ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Lechner, Michael; Melly, Blaise

Working Paper

Partial identification of wage effects of training programs

Working Paper, No. 2010-8

Provided in Cooperation with:
Department of Economics, Brown University

Suggested Citation: Lechner, Michael; Melly, Blaise (2010) : Partial identification of wage effects of training programs, Working Paper, No. 2010-8, Brown University, Department of Economics, Providence, RI

This Version is available at:
https://hdl.handle.net/10419/62614

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

PARTIAL IDENTIFICATION OF WAGE EFFECTS OF TRAINING PROGRAMS

Michael Lechner
Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen

Blaise Melly
Department of Economics, Brown University

First version: July, 2007
This version: April, 2010
Date this version has been printed: 22 April 2010

Abstract: In an evaluation of a job-training program, the influence of the program on the individual wages is important, because it reflects the program effect on human capital. Estimating these effects is complicated because we observe wages only for employed individuals, and employment is itself an outcome of the program. Only usually implausible assumptions allow identifying these treatment effects. Therefore, we suggest weaker and more credible assumptions that bound various average and quantile effects. For these bounds, consistent, nonparametric estimators are proposed. In a reevaluation of a German training program, we find that a considerable improvement of the long-run potential wages of its participants.

Keywords: Bounds, treatment effects, causal effects, program evaluation

JEL classification: C21, C31, J30, J68

Addresses for correspondence

Michael Lechner, Swiss Institute for Empirical Economic Research (SEW), University of St. Gallen, Varnbüelstrasse 14, CH-9000 St. Gallen, Switzerland; Michael.Lechner@unisg.ch; www.sew.unisg.ch/lechner.

Blaise Melly, Department of Economics, Box B, Brown University, Providence, RI 02912, USA; Blaise_Melly@brown.edu.

1 Introduction*

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 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). This literature has studied the effects of the various programs on employment and *realized* earnings usually by setting earnings of the non-employed to zero. On the other hand, this literature was not able to measure the treatment effects on human capital because realized earnings are the product of the individual earnings potential times the probability of employment. Therefore, labor demand and labor supply influence them and we cannot distinguish how much of the effects is due to human capital changes.

However, the aim of many countries when running these programs is not to bring the program participants into (potentially bad) jobs quickly but to increase long-term employability. Training programs should increase the human capital of the participants substantially and therefore help unemployed finding stable and qualified jobs.[1] For this reason, the effect of the

* The first author has further affiliations with ZEW, Mannheim, CEPR, London, IZA, Bonn, PSI, London, IAB, Nuremberg, and Ifo, Munich. 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, Alan Manning, seminar participants at Boston University and MIT and two anonymous referees for helpful comments on a previous version of the paper.

[1] In the case of Germany, § 1 of the Work Promotion Act (Arbeitsförderungsgesetz) explicitly states that the programs should improve the human capital stock of the individuals and counter low-quality employment.

programs on potential wages is a highly policy relevant parameter and measures the success of these programs. Furthermore, in the absence of long-term employment outcomes over 10 to 15 post-program years, changes in human capital are probably the best predictor for long-term employability in these countries.

Evaluating the effect of a training program on the potential wage is, however, a complicated econometric problem because of the selective observability of wages. Participants in training programs are typically low skilled unemployed with 'bad' employment histories and low re-employment rates. Therefore, if we are interested in the wage effects of such programs we have to deal with the fact that many participants as well as comparable non-participants will not receive any wage 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.

A convenient, but generally incorrect, approach to estimate the potential wage effects is to compare the earnings for employed participants and employed non-participants. An alternative popular strategy is to use classical sample selection models (Heckman, 1979). Unfortunately, the identification of such models either requires a distributional assumption or relies on a continuous instrument that determines the employment status but does not affect wages. Finding such a variable, however, is usually very difficult. It is impossible in our application.

Therefore, we follow another strategy: we bound average and quantile program effects on potential wages. 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. The suggested bounds do not depend on the way the selection problem related 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 retraining programs in West Germany.[2] Such programs are an important (and expensive) tool of the German labor market policy. Recent German administrative databases are very informative and allow us to credibly control for selective participation into programs, to capture important aspects of the effect heterogeneity, and to follow the effects of training over a longer period.

The methodological part of this paper builds on the existing literature on partial identification. Manski (1989, 1990, 1994, and 2003) and Robins (1989) contribute prominently to this approach consisting in bounding the effects of interest using only weak assumptions. Horowitz and Manski (2000) bound treatment effects with missing covariate and outcome data. Blundell, Ichimura, Gosling, and Meghir (2007) introduce a restriction imposing positive selection into work, while Lee (2009) 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 assumptions. Angrist, Bettinger, and Kremer (2006) use a similar combination of assumptions to bound the effects of school vouchers on test scores.[3]

This paper contributes to the existing literature in four ways. First, we bound not only average but also quantile treatment effects. The effects of a treatment on the distribution of the outcome are of interest in many areas of empirical research. Policy-makers 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

[2] We concentrate on West Germany only, because East Germany faces unique transition problems.

[3] They assume directly that the effects of the treatment on the potential test-taking status and on the potential score are positive.

mean. 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 second contribution of this paper is to allow for a more general first step selection process. 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. Our theoretical results only require that the first step selection problem is solved for some subpopulation. They do not depend on the specific method used to solve this problem. In our application, we assume unconfoundedness, i.e., we assume that the treatment status is independent of the outcome variables conditionally on a set of covariates.

Third, in our application we bound the treatment effects for the observable population consisting of the employed participants.[4] Most of the existing literature bounds the effects for the unobserved population of individuals working irrespectively of their program participation status. However, results for an unobserved population are less intuitive and more difficult to communicate. For example, simple descriptive statistics cannot characterize such a population. Furthermore, from an economic point of view, the population of working participants is clearly the most interesting population because only they have realized potential wage effects that actually affect their consumption possibilities.

Finally, we apply our results to a policy relevant question. Using our preferred combination of assumptions, we find substantial increases in the earnings capacity for the training program we consider. In fact, both average treatment effects and most quantile treatment effects significantly exclude a zero potential wage effect. This shows that our bounding strategy is

[4] More generally, our theoretical results apply to the employed part of the population for which the first step selection is solved. This population may (e.g. for selection on observables) or may not (e.g. for instrumental variables) be observed but its size is known and it can always be described by summary statistics.

not only credible because it makes weak assumptions, but that it can be 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. An appendix contains the main proofs of the theorems. Further proofs are relegated to an appendix that can be downloaded from the web pages of the authors at *www.sew.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.[5] Table 1 displays the expenditures for active and passive labor market policies and especially for training programs for the unemployed in West Germany for the years 1991-2003. As usual, training has the objective of updating and increasing the human capital. It is the most utilized instrument and represents almost 50% of the total expenditure devoted to the active labor market policy.

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.[6] In our

[5] See Wunsch (2005) for a detailed account of the German labor market policy.

[6] For a recent (classical) evaluation study of the German training policy, see Lechner, Miquel, and Wunsch (2005). This paper contains also more extensive description of the institutional environment and many details on the data.

empirical application, we concentrate on so-called *Re-training* courses. They are substantive investments supposed to enable unemployed of working in a different profession than the one currently held by awarding new vocational degrees. Their mean full-time duration is about 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 re-training 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 non-participants and 4000 participants in re-training courses.[7]

[7] 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.

Our outcome variables are annual employment and earnings during the seventh year after program start. It is a weakness of this database that there is no information on hours available. Therefore, we cannot construct wages but have to stick to annual (or monthly) earnings. Looking at the effects seven years after program start allows us to concentrate on the long-run effects, which are more interesting policy parameters than the short-term effects, because the former are closer to the permanent effects of the program. Particularly for this rather long program, 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 training participation and employment (employed / non-employed) status. This illustrates the 'double selection problem' for the estimation of program effects on wages.

Concerning selection into the programs (compare the two columns *total*), the results can be summarized as follows: Participants in re-training are younger compared to non-participants, 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 re-training is only about half the share of foreigners in the group of non-participants. Participants in re-training are less educated and less skilled. Nevertheless, past earnings are somewhat higher for participants in re-training than for non-participants.

