings equations for each month during the 6<sup>th</sup> and 7<sup>th</sup> year after program start. We then split the sample into those who were employed during the whole period and those who were not employed at least one month. Figure 1 shows that the residuals distribution of those always employed lies

Figure 1: Distribution of residual earnings by gender, age, and employment histories



below the distribution of those sometimes unemployed. Although this is not a true test of the, formally untestable, positive selection assumption, it does provide some credibility for this assumption.

The uniformity assumption will be satisfied if we capture the heterogeneity of the treatment effects on employment. Lechner, Miquel, and Wunsch (2005) find four variables related significantly to the heterogeneity of the employment effects for at least one of the four programs: the regional unemployment rate, residence in big towns, sex, and long-term unemployment before the program. Therefore, we control for these four variables in addition to the propensity score.

## **6.2 Implementation of the estimation and inference procedures**

We use the estimators presented in Section 5 with the propensity scores based on binary probits.<sup>23</sup> All bandwidths necessary to implement the nonparametric regressions are chosen by cross-validation. The bandwidths depend on the program, the dependent variable (employment or earnings) and on the number of regressors. The same bandwidths are used for mean and quantile regressions.

For most cases average treatment effects are unbounded if the support of earnings is unbounded. Here the support is naturally bounded: Due to the regulations of the social security system, from which database results, earnings are top-coded. This ceiling is however high, particularly for the low-earnings population we consider. It is attained by less than 1% of the observations in our sample. Thus, it is used as an upper bound together with zero as the lower bound.

---

<sup>23</sup> Lechner, Miquel, and Wunsch (2005) use a multinomial probit. We use binomial probits to reduce the computation time. Furthermore, the correlations between the estimated probabilities resulting from both estimators are higher than 98%.

We estimate the variance of the estimators by the standard nonparametric bootstrap. The heuristic motivation for the bootstrap is the following: First, note that the bounds implied by the exclusion restriction involve maximum and minimum operators. Thus, it is not clear whether the bootstrap is consistent (e.g., see Horowitz, 2001). However, in the application there are no plausible exclusion restrictions. Therefore, this potential problem is not an issue.

The other conditional bounds are free from any discontinuity and the estimators for them are continuous functions of estimators for which the regularity conditions of the bootstrap hold. The bounds for the treatment effects are estimated by integrating over the conditional bounds that are estimated by local linear methods. This is very similar to the estimator suggested by Heckman, Ichimura, and Todd (1998) for the average treatment effect, for which the bootstrap is known to be consistent.

### **6.3 *Standard earnings and employment effects***

We investigate the long-run effects of the training programs on earnings (and employment) by estimating the effects on annual earnings in the seventh year after program start. Before presenting the results for the potential earnings, we show standard employment and earnings effects as benchmark. The upper panel of Table 3 presents the means of the outcome variables for the non-participants and the participants to the four programs considered. Of course, the differences between these means have no causal interpretation because they are computed for different population. Therefore, Table 3 presents also the estimated ATET. They are similar to the results of Lechner, Miquel, and Wunsch (2005) but are not exactly identical, because we use local linear regression estimators and they used a matching estimator, and because they consider monthly instead of yearly outcome variables.

The results for employment show that all programs have a positive effect on employment. The effects on total earnings are positive for all programs, but it is impossible to know whether

they are only driven by the effects on employment or whether they reflect an improved earnings capacity. The estimated effects on earnings for the sub-samples of employed individuals are only valid if employment and earnings are independent. This assumption is probably not satisfied and these results are, therefore, difficult to interpret.

Table 3: Average employment and earnings effects ( $Y(1)-Y(0)$ ,  $S(1)-S(0)$ )

Population	Non-participants	Practice firms	Short training	Long training	Re-training
Mean:					
Employment	0.45	0.58	0.63	0.62	0.73
Earnings	8619	10429	13601	15745	15920
Earnings given employment	19272	17897	21738	25254	21743
ATET on:					
Employment		<b>0.08</b> (0.04)	<b>0.10</b> (0.03)	<b>0.09</b> (0.04)	<b>0.14</b> (0.03)
Earnings		1234 (877)	<b>3117</b> (690)	<b>3597</b> (1122)	<b>4816</b> (849)
Earnings given employment		-325 (1098)	<b>2031</b> (708)	<b>2926</b> (1113)	<b>2972</b> (763)

Note: The employment indicator is one if an individual worked at least one month in year 7. Earnings are defined as gross yearly earnings in year 7. Earnings for non-employed are coded as zero. Effects for "earnings given employment" are estimated on the subsamples of individuals with non-zero earnings. Bold numbers indicate significance at the 5% level.