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' human capital effects of these training programs.

3 Notation, definitions, and effects

3.1 The standard model of potential outcomes

We observe *N* random draws, indexed by *i*, from a large population. The binary variable D_i indicates participation of individual *i* in the training program. Individual characteristics are captured by X_i which is defined over the support χ. We follow the standard approach in the microeconometric literature to use potential outcomes to define causal effects of interest. Rubin (1974), among others, popularized this approach. Let $Y_i(1)$ be the earnings individual *i* would earn after program participation and let $Y_i(0)$ be the earnings individual *i* would obtain otherwise. Similarly, we define the potential employment statuses $S_i(1)$ and $S_i(0)$, where $S=1$ for employed individuals and 0 otherwise.

We assume that the four potential outcomes are well-defined even if we do not observe all of them in practice. Assuming the validity of the stable-unit-treatment-value assumption (see Rubin, 1980) allows us to relate the different potential outcomes to the observable outcomes:

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

$$Y = DY(1) + (1-D)Y(0) \text{ if } S=1 \text{ and } Y \text{ is missing if } S=0.$$

While the definition of *S* is standard in the treatment effect literature, the double selection problem appears in the definition of *Y*. For instance, we observe $Y_i(1)$ only if $D_i = 1$ *and* $S_i(1)=1$. This explains why most of the literature has considered only effects on gross or total earnings, $Y(d)\cdot S(d)$, that are always observed.[8]

[8] Here earnings are simply set to zero for non-working individuals. Alternatively, they could also contain 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

Following the literature, we base our analysis on causal parameters that can be deduced from the differences of the marginal distributions of potential outcomes.[9] First, consider average and quantile treatment effects on $Y \cdot S$ caused by *D*. To define the quantile effects, let $F_{V|W}(v;w)$ be the distribution function of the random variable *V* conditional on *W* evaluated at *v* and *w*. *W* may be a vector of random variables. The corresponding θ^{th} ($0 \le \theta \le 1$) quantile of $F_{V|W}(v;w)$ is denoted by $F^{-1}_{V|W}(\theta;w)$. Using this definition, we obtain the following earnings effects of participating in a program:

$$ATE^{YS}(T=t) = E\left(Y(1)S(1)\middle|T=t\right) - E\left(Y(0)S(0)\middle|T=t\right);$$

$$QTE^{YS}_{\theta}(T=t) = F^{-1}_{Y(1)S(1)|T}(\theta;t) - F^{-1}_{Y(0)S(0)|T}(\theta;t).$$

The random variable *T* defines the target population for which we define the effect. For instance, if *T* is the treatment variable *D* and $t=1$ we obtain the so-called treatment effect on the treated, whereas for $t=0$ we obtain the treatment effect on the non-treated. If we identify the treatment effects using an instrumental variable strategy, *T* can be an indicator for being a complier as defined in Imbens and Angrist (1994) and we obtain the local average or local quantile treatment effect parameter. In the framework of Heckman and Vytlacil (2005, 2007), *T* can be the error term in the treatment selection equation and we obtain a family of marginal treatment effects.

These parameters, which are defined for various outcome variables, are the usual objects of investigation in empirical evaluation studies. They are interesting in their own right and are

---

of the program on a measure of disposable income. In some applications, setting the unobserved outcome to any value may not make sense, for instance when the outcome is a test score or a measure of health.

[9] 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 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.

frequently estimated in empirical studies (e.g. Lechner, Miquel, and Wunsch, 2005). However, they fail to answer the important question whether the program lead to productivity increases. The failure of answering this important policy question comes from the fact that these earnings outcomes mix employment and pure earnings effects. Therefore, to answer questions about the potential earnings effects, we compare potential outcomes for different participation states in a (potential) world in which all individuals had found a job, which is not observable for non-working individuals. In particular, we investigate the potential earnings effects for those individuals who would find a job under the treatment:

$$ATE^{Y}\left(T=t\right)=E\left[Y\left(1\right)\middle|T=t,S\left(1\right)=1\right]-E\left[Y\left(0\right)\mid T=t,S\left(1\right)=1\right];$$

$$QTE_{\theta}^{Y}\left(T=t\right)=F_{Y(1)|T,S(1)}^{-1}\left(\theta;t,1\right)-F_{Y(0)|T,S(1)}^{-1}\left(\theta;t,1\right).$$

Note that the problem is symmetric with respect to the definition of the treatment. The technical arguments would be almost identical if we were interested in the same effects for those individuals who would find job when non-treated.

Note also that we consider treatment effects for subpopulations defined by the potential employment status $S\left(1\right)$, not by the observed *S*. The causal interpretation of effects conditional on *S* is unclear, because part of the effect of *D* 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.

In section 4, we derive the bounds for a generic population defined by $T=t$ and $S\left(1\right)=1$. In our application, we are more specific and consider the effects for the “doubly treated” population with $D=1$ and $S\left(1\right)=1$. We could also consider the treatment effects for the whole population (irrespectively 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 a narrowly targeted program.

The effects for other populations have been considered in the literature as well. 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 (2009) consider wage 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 (2009) 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 (2009) 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.

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 (partial) identification of $E\left[g\left(Y\left(d\right)\right)\middle|X=x,T=t,S\left(1\right)=1\right]$ for $d\in\left\{0,1\right\}$ and $x\in\chi_{T=t}$ to examine the identification of the average and quantile treatment effects. Letting $g\left(Y\right)=Y$, we obtain the *ATEs* defined above. Letting $g\left(Y\right)=1\left(Y\leq\tilde{y}\right)$, we identify the

distribution function of *Y* evaluated at $\tilde{y}$.[10] The distribution function can then be inverted to get the quantiles of interest and to obtain all *QTE*s defined above.

Define $\underline{b}_g \equiv \inf_y g(y)$ as lower bound of $g(\cdot)$ and $\overline{b}_g \equiv \sup_y g(y)$ 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 1. If we estimate the expected value of *Y*, $g(\cdot)$ is the identity function and $\underline{b}_Y$ and $\overline{b}_Y$ 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 $0.25(\overline{b}_Y - \underline{b}_Y)^2$.

Lemma 1 shows that we obtain sharp unconditional bounds by integrating sharp conditional bounds.[11] The proofs of all lemmas can be found in the Internet Appendix.

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

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

Naturally, if $\underline{b}_g(x) = \overline{b}_g(x)$ for $\forall x \in \chi$, then $E(g(Y))$ is identified. For instance, in the Assumption 1 defined below, we assume that $E[Y(1)|X = x, T = t, S(1) = 1]$ is point identified. Therefore, we obtain the following bounds on $ATE^Y(T = t)$:

[10] The indicator function $1(\cdot)$ equals one if its argument is true.

[11] We define *sharp* (or *tight*) bounds as bounds that cannot be improved upon without further information.

$$\underset{X|T=t,S(1)=1}{E}\left[E\left(Y(1)-\overline{b}_{Y(0)}(X)\middle|X=x,T=t,S(1)=1\right)\right]$$
$$\leq ATE^{Y}(T=t)\leq \underset{X|T=t,S(1)=1}{E}\left[E\left(Y(1)-\underline{b}_{Y(0)}(X)\middle|X=x,T=t,S(1)=1\right)\right].$$

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.

Lemma 2 (bounds on the quantile function)

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

$$\underline{b}_{QY}(\theta)\equiv\inf_{\tilde{y}}\left\{\overline{b}_{1(Y\leq\tilde{y})}\geq\theta\right\} \qquad \text{if } \lim_{\tilde{y}\to-\infty}\overline{b}_{1(Y\leq\tilde{y})}>\theta,$$
$$\equiv\underline{b}_{Y} \qquad \text{otherwise;}$$
$$\overline{b}_{QY}(\theta)\equiv\sup_{\tilde{y}}\left\{\underline{b}_{1(Y\leq\tilde{y})}\leq\theta\right\} \qquad \text{if } \lim_{\tilde{y}\to\infty}\underline{b}_{1(Y\leq\tilde{y})}<\theta,$$
$$\equiv\overline{b}_{Y} \qquad \text{otherwise.}$$

The sharp lower and upper bounds for the θth quantile of Y are $\underline{b}_{QY}(\theta)$ and $\overline{b}_{QY}(\theta)$.

The implication of Lemmas 1 and 2 is that we only need to determine tight bounds of the conditional expected value of $g(Y(0))$ for the population of interest to bound sharply the *ATE*s and *QTE*s. This is what we do in the next section.

4 Identification

4.1 First step assumptions

To concentrate on the special issues related to the 'double selection problem' into programs and employment, we assume that the treatment effects would be point identified if there was

no sample selection. We state Assumption 1 in general terms because the bounding strategy that we propose does not depend on the way this first selection problem is solved. Below, we discuss two examples that satisfy this assumption: selection on observables and instrumental variables.

Assumption 1 (Identification of conditional employment and wages given employment)

$p_{S(d)|X,T}(x,t)$, $F_{Y(d)|X,T,S(d)}(\tilde{y};x,t,1)$ and $F_{X|T}(x;t)$ are identified for $\forall d \in \{0,1\}$, for $\forall x \in \chi_{T=t}$ and $\forall \tilde{y} \in (-\infty,\infty)$.[12]

Lemma 3 illustrates the empirical content of Assumption 1. It 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 employment and earnings)

If Assumption 1 holds, then $E\left[S(1)-S(0)|T=t\right]$, $ATE^{YS}(T=t)$, and $QTE_{\theta}^{YS}(T=t)$ for $\forall \theta \in (0,1)$ are identified.

Example 1: effects on the treated, selection on observables This example corresponds to our application in section 6. Here, we assume independence of treatment, *D*, and potential outcomes, $Y(d)$, $S(d)$, conditional on confounders, *X*, as in the standard matching literature (see Section 6 for a brief justification in the specific context of that application). *T* is equal to the treated (we could as well consider the effect on the non-treated, the whole population, etc.). Furthermore, to be able to recover the necessary information from the data, common support assumptions are added in part b) of Assumption 1'. As we are interested in the effects on the double treated, combinations of characteristics that lead to a zero treatment probability

[12] When the elements of *V* are binary, the probability of all elements of *V* jointly being equal to one conditional on $W=w$ is denoted by $p_{V|W}(w)$.

do not matter. Lemma 4 verifies that this traditional matching assumption satisfies the conditions of Assumption 1.

Assumption 1' (conditional independence assumption for first stage)

a) Conditional independences: $\{Y(0)S(0), S(0)\} \perp D | X = x$ for $\forall x \in \chi_{D=1}$;[13]

b) Common support: $P(D = 1 | X = x) < 1$ for $\forall x \in \chi$.

Lemma 4: Assumption 1' implies Assumption 1 for the treated population.

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 (2009).

Example 2: binary instrumental variable, effects for the compliers In this example, we show that the assumptions of Imbens and Angrist (1994) made for both outcomes (employment and earnings) satisfy the conditions stated in Assumption 1 for the population of compliers. Let Z be a binary variable and define the potential treatment variables indexed against Z, $D_i(0)$ and $D_i(1)$. We follow the original article and call compliers the individuals who react to a change in the value of Z: $C = 1(D(1) > D(0))$.

Assumption 1'' states standard conditions as found in Abadie, Angrist, and Imbens (2002), for instance. Lemma 5 shows that these assumptions satisfy the requirements of Assumption 1 for the population of compliers (T is C, and $t = 1$)

[13] This notation means that the joint distribution of $Y(0)S(0)$ and $S(0)$ is independent of D conditional on X.

Assumption 1'' (conditional instrumental variable assumption for first stage)

a) Conditional independences: $\{Y(0)S(0), Y(1)S(1), S(0), S(1), D(0), D(1)\} \perp Z | X = x$ for $\forall x \in \chi$;

b) Monotonicity and first-stage: $\Pr(D_i(1) \geq D_i(0)) = 1$ and $E[D(1)] > E[D(0)]$;

c) Common support: $0 < P(Z = 1 | X = x) < 1$ for $\forall x \in \chi$;

Lemma 5: Assumption 1'' implies Assumption 1 for the compliers.

Therefore, the bounds derived in this paper are also useful in cases where the earnings effects are identified using an instrumental variable strategy. A recent example of such a case occurs in Engberg, Epple, Imbrogno, Sieg, and Zimmer (2009). They are interested in the effects of magnet schools on education outcomes, where a lottery determines magnet school assignment. The problem is that some households move to another school district if they lose the lottery and the outcomes are no longer observed in this case. We could bound the causal effect of magnet schools on the schooling outcomes using the strategy suggested in this paper.

These two examples do not exhaust all cases where Assumption 1 is satisfied. For instance, in the presence of a continuous instrument, Heckman and Vytlacil (2005, 2007) discuss the identification and estimation of the marginal treatment effects. Our results can be applied to bound these effects when the outcome is not observed only for a selected population.

4.2 Worst case bounds

Assumption 1 is not sufficient to identify the effects of *D* on the potential earnings, $Y(d)$. Of course, we point identify these effects if we assume that selection into employment is independent from potential earnings. Another 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). However, since assuming independent sample selection is not plausible in our application (and probably the majority of applications), and no continuous instrument is available for the second stage selection process, we give up on trying to achieve 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 employment and realized earnings reduces the uncertainty. The bounds given in Theorem 1 – as well as all other bounds derived in this paper – are *sharp* (see Internet Appendix for proofs). To state this theorem concisely, we denote the expected value of *Y* in the *p* fraction of the population with the smallest value of *Y* by $\underset{\min|p}{E}(Y)$. If *Y* is a continuous random variable, $\underset{\min|p}{E}(Y)=\int_{-\infty}^{F_Y^{-1}(p)} y\cdot\frac{f_Y(y)}{p}\cdot dy$. If *Y* is an indicator variable, $\underset{\min|p}{E}(Y)=0$ if $\Pr(Y=1)<1-p$ and $\underset{\min|p}{E}(Y)=\left(\Pr(Y=1)-1+p\right)/p$ otherwise. Similarly, $\underset{\max|p}{E}(Y)$ denotes the expected value of *Y* in the *p* fraction of the population with the largest value of *Y*.[14]

Theorem 1 (worst-case bounds)

Assumption 1 holds. If $p_{S(0)|X,T}(x,t)+p_{S(1)|X,T}(x,t)>1$, then the lower and upper bounds on $E\left[g\left(Y(0)\right)\middle|X=x,T=t,S(1)=1\right]$ are given respectively by

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

$$+\underline{b}_g\frac{1-p_{S(0)|X,T}(x,t)}{p_{S(1)|X,T}(x,t)},\quad\text{and}$$

[14] We use the notation introduced by Zhang and Rubin (2003), who bounded the effects for the unobserved population of individuals working irrespectively of their program participation status.

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

$$+\bar{b}_g\frac{1-p_{S(0)|X,T}(x,t)}{p_{S(1)|X,T}(x,t)}.$$

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

Note that the bounds are identified by Assumption 1. Theorem 1 shows that we can learn part of the nonparticipation-employment outcome of employed participants from the employment outcomes of non-participants. However, there remains uncertainty because the employed participants could be unemployed when non-treated. In addition, Assumption 1 is silent about which wage the always-employed participants would receive when non-treated. The importance of these two sources of uncertainty decreases as the employment probabilities increase.