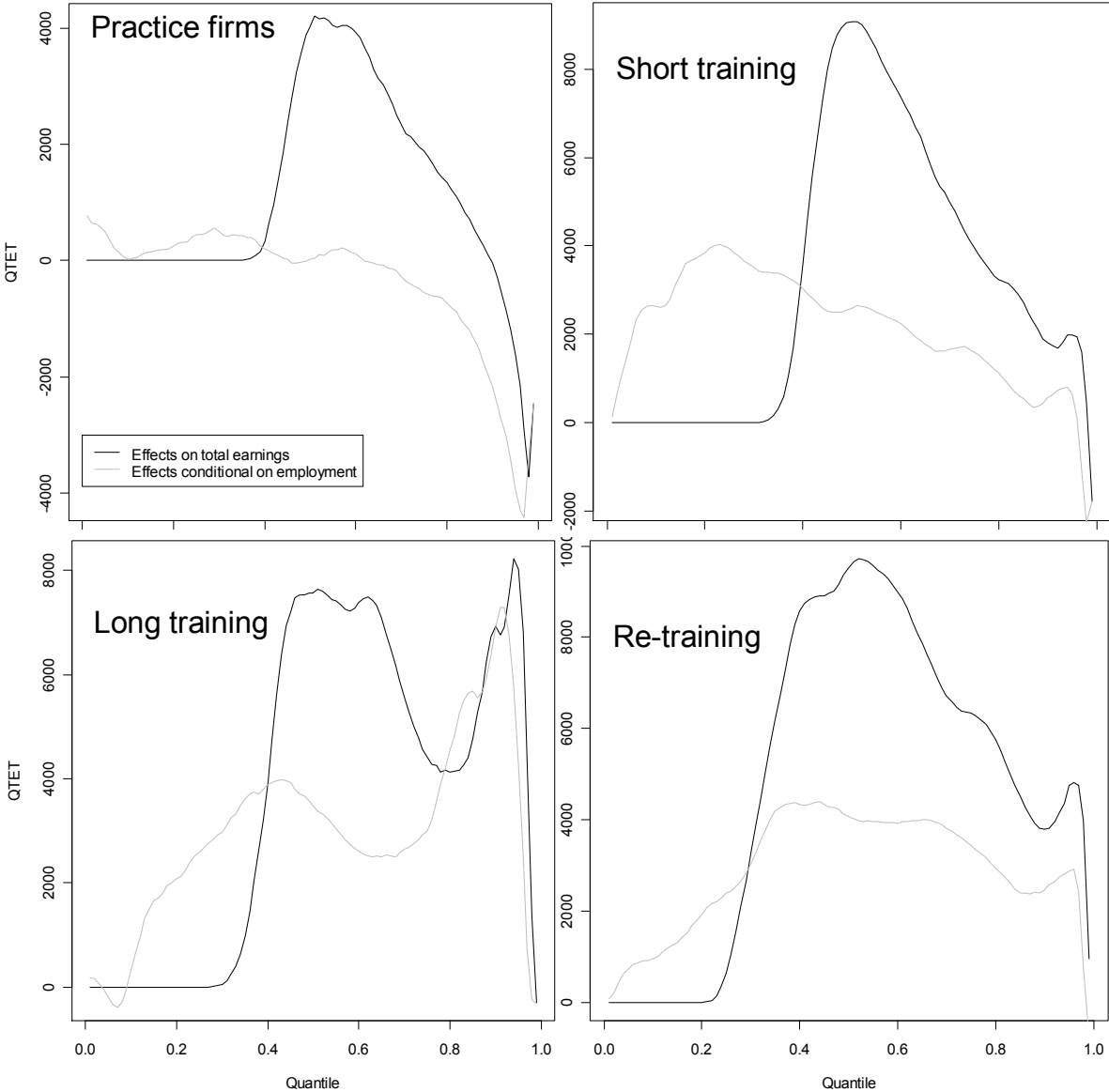
The quantile treatment effects on earnings, that are new, are presented in Figure 2.<sup>24</sup> They are even more difficult to interpret than the ATET. A substantial proportion of individuals are still unemployed whether they participated in a training program or not. Therefore, quantile treatment effects are zero for the lower part of the distribution. After that, participants are em-

<sup>24</sup> Since the sample objective function defining quantiles is non-differentiable, some bounds may slightly jump from one quantile to the other. Therefore, we use bagging (*bootstrap aggregating*) to smooth the results by defining the estimator of the bound to be the mean of the estimates obtained in 200 bootstrap samples. Lee and Yang (2006) and Knight and Bassett (2002) provide justification for bagging quantile regressions.



ployed and the non-participants are unemployed. Consequently, the quantile treatment effects increase strongly but this is a pure employment effect. Finally, the quantile treatment effects stabilize when both participants and non-participants are employed in the upper part of the distribution. Furthermore, Figure 2 also shows the effects conditionally on being employed, but they are probably biased because of the sample selection issue that is the key topic of this paper.