Obviously, these bounds will be very wide if the employment probabilities are low. In our application, the employment probabilities are small because we consider a sample of persons who are unemployed when treatment starts. The employment probability at the end of our sample period is below 60%. Therefore, we cannot obtain informative bounds without further restricting the selection process into employment. Such restrictions will be imposed below.[15]

4.3 Positive selection into employment

In the standard labor supply model individuals accept a job offer if the offered wage is higher than the reservation wage, denoted by Y^R, i.e. $S(0)=1\left(Y(0)\geq Y^R\right)$. If Y^R is independent from $Y(0)$ this model trivially implies that employed individuals are positively selected from the population. This is more generally true as long as the difference between $Y(0)$ and Y^R is

[15] The presence of a discrete instrument for employment is an example of such a restriction. We do not present the implied bounds because we do not have any plausible exclusion restriction in our application. Moreover, the bounds are straightforward to derive as the intersections of the bounds conditionally on the instrument.

positively associated with $Y(0)$.[16] This relation motivates the assumption, suggested by Blundell, Gosling, Ichimura, and Meghir (2007), that $\Pr\left(S(0)=1 \middle| X=x, Y(0)\le \tilde{y}\right) \le \Pr\left(S(0)=1 \middle| X=x, Y(0)>\tilde{y}\right)$. Such a condition is equivalent to assuming that the wage distribution of the employed first-order stochastically dominates the wage distribution of non-employed:[17]

Assumption 2 (positive selection into employment)

$F_{Y(0)|X,T,S(0)}(\tilde{y};x,t,0) \ge F_{Y(0)|X,T,S(0)}(\tilde{y};x,t,1)$ for $\forall x \in \chi_{T=t}$ and $\forall \tilde{y}$.

Blundell, Gosling, Ichimura, and Meghir (2007) are interested in the conditional distribution of wages for the whole population. We are more specifically interested in the (unconditional) treatment effects for the employed treated population. In our application, their bounds are uninformative while we obtain tighter bounds because we consider a population with a higher employment probability and we use the treatment effect structure. We obtain especially informative bounds by combining positive selection with an assumption about the treatment choice defined below that cannot be used to bound outcomes for the whole population.

Assumption 2 tightens one of the bounds derived in Theorem 1:

Theorem 2 (positive selection into employment)

a) If Assumptions 1 and 2 hold, and $g(\cdot)$ is a monotone increasing function, then:

$$E\left[g(Y(0)) \middle| X=x, T=t, S(1)=1\right] \le \underset{\max \left| p_{S(1)|X,T}(x,t)\right.}{E}\left[g(Y(0)) \middle| X=x, T=t, S(0)=1\right].$$

b) If Assumptions 1 and 2 hold and $g(\cdot)$ is a monotone decreasing function, then:

$$E\left[g(Y(0)) \middle| X=x, T=t, S(1)=1\right] \ge \underset{\min \left| p_{S(1)|X,T}(x,t)\right.}{E}\left[g(Y(0)) \middle| X=x, T=t, S(0)=1\right].$$

---

[16] Y^R is of course allowed to be positively correlated with $Y(0)$.

[17] See Blundell, Gosling, Ichimura, and Meghir (2007) for a proof.

Thus, assuming positive selection allows tightening the upper bound on the expected value and the lower bound on the distribution function of the potential outcome. There is no contradiction because the distribution function must be inverted to obtain the quantile function, which means that adding positive selection has a similar effect on the quantiles and the mean.

4.4 Conditional monotonicity of the treatment effect on employment

Lee (2009) restricts the individual treatment effect on the employment probability to have the same sign for all of the population.[18] Lee's assumption is similar to the *monotonicity* assumption of Imbens and Angrist (1994), but they restrict the effect of the instrument on the treatment status, while Lee restricts the effect of the treatment on sample selection.[19]

Lee's (2009) 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 (2009) 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.

A third difference with Lee (2009) is that we bound the effect for an observable population as long as assumption 1 is satisfied for an observable population. In our application, we bound

[18] The same assumption is also made by Zhang and Rubin (2003).

[19] 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.

the effects for the employed treated while Lee bounds the effects 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*, he estimates the effects for a population that is a mixture of treated and non-treated, which is difficult to interpret.

The formal definition of *monotonicity* is given in Assumption 3:

Assumption 3 (conditional monotonicity of the treatment effect on employment)

For each $x \in \chi_{T=t}$, either a) $P\left(S(1) \geq S(0) \mid X = x, T = t\right) = 1$,

or b) $P\left(S(1) \leq S(0) \mid X = x, T = t\right) = 1$.

Theorem 3 shows that Assumption 3 allows tightening the bounds considerably:

Theorem 3 (conditional monotonicity of the treatment effect on employment)

a) Assumptions 1 and 3-a) hold. The bounds are given by the following expressions:

$$E\left[g\left(Y(0)\right)\middle|X = x, T = t, S(0) = 1\right]\frac{p_{S(0)|X,T}(x,t)}{p_{S(1)|X,T}(x,t)} + \underline{b}_g \frac{p_{S(1)|X,T}(x,t) - p_{S(0)|X,T}(x,t)}{p_{S(1)|X,T}(x,t)}$$
$$\leq E\left[g\left(Y(0)\right)\middle|X = x, T = t, S(1) = 1\right] \leq$$
$$E\left[g\left(Y(0)\right)\middle|X = x, T = t, S(0) = 1\right]\frac{p_{S(0)|X,T}(x,t)}{p_{S(1)|X,T}(x,t)} + \overline{b}_g \frac{p_{S(1)|X,T}(x,t) - p_{S(0)|X,T}(x,t)}{p_{S(1)|X,T}(x,t)}.$$

b) Assumptions 1 and 3-b) hold. The bounds are given by the following expressions:

$$\min_{\left|\frac{p_{S(1)|X,T}(x,t)}{p_{S(0)|X,T}(x,t)}\right|} E\left[g\left(Y(0)\right)\middle|X = x, T = t, S(0) = 1\right] \leq E\left[g\left(Y(0)\right)\middle|X = x, T = t, S(1) = 1\right] \leq$$
$$\leq \max_{\left|\frac{p_{S(1)|X,T}(x,t)}{p_{S(0)|X,T}(x,t)}\right|} E\left[g\left(Y(0)\right)\middle|X = x, T = t, S(0) = 1\right].$$

Interestingly, we obtain point identification if $p_{S(1)|X,T}(x,t) = p_{S(0)|X,T}(x,t)$. The reason is that under the monotonicity 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 procedures are similar in this respect because they condition on the participation probability, as discussed by Angrist (1997). 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 1 in Lee (2009) 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.

4.5 Combination of assumptions

Adding positive selection as defined in Assumption 2 to conditional monotonicity as defined in Assumption 3 tightens one of the bounds on $E\left[g\left(Y(0)\right)\middle|X=x,T=t,S(1)=1\right]$:

Theorem 4 (positive selection into employment and conditional monotonicity)

a) Assumptions 1, 2, and 3-a) hold. If $g(\cdot)$ is an monotone increasing function, then, the upper bound given in Theorem 3-a) tightens to:

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

b) Assumptions 1, 2, and 3-a) hold. If $g(\cdot)$ is a monotone decreasing function, then the lower bound given in Theorem 3-a) tightens to:

$$E\left(g\left(Y(0)\right)\middle|X=x,T=t,S(0)=1\right)\frac{p_{S(0)|X,T}(x,t)}{p_{S(1)|X,T}(x,t)}$$

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

The bounds obtained in Theorem 4 are only slightly more informative than those of Theorem 3. The reason is that the positive selection assumption compares – suppressing the dependence on $T=t$ - $F_{Y(0)|X,S(0)}(\tilde{y};x,0)$ and $F_{Y(0)|X,S(0)}(\tilde{y};x,1)$, but not $F_{Y(0)|X,S(0),S(1)}(\tilde{y};x,0,1)$ and $F_{Y(0)|X,S(0),S(1)}(\tilde{y};x,1,1)$. Therefore, it is possible that $F_{Y(0)|X,S(0),S(1)}(\tilde{y};x,0,1)$ dominates $F_{Y(0)|X,S(0),S(1)}(\tilde{y};x,1,1)$, although at the same time $F_{Y(0)|X,S(0)}(\tilde{y};x,1)$ dominates $F_{Y(0)|X,S(0)}(\tilde{y};x,0)$. We rule out this implausible scenario in Assumption 4:

Assumption 4 (positive selection into employment conditionally on $S(1)=1$)

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

Combining Assumption 4 with the conditional monotonicity assumption leads to simple and intuitive bounds that are given in Theorem 5:

Theorem 5 (positive selection into employment conditionally on $S(1)=1$ and monotonicity)

a) Assumptions 1, 3-a), and 4 hold. If $g(\cdot)$ is a monotone increasing function, then: $E\left[g\left(Y(0)\right)\middle|X=x,T=t,S(0)=1\right]\geq E\left[g\left(Y(0)\right)\middle|X=x,T=t,S(1)=1\right]$.

b) Assumptions 1, 3-a), and 4 hold. If $g(\cdot)$ is a monotone decreasing function, then: $E\left[g\left(Y(0)\right)\middle|X=x,T=t,S(0)=1\right]\leq E\left[g\left(Y(0)\right)\middle|X=x,T=t,S(1)=1\right]$.

The intuition for this result is simple. Suppose that a program has a positive effect on employment. This means that the potential wage of the employed participants is lower than that

of the employed non-participants by the positive selection assumption. If, despite this positive effect on employment, the program has also a positive effect on the observed wages, this implies that the program has a positive effect on potential wages. Formally,

$$E\left[Y(1)-Y(0)\middle|X=x,T=t,S(1)=1\right]$$
$$\geq E\left[Y(1)\middle|X=x,T=t,S(1)=1\right]-E\left[Y(0)\middle|X=x,T=t,S(0)=1\right].$$

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 but keep this part of the paper brief. We consider only selection on observables as defined by Assumption 1' and the effects on the treated because we use this strategy in our application. 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 5. 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 $g(Y)=1(Y\leq\tilde{y})$

.

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 nonli-

near estimator (Fan, Heckman, and Wand, 1995), like a local probit for instance, should be more suited for binary dependent variables.[20]

$E\left[1\left(Y \le \tilde{y}\right)\middle| X = x, D = 0, S = 1\right]$ could be estimated by a local probit as well. However, for the *QTE*s we need to estimate the conditional distribution function evaluated at a large number of $\tilde{y}$, which is computationally very intensive. Moreover, since we need to estimate the complete conditional distribution anyway, it is natural and faster to estimate the whole distribution 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 Chernozhukov, Fernández-Val and Galichon (2009).

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

The majority of the bounds for the mean contain conditional, asymmetrically trimmed means like $\underset{\max|p}{E}\left(Y\middle| X = x, D = 0, S = 1\right)$ and $\underset{\min|p}{E}\left(Y\middle| X = x, D = 0, S = 1\right)$.[21] Lee (2009) 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 θth quantile regression coefficient vector. We apply a nonparametric version of their estimator with a particular weight function, $J(\theta)$, that selects

[20] Moreover, in Frölich (2006) the local parametric estimator appears to have better small sample properties.

[21] For the distribution function, $\underset{\max|p}{E}\left(1\left(Y \le \tilde{y}\right)\right) = 1$ if $p < E\left(1\left(Y \le \tilde{y}\right)\right)$ and $\underset{\max|p}{E}\left(1\left(Y \le \tilde{y}\right)\right) = E\left(1\left(Y \le \tilde{y}\right)\right)/p$ otherwise, such that we do not need to estimate a trimmed mean. A similar result holds for the lower bound.

quantiles only above or below p. We estimate $\underset{\max|p}{E}(Y|X=x, D=0, S=1)$ by $\int_{1-p}^{1} x\hat{\beta}(x,\theta)d\theta$ and $\underset{\min|p}{E}(Y|X=x, D=0, S=1)$ by $\int_{0}^{p} x\hat{\beta}(x,\theta)d\theta$ where $\hat{\beta}(x,\theta)$ is the θth local linear quantile regression evaluated at x in the subpopulation *D=0, S=1*.

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 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 1'a) holds conditional on the (one-dimensional) propensity score as well:

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

Similarly, if Assumption 2 and 4 (positive selection into employment) are valid conditionally on *X*, they are also valid conditionally on the propensity score. In fact, these two 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 3. The monotonicity 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.

We estimate the propensity scores (for each program compared to nonparticipation) with parametric binary probits. In a second step, we estimate nonparametrically the response functions and bounds conditional on the propensity score and, when necessary, on the variables suspected to be related to employment effect heterogeneity. A potential drawback is that the bounds may be asymptotically less informative when we do not condition on all covariates.

6 Wage effects of training programs in West Germany

6.1 Validity of the identifying assumptions

As explained in Section 2, we use similar 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, position in previous job, specific occupation, and industry. Some potentially important factors are still missing such as ability, motivation, jail and detailed health histories, but we are confident to capture these factors indirectly with almost 20 years of employment histories and the other covariates.

Assumptions 2 and 4 (positive selection into employment) could be violated by a strong enough positive relationship between actual wages and reservation wages. This could particularly be the case for older unemployed with high productivity, if they have saved enough to

retire. We should not be affected by this problem because our population of interest is young (90% below 40 years old).

The monotonicity 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 5.2 with the propensity scores based on binary linear probits.[22] 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 the regressors included. 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 bounded by construction: Due to the regulations of the social security system, from which the 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. Thus, it is used as an upper bound together with zero as the lower bound.

While estimation of the bounds is a relatively standard problem, inference on partially identified parameters raises a number of issues that are the subject of a currently active literature.

[22] Lechner, Miquel, and Wunsch (2005) use a multinomial probit. We use the binomial probit because we are only interested in one particular program. Furthermore, the correlations between the estimated probabilities resulting from both estimators are higher than 98%.

Since it is outside the scope of this paper to derive the asymptotic distribution of our estimators, we will motivate the inference procedure only heuristically.

The identified sets are intervals whose boundaries have closed form expressions. Therefore, we follow an approach suggested by Horowitz and Manski (2000) and Imbens and Manski (2004) and refined by Stoye (2009). We estimate these boundaries by averages of conditional bounds, which are themselves estimated by local parametric estimators. This is very similar to the matching estimators suggested by Heckman, Ichimura, and Todd (1998) that are asymptotically normally distributed and have asymptotically linear score functions.[23] In addition, by construction, the estimated upper bound is always larger than the estimated lower bound. By Lemma 3 in Stoye (2009), this justifies using the procedure suggested by Imbens and Manski (2004). We implement their method to obtain confidence intervals that cover the true value of the parameters with a fixed probability.

We use the standard nonparametric bootstrap to estimate the joint distribution of the bounds. It is well-known that the bootstrap fails when the asymptotic distribution of the estimators is discontinuous as a function of the parameters. This is the case for our estimators only if $p_{S(1)|X,T}(x,t) = p_{S(0)|X,T}(x,t)$ for *all* $x \in \chi_{T=t}$. In our application, we exclude this possibility based on a conservative pre-test.[24] This implies that the parameters are not point identified and that the asymptotic distributions of our estimators are continuous as a function of the parameters of interest.[25]

The procedure is evaluated in a Monte Carlo study, whose data generating process is calibrated to match many characteristics of the data in our application. The bootstrap performs

[23] The estimator implementing the result of theorem 5 is exactly the estimator of Heckman, Ichimura, and Todd (1998). The other bounds are small variations of their estimator.

[24] The *p*-value of this test is 0.000003. Note that we don't have a discontinuity if $p_{S(1)|X,T}(x,t) = p_{S(0)|X,T}(x,t)$ for some *x* because we bootstrap the unconditional and not the conditional bounds.

[25] See assumption 3 in Hoderlein and Stoye (2009) for a similar assumption.

reasonably well and the confidence intervals are either slightly conservative or accurate. The detailed results are available in the Internet Appendix.

6.3 Standard earnings and employment effects

We investigate the long-run effects of the re-training program on earnings (and employment) by estimating the effects on annual earnings in the seventh year after program start.[26] Before presenting the results for the potential earnings, we show standard employment and realized earnings effects as benchmark. The upper panel of Table 3 presents the means of the outcome variables for the non-participants and the participants in the re-training program. 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 modified radius matching estimator, and because they consider monthly instead of yearly outcome variables.