Figure 2: Quantile earnings effects (Y(1)-Y(0))



Note: See note below Table 2. The standard errors, not plotted to avoid overloading the figure, amount to about 1200 such that most of the positive quantile treatment effects on total earnings are significant. The quantile treatment effects conditional on employment are mainly significantly different from zero for short, long and re-training.

To conclude, it is obvious that such results usually estimated and reported in evaluation studies are unable to reveal the effects of the training programs on the earnings capacity of the unemployed. Next, we present the results that are informative on that issue.

#### **6.4 Bounds on the potential earnings effects**

Table 4 shows the bounds for the *ATETs* and the *QTETs*. For the latter we present three selected quantiles (0.25, 0.5, and 0.75). As discussed in Imbens and Manski (2004), for inference we may be interested in estimating confidence intervals that cover the entire identified interval with fixed probability or in confidence intervals that cover the true value of the parameter with a fixed probability. Since we are ultimately interested in the treatment effects and not in the bounds, the second type of confidence interval appears to be more appropriate, while the first one is more conservative. Therefore, Table 4 shows both types of confidence intervals, but the conclusions from the empirical results are robust to the type of confidence interval used.

For all programs and all treatment effects, the worst-case bounds are extremely wide. Imposing positive selection increases significantly the lower bounds but, with possibly one exception, is not sufficient to exclude zero effects. The uniformity assumption allows tightening upper and lower bounds compared to the worst-case bounds, but it is not powerful enough to reject the absence of earnings effect. However, the combination of positive selection with uniformity leads to informative bounds. We present the results for the three different definitions of positive selection as discussed in Section 4. Note that the definition based on Theorem 7 is more restrictive than the other two versions, because it relates potential outcomes for different treatment. Moreover, in our case the gain from imposing this assumption is small, because the employment effects of the programs are positive for the majority of the individuals. Therefore, our preferred bounds are those implied by Theorem 6 that are now discussed in more detail.

Table 4: Bounds on the ATETs and QTETs of the programs practice firms and short training

	ATET		QTET(0.25)		QTET(0.5)		QTET(0.75)	
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
Practice firms								
Worst case	-32086	17521	-43333	10067	-34581	18819	-27446	25954
	[-35042	19552]	[-47466	13303]	[-37736	21974]	[-29890	28420]
	[-34567	19226]	[-46801	12783]	[-37229	21467]	[-29497	28024]
Positive selection into employment	-7543	17521	-10813	10067	-6741	18819	-4646	25954
	[-9961	19552]	[-14607	13303]	[-10257	21974]	[-7584	28420]
	[-9572	19226]	[-13997	12783]	[-9692	21467]	[-7112	28024]
Uniformity	-7719	3215	-2413	7187	-4221	4659	-12206	1594
	[-11053	5647]	[-6341	11647]	[-8229	8925]	[-27080	4732]
	[-10517	5256]	[-5710	10930]	[-7585	8239]	[-24708	4232]
P.S. and uniformity (Theorem 5)	-2847	3215	-2293	7187	-2781	4659	-3566	1594
	[-5085	5647]	[-6066	11647]	[-6371	8925]	[-6518	4732]
	[-4725	5256]	[-5460	10930]	[-5794	8239]	[-6043	4228]
P.S. and uniformity (Theorem 6)	-718	3215	107	7187	-381	4659	-1166	1594
	[-2995	5647]	[-3605	11647]	[-3984	8925]	[-4123	4732]
	[-2629	5256]	[-3008	10930]	[-3405	8240]	[-3653	4233]
P.S. and uniformity (Theorem 7)	-718	3019	[107	7067	-381	4419	-1166	1234
	[-2995	5550]	[-3605	11630]	[-3984	8803]	[-4123	4396]
	[-2629	5143]	[-3008	10896]	[-3406	8099]	[-3660	3901]
Short training								
Worst case	-25846	19657	-38436	14364	-30869	22531	-23657	28663
	[-28142	21313]	[-48587	15879]	[-32824	24487]	[-25239	31785]
	[-27773	21046]	[-46955	15636]	[-32510	24172]	[-24985	31283]
Positive selection into employment	-4997	19657	-6516	14364	-3389	22531	-2297	28663
	[-6672	21313]	[-8473	15879]	[-5544	24487]	[-4334	31785]
	[-6402	21046]	[-8158	15636]	[-5197	24172]	[-4007	31283]
Uniformity	-5784	6840	1164	12924	-2309	8851	-21737	4303
	[-8218	8531]	[-908	15706]	[-4893	11764]	[-31066	6128]
	[-7827	8259]	[-575	15259]	[-4478	11296]	[-29566	5835]
P.S. and uniformity (Theorem 5)	-703	6840	1164	12924	-389	8851	-977	4303
	[-2138	8531]	[-696	15706]	[-2311	11764]	[-2849	6128]
	[-1907	8259]	[-397	15259]	[-2002	11296]	[-2548	5835]
P.S. and uniformity (Theorem 6)	<b>1863</b>	<b>6840</b>	<b>3684</b>	<b>12924</b>	<b>2611</b>	<b>8851</b>	<b>1543</b>	<b>4303</b>
	[431	8531]	[1978	15706]	[666	11764]	[-131	6128]
	[661	8259]	[2253	15259]	[979	11296]	[138	5835]
P.S. and uniformity (Theorem 7)	<b>1863</b>	<b>6794</b>	<b>3684</b>	<b>12924</b>	<b>2611</b>	<b>8731</b>	<b>1543</b>	<b>4303</b>
	[431	8603]	[1978	15742]	[666	11762]	[-131	6198]
	[661	8312]	[2253	15289]	[979	11275]	[138	5893]