The results for employment show that the program has a strongly significant positive effect on employment. This allows us to exclude point identification when we consider making inference about the bounds. The effect on total earnings is also positive, but it is impossible to know whether it is only driven by the effect on employment or whether it reflects an improved productivity. The estimated effect on earnings for the sub-samples of employed individuals is only valid if employment and earnings are independent. This assumption is probably not satisfied and this result is, therefore, difficult to interpret.

[26] The reason why we define long run as corresponding to 7 years (and not more) is driven by data availability. However, being able to observe individual outcomes for up to 7 years after program start is unusual for evaluation studies. Moreover, Lechner, Miquel, and Wunsch (2005) show that the effects were quite stable starting from the 5th year after the program start.

Figure 1 presents the quantile treatment effects on earnings, which are new.[27] They are even more difficult to interpret than the ATET. A substantial proportion of individuals are still unemployed whether they participated in a re-training program or not. Therefore, quantile treatment effects are zero for the lower part of the distribution. After that, participants are employed 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 1 also shows the effects conditionally on being employed, but they are probably biased estimates of the wage effects because of the sample selection issue that is the key topic of this paper.

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 productivity of the treated individuals. Next, we present the results that are informative on that issue.

6.4 Bounds on the potential wage effects

Table 4 shows the bounds for the *ATET* and three *QTET*s. It also reports confidence intervals obtained by the procedure described in section 6.2. They cover the true treatment effects with 95% probability. Confidence intervals for the bounds, not reported, are slightly more conservative, but lead to similar conclusions regarding the significance of the reported effects.

It is clear from the results, that the worst-case bounds are extremely wide. Imposing positive selection increases significantly the lower bounds but, with the exception of one quantile effect, is not sufficient to exclude zero effects. The monotonicity assumption allows tightening upper and lower bounds compared to the worst-case bounds, but it is not powerful enough to

[27] Since the sample objective function defining quantiles is non-differentiable, some estimates may slightly jump from one quantile to the other. Therefore, we use bagging (*b*ootstrap *agg*regat*ing*) to smooth the results by defining the estimator 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.

reject the absence of a human capital effect. However, the combination of positive selection with monotonicity leads to informative bounds. We present the results for the two different definitions of positive selection as discussed in Section 4.5. Interestingly, the bounds implied by Theorem 5 are more informative without making stronger behavioral restrictions: For them we find significant positive effects at the mean and most of the quantiles. The magnitudes of the effects are not small compared to the median observed earnings of about 19'000 Euros: the lower bound on the median effect is about 3'700 Euros, which is almost 20% of the median observed earnings. The average effect is somewhat smaller but still sizeable. These results indicate that participating in the re-training programs significantly increases the potential earnings of the participants.

While Table 4 shows the results for three selected quantiles only, Figure 2 gives a more complete picture of the quantile treatment effects by considering 99 percentiles. It presents the bounds for our preferred combination of assumptions (Theorem 5) along with confidence intervals covering the true parameters with 95% probability.

Figure 2 shows that the re-training program has positive effects at most percentiles of the potential earnings distribution. This effect is significant at 54 percentiles. It also 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.

One of the initial motivations for considering potential earnings effects was that total earnings effects mix the employment effects and the human capital effects. We bound now the respec-

tive importance of each component. Abstracting from the conditioning set X and T, by the monotonicity assumption and assuming that $\Pr(S(1)) > \Pr(S(0))$, we obtain,

$$E[Y(1)-Y(0)] = E[Y(1)-Y(0)|S(1)=1]\Pr(S(1)=1) \\ +\left(E[Y(0)|S(0)<S(1)]\right)\Pr(S(0)<S(1)).$$

The first term is the human capital effect and the second term is the employment effect. $E[Y(1)-Y(0)]$, $\Pr(S(0)=1)$ and $\Pr(S(1)=1)$ are identified by Assumption 1. Theorems 3, 4, and 5, provide us with the bounds for $E[Y(1)-Y(0)|S(1)=1]$. Therefore, we can bound the relative importance of the employment effect. A similar decomposition can be derived for the case where the program has a negative effect on employment. Applying these results to our data, at least 46% of the total earnings effects are pure human capital effects.

7 Conclusion

Using our preferred combination of assumptions, we find substantial increases in the earnings capacity due to participation in the German re-training program. 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 indeed helpful for their participants (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, 2008). 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 suggested in this paper are specific neither to problems of selection into employment nor to training program evaluation. Non-random sample selection often limits our ability to analyze the effect of a treatment. This is the case with sample attrition in panel econometrics. Evaluating the effects of a drug on an outcome observed only if the patient survives, as in Zhang and Rubin (2003), represents a second example. The evaluation of the effects of an educational program on exam grades is a third example. Some of the students will probably not write the exam and these are probably less good than the students taking the exam. Finally, Engberg, Epple, Imbrogno, Sieg, and Zimmer (2009) are interested in the effects of magnet schools on education outcomes, where a lottery determines magnet school assignment. The problem is that we do not observe the outcome if the household moves to another school district.

Literature

Abadie, A., J. Angrist and G. Imbens (2002): “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings”, *Econometrica*, 70, 91-117.

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.

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.

Chernozhukov, V., I. Fernández-Val and A. Galichon (2009): “Quantile and probability curves without crossing”, forthcoming in *Econometrica*.

Das, M., W. K. Newey, and F. Vella (2003): "Nonparametric Estimation of Sample Selection Models", *Review of Economic Studies*, 70, 33-58.

Engberg, J., D. Epple, J. Imbrogno, H. Sieg and R. Zimmer (2009): „Evaluating Education Programs That Have Lotteried Admission and Selective Attrition“, mimeo.

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.

Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects", *Econometrica*, 66, 315-331.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error", *Econometrica*, 47, 153-162.

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., 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.

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., and E. Vytlacil (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation", *Econometrica*, 73, 669-738.

Heckman, J. J., and E. Vytlacil (2007): “Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment effect to Organize Alternative Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments”, in: J. J. Heckman and E. E. Leamer (eds.), *Handbook of Econometrics, Vol. 6B*, 4875-5143.

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., and C. F. Manski (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data", *Journal of the American Statistical Association*, 95, 77-84.

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.

Kluve, 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. (2008): "A Note on Endogenous Control Variables in Evaluation Studies", *Statistics and Probability Letters*, 78, 190-195.

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, forthcoming in the *Journal of the European Economic Association*.

Lee, D. S. (2009): "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects", *Review of Economic Studies*, 76, 1071–1102.

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.

Robins, J. (1989): "The Analysis of Randomized and Non-Randomized AIDS Treatment Trials Using a New Approach to Causal Inference in Longitudinal Studies," in L. Sechrest, H. Freeman, and A. Mulley (eds), *Health Service Research Methodology: A Focus on AIDS*, Washington DC: US Public Health Service.

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.

Stoye, J. (2009): “More on Confidence Regions for Partially Identified Parameters”, *Econometrica*, 77, 1299-1315.

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 (2008): "What Did All the Money Do? On the General Ineffectiveness of Recent West German Labour Market Programmes," *Kyklos: International Review for Social Sciences*, 134-174.

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. L., 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.

Tables

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).

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

	Nonparticipation			Re-training		
	E	NE	Total	E	NE	Total
Number of observations	3211	5717	8982	254	153	407
Monthly earnings (EUR)	1561	1462	1497	1637	1519	1592
Age (years)	34	39	37	30	31	30
Women (share in %)	38	44	42	38	37	38
Nationality: German	83	81	82	90	88	91
Urbanization: Big city	25	27	26	19	22	20
Education: no degree	21	27	25	22	27	25
University degree	6	5	5	3	3	3
Salaried worker	30	28	29	25	20	23
Unskilled worker	37	41	40	51	54	53

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 the current unemployment spell.

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

	Sample averages						Program effects for participants		
Outcome	Employment		Earnings		Earnings if employed		Employment	Earnings	Earnings if employed
Population	Non-part.	Part.	Non-part.	Part.	Non-part.	Part.			
Mean / effect (std.)	0.45	0.73	8619	15920	19272	21743	**0.14** (0.03)	**4816** (849)	**2972** (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.

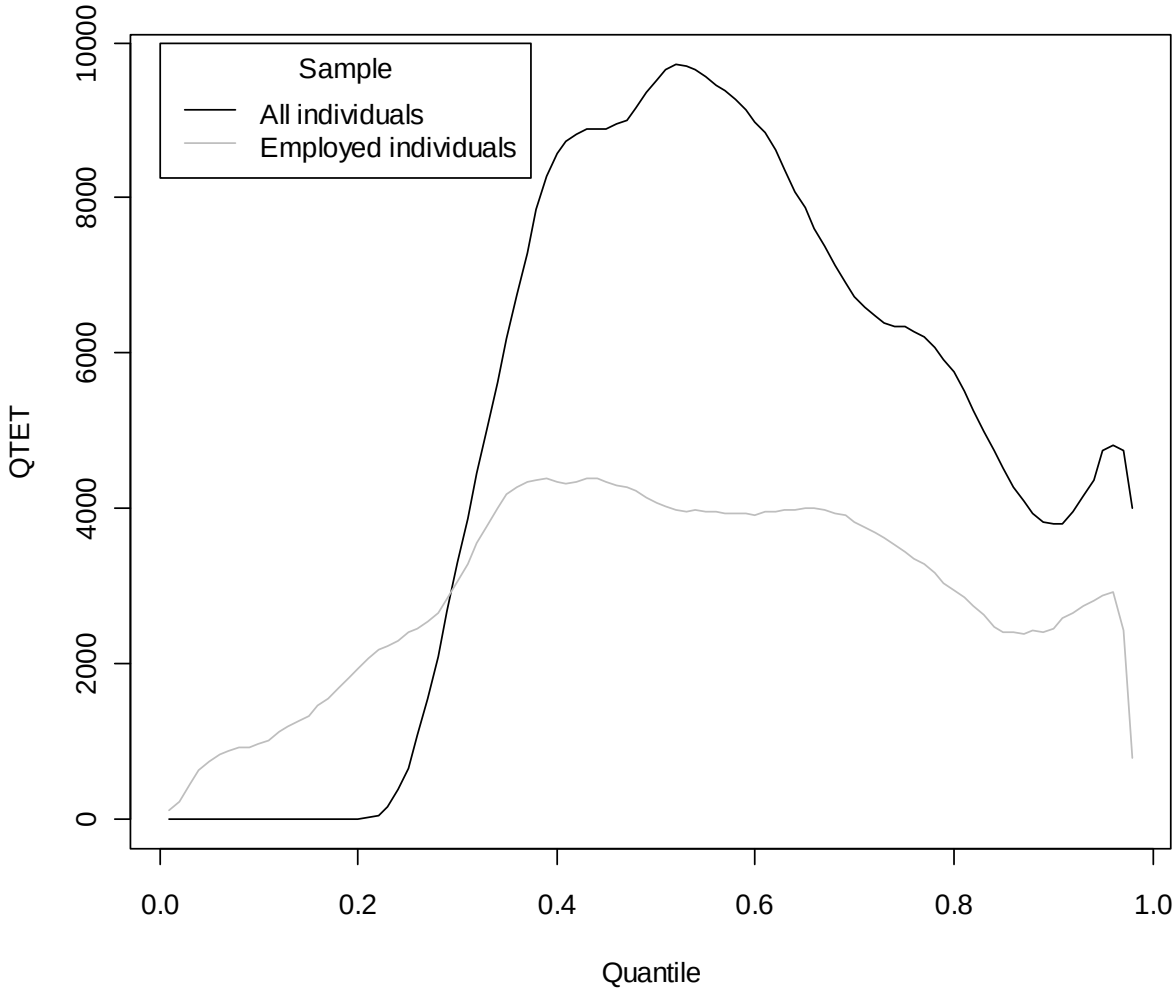
Table 4: Bounds on the ATETs and QTETs of the re-training

	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
Worst case	-18861	16473	-14430	11730	-30505	22895	-23210	19870
	[-21201	18000]	[-17593	13659]	[-34154	24385]	[-24351	22661]
Positive selection	-1650	16473	-5070	11730	335	22895	**1390**	**19870**
	[-3169	18000]	[-7542	13659]	[-1432	24385]	**[43**	**22661]**
Monotonicity	-4770	7039	-1350	9570	-265	8735	-9290	6310
	[-6792	8508]	[-3833	12266]	[-2278	11100]	[-20402	7921]
P.S. and monotonicity (Theorem 4)	785	7039	-990	9570	1655	8735	**1750**	**6310**
	[-647	8508]	[-3304	12266]	[-107	11100]	**[287**	**7920]**
P.S. and monotonicity (Theorem 5)	**2695**	**7039**	1410	9570	**3695**	**8735**	**3430**	**6310**
	[1289	**8508]**	[-863	12266]	**[1898**	**11100]**	**[1875**	**7920]**

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

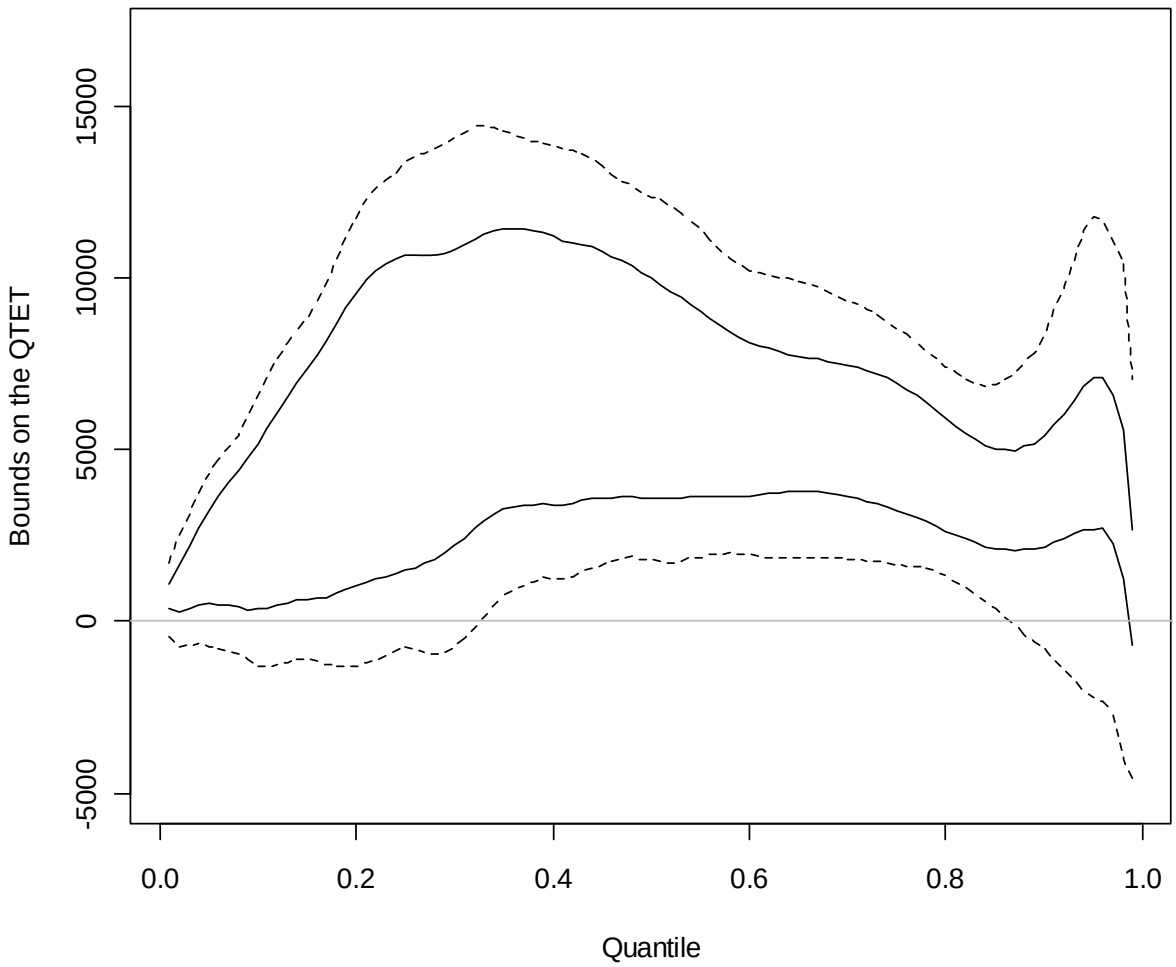
Figures

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



Note: See note below Table 3. The standard errors, not plotted to avoid overloading the figure, amount to about 1200 such that most of the QTEs on total earnings and the majority of the QTEs conditional on employment are significant.