Table 4 to be continued ...

Table 4 continued ...

	ATET		QTET(0.25)		QTET(0.5)		QTET(0.75)	
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
Long training								
Worst case	-23200	21508	-33874	13646	-28845	24555	-20324	29596
	[-26091	24017]	[-44901	16741]	[-30879	26588]	[-24530	35699]
	[-25626	23614]	[-43128	16244]	[-30552	26261]	[-23854	34718]
Positive selection into employment	-4243	21508	-7954	13646	-3045	24555	-2684	29596
	[-6723	24017]	[-11575	16741]	[-5577	26588]	[-7082	35699]
	[-6324	23614]	[-10993	16244]	[-5170	26261]	[-6375	34718]
Uniformity	-3323	7953	86	10886	-765	9315	-14804	6076
	[-6695	10503]	[-3495	15710]	[-4435	12605]	[-25814	10545]
	[-6153	10094]	[-2920	14934]	[-3845	12076]	[-24044	9827]
P.S. and uniformity (Theorem 5)	863	7953	566	10886	795	9315	-284	6076
	[-1538	10503]	[-2993	15710]	[-1750	12605]	[-4762	10545]
	[-1152	10094]	[-2421	14934]	[-1341	12076]	[-4043	9827]
P.S. and uniformity (Theorem 6)	<b>3389</b>	<b>7953</b>	2846	10886	<b>3795</b>	<b>9315</b>	2956	6076
	[986	10503]	[-644	15710]	[1294	12605]	[-1389	10545]
	[1372	10094]	[-83	14934]	[1696	12076]	[-718	9854]
P.S. and uniformity (Theorem 7)	<b>3389</b>	<b>7872</b>	2846	10886	<b>3795</b>	<b>9315</b>	2956	5956
	[986	10594]	[-644	15808]	[1294	12821]	[-1389	10398]
	[1372	10157]	[-83	15017]	[1696	12257]	[-722	9716]
Re-training								
Worst case	-18861	16473	-14430	11730	-30505	22895	-23210	19870
	[-21649	18293]	[-18198	14028]	[-34853	24671]	[-24570	23196]
	[-21201	18000]	[-17593	13659]	[34154	24385]	[-24351	22661]
Positive selection into employment	-1650	16473	-5070	11730	335	22895	<b>1390</b>	<b>19870</b>
	[-3460	18293]	[-8016	14028]	[-1770	24671]	[-215	23196]
	[-3169	18000]	[-7542	13659]	[-1432	24385]	[43	22661]
Uniformity	-5161	7221	-1350	10170	-505	9215	-12770	6430
	[-7814	9085]	[-4299	13377]	[-3172	12318]	[-26071	8371]
	[-7387	8786]	[-3825	12861]	[-2744	11819]	[-23933	8059]
P.S. and uniformity (Theorem 5)	783	7221	-990	10170	1655	9215	<b>1870</b>	<b>6430</b>
	[-934	9085]	[-3782	13377]	[-449	12318]	[94	8371]
	[-658	8786]	[-3333	12861]	[-111	11819]	[379	8059]
P.S. and uniformity (Theorem 6)	<b>2744</b>	<b>7221</b>	1410	10170	<b>3695</b>	<b>9215</b>	<b>3430</b>	<b>6430</b>
	[1069	9085]	[-1318	13377]	[1549	12318]	[1567	8371]
	[1339	8786]	[-880	12861]	[1894	11819]	[1866	8059]
P.S. and uniformity (Theorem 7)	<b>2744</b>	<b>7049</b>	1410	10170	<b>3695</b>	<b>8855</b>	<b>3430</b>	<b>6070</b>
	[1069	9051]	[-1318	13453]	[1549	12119]	[1567	8143]
	[1339	8729]	[-880	12925]	[1894	11594]	[1866	7810]

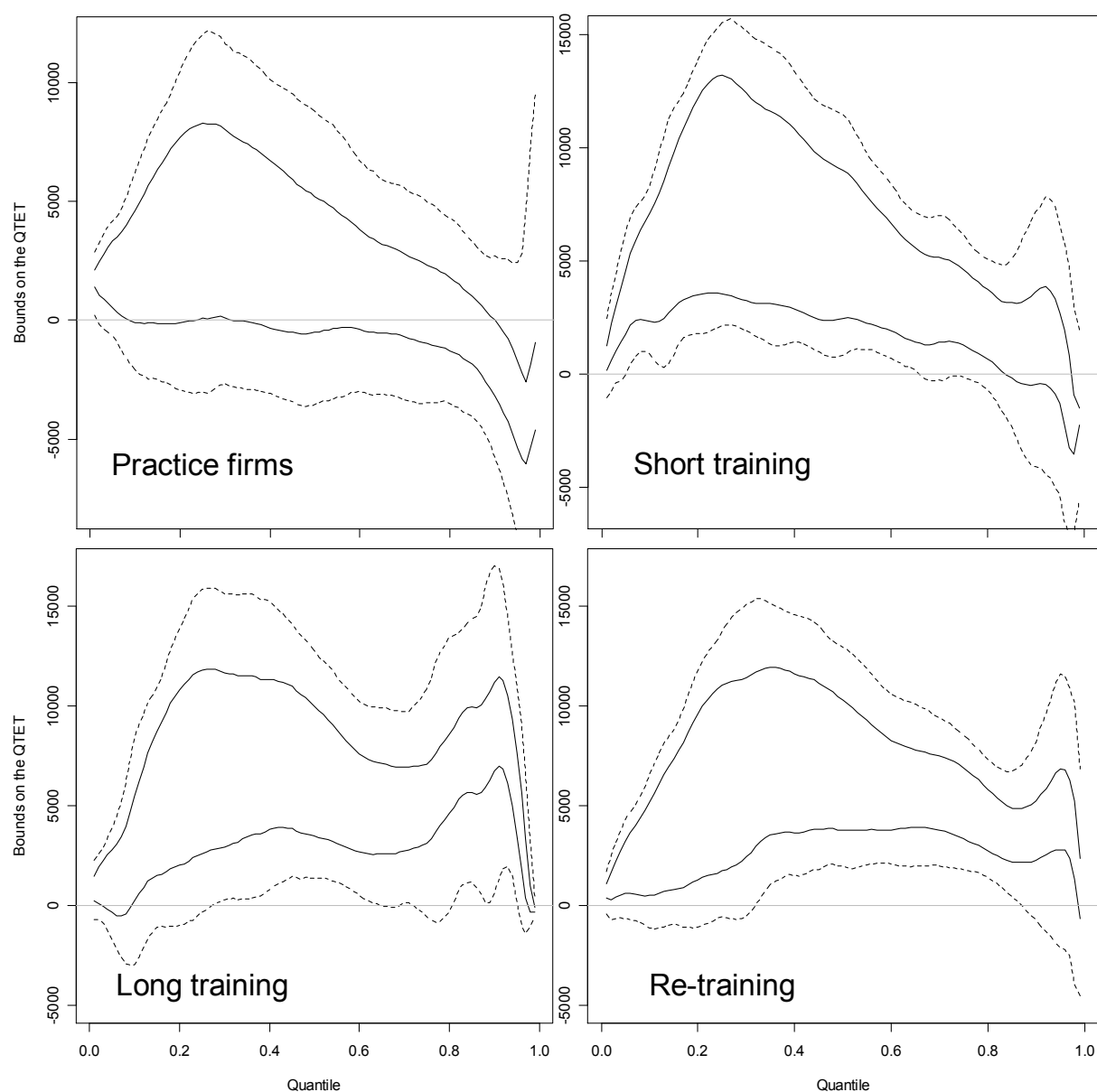
Note: The second line in each cell gives a 95%-confidence interval for the bounds estimated by bootstrapping the results 200 times. The third line in each cell gives a 95%-confidence interval for the treatment effects obtained by the method of Imbens and Manski (2004) and by bootstrapping the results 200 times. P.S. means positive selection into employment. Intervals in bold significantly exclude a zero effect at the 5% level. See also note below Table 3.

For practice firms, none of the bounds can significantly exclude zero effects. For all other programs, significant positive effects are found for the average effects and most quantile effects. The magnitudes of the effects are not small compared to median observed earnings of about 19'000 Euros. For instance, the lower bound on the median effect for re-training and long training is about 3'700 Euros, which is almost 20% of the median observed earnings. The average effects are somewhat smaller but still sizeable. All these results indicate that participating in one of the three training programs significantly increases the earnings capacity of the participants.

While Table 4 shows the results for three selected quantiles only, Figures 3 and 4 give a more complete picture of the quantile treatment effects by considering 99 percentiles. Figure 3 presents the bounds for our preferred combination of assumptions (Theorem 6) along with confidence intervals covering the true parameters with 95% probability (Imbens and Manski, 2004). To get an impression about the identifying power of the other assumptions that have potential power to exclude zero effects in this application, Figure 4 shows the bounds resulting from the combination of the uniformity assumption with the three types of positive selection. To avoid overloading this figure we do not present confidence intervals.

Figure 3 shows that short training has a significantly positive effect on earnings capacity from the 10<sup>th</sup> to the 65<sup>th</sup> percentile, long training from the 30<sup>th</sup> to the 60<sup>th</sup> percentile and re-training from the 35<sup>th</sup> to the 85<sup>th</sup> percentile, while no such effects appear for practice firms. It seems that we cannot reject the null hypothesis that the effects are the same for all quantiles. Note however that these are effects in absolute value (in Euros) and the same absolute effect represents a much higher relative effect for low quantiles than for high quantiles. A statistically valid test of this hypothesis requires developing new test procedures based on the entire partially identified quantile process. This is outside the scope of this paper.

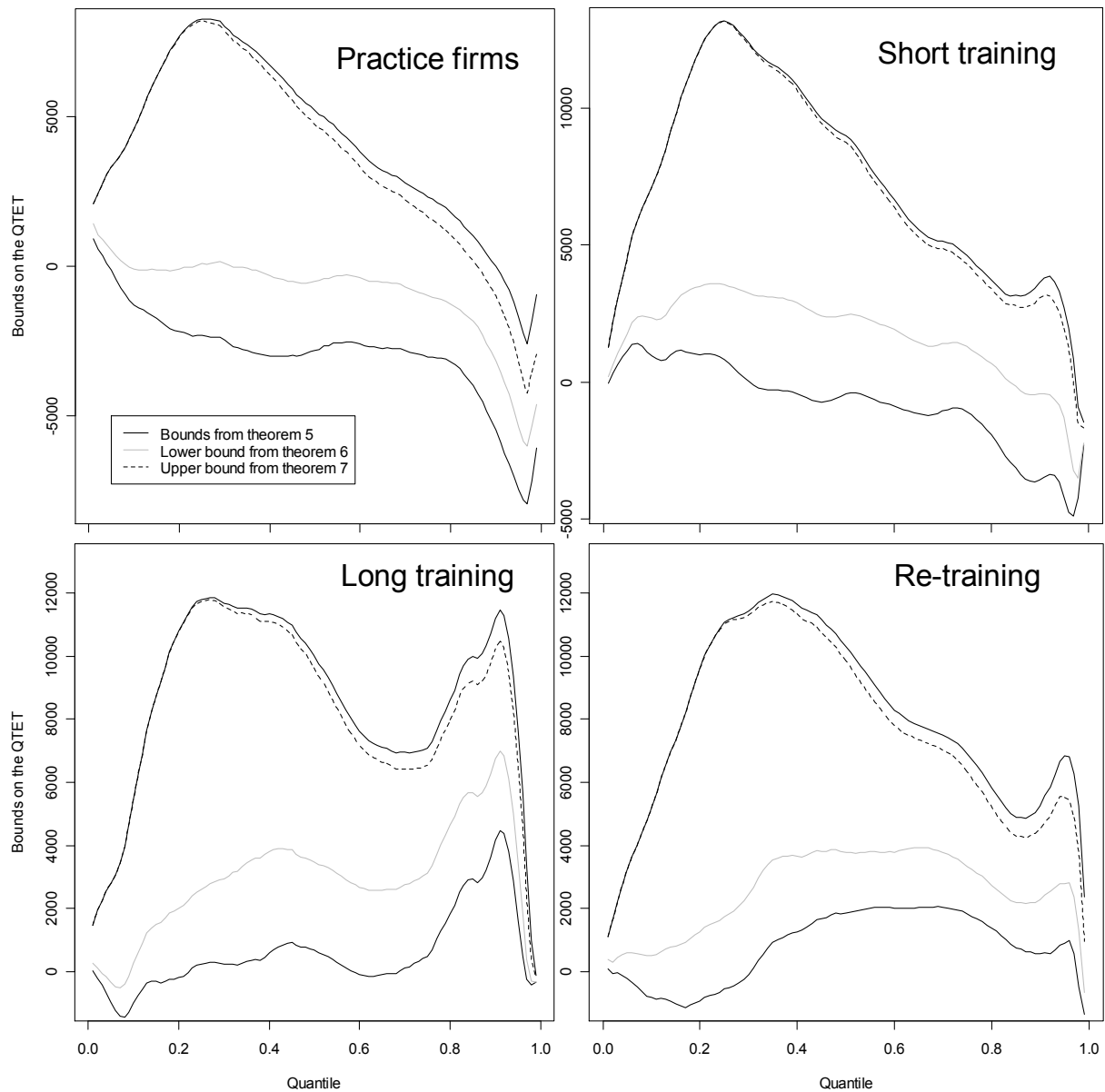
Figure 3: Bounds on the QTETs resulting from Theorem 6



Note: "Bagged" results obtained by taking the mean of the estimates over 200 bootstrap replications. Solid lines give the estimated bounds, dashed lines give the 95% confidence intervals, grey lines give the 0 line. The 95% confidence intervals are obtained by implementing Imbens and Manski (2004) results with a bootstrap based on 200 replications.

Figure 4 compares the power of the different combinations of positive selection with uniformity. It shows that the lower bound implied by Theorem 5 allows rejecting the absence of any effects only for re-training. The more informative lower bounds of Theorem 6 are needed to obtain significantly positive effects for short and long training.

Figure 4: Bounds on the QTETs for three combinations of uniformity and positive selection



Note: "Bagged" results obtained by taking the mean of the estimates over 200 bootstrap replications.

## 7 Conclusion

Using our preferred combination of assumptions, we find substantial increases in the earnings capacity for three of the four groups of German training programs we consider. Although the assumptions used to obtain these results are rather weak and do not allow point identification of the effect, the effects are large enough and the assumptions powerful enough to reject the hypotheses that the average program effects and most quantile program effects on potential

earnings are zero. This adds further evidence to previous findings suggesting that the West German training programs as run in the years after unification are a success (e.g., Fitzenberger, Osikominu, and Völter, 2007, and Lechner, Miquel, and Wunsch, 2005) in sharp contrast to the programs run a decade later (see Wunsch and Lechner, 2007). From a methodological point of view, these results indicate that our bounding strategy is not only credible because it makes weak assumptions, but the strategy can be very informative for policy makers as well.

The methods introduced in this paper are specific neither to problem of selection into employment nor to program evaluation. A first new potential application for the bounds we have obtained is the problem of sample attrition in panel econometrics. Evaluating the effects of a drug on an outcome different from the survival probability, quality of life for instance, represents a second example. Death will truncate the observed distribution of the life quality indicator. Moreover, we cannot assume that death and life quality are independent. A third example is given by the effects of an educational program on exam grades. Some of the students will probably not write the exam and these are probably less good than the students taking the exam.

## **Literature**

Angrist, J. D. (1997): "Conditional independence in sample selection models", *Economics Letters*, 54, 103-112.

Angrist, J. D., E. Bettinger, and M. Kremer (2006): "Long-Term Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia", *American Economic Review*, 96, 847-862.

Angrist, J. D., E. Bettinger, E. Bloom, E. King, and M. Kremer (2002): "Vouchers for Private Schooling in Colombia: Evidence from Randomized Natural Experiments", *American Economic Review*, 92, 1535-1558.



- Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition Using Bounds", *Econometrica*, 75, 323–363.
- Card, D., C. Michalopoulos, and P. K. Robins (2001): "Measuring Wage Growth Among Former Welfare Recipients", NBER Working Paper, 8444.
- Chaudhuri, P. (1991): "Nonparametric Estimates of Regression Quantiles and their Local Bahadur Representation", *Annals of Statistics*, 19, 760-777.
- Das, M., W. K. Newey, and F. Vella (2003): "Nonparametric Estimation of Sample Selection Models", *Review of Economic Studies*, 70, 33-58.
- Drake, C. (1993): "Effects of Misspecification of the Propensity Score on Estimators of Treatment Effect", *Biometrics*, 49, 1231-1326.
- Fan, J., N. Heckman, and M. Wand (1995): "Local polynomial kernel regression for generalized linear models and quasi-likelihood functions", *Journal of the American Statistical Association*, 90, 141-150.
- Fitzenberger, B., A. Osikominu, and R. Völter (2007): "Get Training or Wait? Long-Run Employment Effects of Training Programs for the Unemployed in West Germany," Department of Economics, Goethe-University, Frankfurt am Main.
- Firpo, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects", *Econometrica*, 75, 259-276.
- Friedlander, D., D. Greenberg and P. Robins (1997): "Evaluating Government Training Programs for the Economically Disadvantaged," *Journal of Economic Literature*, 35, 1809-1855.
- Frölich, M. (2006): "Non-parametric regression for binary dependent variables", *Econometrics Journal*, 9, 511-540.

- Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects", *Econometrica*, 66, 315-331.
- Ham, J. C., and R. J. LaLonde (1996): "The Effects of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training", *Econometrica*, 64, 175-205.
- Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153-162.
- Heckman, J. J., and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation", *Econometrica*, 73, 669-738.
- Heckman, J. J., H. Ichimura, and P. Todd (1998): "Matching as an Econometric Evaluation Estimator", *The Review of Economic Studies*, 65, 261-294.
- Heckman, J. J., J. Smith and N. Clements (1997): "Making the most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts", *The Review of Economic Studies*, 64, 487-535.
- Heckman, J. J., R. LaLonde, and J. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs", in: O. Ashenfelter and D. Card (eds.), *Handbook of Labour Economics*, Vol. 3, 1865-2097, Amsterdam: North-Holland.
- Hirano, K., G. W. Imbens and G. Ridder (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score", *Econometrica*, 71, 1161-1189.
- Horowitz, J. L. (2001): "The Bootstrap", in J. J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, Vol. 5, chapter 52, 3159-3228.
- Imbens, G. W. (2000): "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, 87, 706-710.
- Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: a Review", *Review of Economic and Statistics*, 86, 4-29.

- Imbens, G. W. and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, 62, 467-475.
- Imbens, G. W., and C. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845-1857.
- Imbens, G. W., W. Newey, and G. Ridder (2005): "Mean-square-error Calculations for Average Treatment Effects," IEPR working paper 05.34.
- Kluge, J. (2006): "The Effectiveness of European Active Labour Market Policy", IZA Discussion Paper 2018.
- Knight, K. and G. W. Bassett (2002): "Second Order Improvements of Sample Quantiles Using Subsamples", University of Toronto and University of Illinois, Chicago.
- Koenker, R. (2005), *Quantile Regression*, Cambridge: Cambridge University Press.
- Koenker, R., and S. Portnoy (1987): "L-Estimation for Linear Models", *Journal of the American Statistical Association*, 82, 851-857.
- Lechner, M. (2001): "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption", in: M. Lechner and F. Pfeiffer (eds.), *Econometric Evaluation of Active Labour Market Policies*, 43-58, Heidelberg: Physica.
- Lechner, M. (2008): "A Note on Endogenous Control Variables in Evaluation Studies", *Statistics and Probability Letters*, forthcoming.
- Lechner, M., R. Miquel, and C. Wunsch (2005): "Long-Run Effects of Public Sector Sponsored Training in West Germany", Discussion paper, Department of Economics, University of St. Gallen.
- Lee, D. S. (2005): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects", NBER, Working Paper, 11721.
- Lee, T. H., and Y. Yang (2006): "Bagging binary and quantile predictors for time series", *Journal of Econometrics*, 135, 465-497.

- Manski, C. (1989): "Anatomy of the Selection Problem", *Journal of Human Resources*, 24, 343-360.
- Manski, C. (1990): "Nonparametric Bounds on Treatment Effects", *American Economic Review, Papers and Proceedings*, 80, 319-323.
- Manski, C. (1994): "The Selection Problem", in C. Sims (editor), *Advances in Econometrics*, Sixth World Congress, Cambridge: Cambridge University Press, pp. 143-170.
- Manski, C. (1997): "Monotone Treatment Response," *Econometrica*, 65, 1311-1334.
- Manski, C. (2003), "Partial Identification of Probability Distributions", New York: Springer.
- Manski, C., and J. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling", *Econometrica*, 68, 997-1010.
- Martin, J. P., and D. Grubb (2001): "What works and for whom: A review of OECD Countries' experiences with active labour market policies", *Swedish Economic Policy Review*, 8, 9-56.
- Melly, B. (2006): "Estimation of Counterfactual Distributions using Quantile Regression", mimeo.
- Roy, A. D. (1951): "Some thoughts on the distribution of earnings", *Oxford Economic Papers*, 3, 135-146.
- Rosenbaum, P. R. and D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects", *Biometrika*, 70, 41-55.
- Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D. B. (1980): "Comment on Randomization Analysis of Experimental Data: The Fisher Randomization Test by D. Basu", *Journal of the American Statistical Association*, 75, 591-593.

- Van Ours, J. (2004): "The Locking-in Effect of Subsidized Jobs", *Journal of Comparative Economics*, 32, 37-52.
- Wunsch, C. (2005): "Labour Market Policy in Germany: Institutions, Instruments and Reforms since Unification," Discussion paper, Department of Economics, University of St. Gallen.
- Wunsch, C., and M. Lechner (2007): "What Did All the Money Do? On the General Ineffectiveness of Recent West German Labour Market Programmes," Discussion paper, Department of Economics, University of St. Gallen.
- Zhang, J. L. and D. B. Rubin (2003): "Estimation of causal effects via principal stratification when some outcome are truncated by death", *Journal of Educational and Behavioral Statistics*, 28, 353-368.
- Zhang J., D. B. Rubin, and F. Mealli (2007): "Evaluating The Effects of Job Training Programs on Wages through Principal Stratification", forthcoming in D. Millimet, J. Smith, and E. Vytlacil (eds.), *Advances in Econometrics: Modelling and Evaluating Treatment Effects in Econometrics*, UK, Elsevier Science Ltd.
- Zhao, Z. (2005): "Sensitivity of Propensity Score Methods to the Specifications", IZA discussion paper 1873.