Figure 2: 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.

Appendix: Proof of the Theorems

To simplify the notation in this appendix, we suppress the dependence on $X = x$ and $T = t$ in all expression. The proof of the sharpness of the bounds is quite repetitive and therefore relegated to the Internet Appendix.

Proof of Theorem 1

By the law of total probability, we obtain the following expression:

$$\begin{aligned} E\left[g\left(Y(0)\right) \mid S(1)=1\right] &= E\left[g\left(Y(0)\right) \mid S(1)=1, S(0)=1\right]\frac{p_{S(1),S(0)}}{p_{S(1)}} \\ &\quad + E\left[g\left(Y(0)\right) \mid S(1)=1, S(0)=0\right]\frac{p_{S(1)} - p_{S(1),S(0)}}{p_{S(1)}}. \end{aligned} \tag{1}$$

$E\left[g\left(Y(0)\right) \mid S(1)=1, S(0)=0\right]$ is unobserved and bounded only by $\underline{K}_g$ and $\overline{K}_g$ without further assumptions. Therefore, the worst case bounds are attained if $p_{S(1),S(0)}$ attains the smallest value compatible with the observed employment probabilities. The following set of equations restricts this probability:

$$\begin{aligned} &p_{S(1),S(0)} + p_{S(1),1-S(0)} = p_{S(1)} \\ &p_{S(1),S(0)} + p_{1-S(1),S(0)} = p_{S(0)} \\ &p_{S(1),S(0)} + p_{S(1),1-S(0)} + p_{1-S(1),S(0)} + p_{1-S(1),1-S(0)} = 1. \end{aligned}$$

We have three restrictions and four unknowns. Solving for the element of interest, we get $p_{S(1),S(0)} = p_{S(0)} + p_{S(1)} - 1 + p_{1-S(1),1-S(0)}$.

We cannot exclude that $p_{S(1),S(0)} = 0$ if $p_{S(0)} + p_{S(1)} \leq 1$. In this case, the observed values do not allow us to tighten the bounds on the support $\underline{K}_g$ and $\overline{K}_g$. If $p_{S(0)} + p_{S(1)} > 1$, the smallest acceptable value for $p_{S(1),S(0)}$ is given by $p_{S(0)} + p_{S(1)} - 1$ and is strictly positive. In this case, we need to bound $E\left[g\left(Y(0)\right) \mid S(1)=1, S(0)=1\right]$.

The distribution of $g\left(Y\left(0\right)\right)|S\left(0\right)=1$ is identified by Assumption 1. The population defined by $S\left(0\right)=1$ is a $\frac{p_{S(1),S(0)}}{p_{S(0)}}$ and $\frac{p_{S(0)}-p_{S(1),S(0)}}{p_{S(0)}}$ mixture of the population with $S\left(0\right)=1$, $S\left(1\right)=1$ and $S\left(0\right)=1, S\left(1\right)=0$. Since in the worst case $p_{S(1),S(0)}=p_{S(0)}+p_{S(1)}-1$, the upper bound will be attained when the population with $S\left(0\right)=1, S\left(1\right)=1$ represents the $\frac{p_{S(0)}+p_{S(1)}-1}{p_{S(0)}}$ fraction of the population with $S\left(0\right)=1$ with the largest value of $g\left(Y\left(0\right)\right)$. Similarly, the lower bound will be attained when population with $S\left(0\right)=1, S\left(1\right)=1$ represents the $\frac{p_{S(0)}+p_{S(1)}-1}{p_{S(0)}}$ fraction of the population with $S\left(0\right)=1$ with the smallest value of $g\left(Y\left(0\right)\right)$. This is the result of Theorem 1.

Proof of Theorem 2

a) If $g\left(Y\right)$ is a monotonic increasing function of *Y*, then Assumption 2 implies that the distribution of $g\left(Y\left(0\right)\right)$ given $S\left(0\right)=1$ stochastically dominates the distribution of $g\left(Y\left(0\right)\right)$ given $S\left(0\right)=0$. The upper bound is attained when this assumption is just satisfied, that is when these two distributions are the same.

The populations defined by $S\left(0\right)=1$ and $S\left(0\right)=0$ are mixtures of the two sub-populations defined by $S\left(1\right)\in\{0,1\}$. Since, in the worst case, $g\left(Y\left(0\right)\right)$ given $S\left(0\right)=0$ has the same

distribution as $g\left(Y(0)\right)$ given $S(0)=1$, the upper bound is attained when the mixture proportions are the same for $S(0)=0$ and $S(0)=1$: $\frac{p_{S(1),S(0)}}{p_{S(0)}}=\frac{p_{S(1),1-S(0)}}{p_{1-S(0)}}=p_{S(1)}$.[28]

Thus, the upper bound will be attained when the population with $S(0)=1, S(1)=1$ represents the $p_{S(1)}$ fraction of the population with $S(0)=1$ with the largest value of $g\left(Y(0)\right)$. Simultaneously, the population with $S(0)=0, S(1)=1$ represents the $p_{S(1)}$ fraction of the population $S(0)=0$ with the largest value of $g\left(Y(0)\right)$. The distribution of $g\left(Y(0)\right)$ is not observed for this last population but, by the positive selection assumption, it is bounded by the distribution of $g\left(Y(0)\right)$ for the population with $S(0)=1$. Therefore,

$$E\left[g\left(Y(0)\right)\middle|S(1)=1,S(0)=1\right]\underset{\max|p_{S(1)}}{\leq}E\left[g\left(Y(0)\right)\middle|S(0)=1\right]$$
$$E\left[g\left(Y(0)\right)\middle|S(1)=1,S(0)=0\right]\underset{\max|p_{S(1)}}{\leq}E\left[g\left(Y(0)\right)\middle|S(0)=1\right].$$

Inserting these two bounds in (1) gives the result of Theorem 2-a).

b) If $g(Y)$ is a monotonic decreasing function of *Y*, then Assumption 2 implies that the distribution of $g\left(Y(0)\right)$ given $S(0)=1$ is stochastically dominated by the distribution of $g\left(Y(0)\right)$ given $S(0)=0$. This implies that the positive selection assumption allows tightening the lower bound instead of the upper bound. The rest of the proof is along the lines of part a).

Proof of Theorem 3

Part a) $P\left(S(1)\geq S(0)\right)=1$

[28] If these mixture proportions were not the same, then it would be possible to get a higher upper bound by increasing slightly the smallest mixing proportion and decreasing slightly the highest mixing proportion.

Assumption 3a) excludes the existence of observations with $S(0)=1$ and $S(1)=0$. Therefore, $p_{S(0),1-S(1)}=0$, $p_{S(0),S(1)}=p_{S(0)}$, $p_{1-S(0),S(1)}=p_{S(1)}-p_{S(0)}$ and $p_{1-S(0),1-S(1)}=1-p_{S(1)}$.

By the law of total probability, Bayes' rule, and the implications of Assumption 3a) derived above:

$$
\begin{aligned}
&E\left[g\left(Y(0)\right)\middle|S(1)=1\right]\\
&\overset{TP}{=}E\left[g\left(Y(0)\right)\middle|S(1)=1,S(0)=1\right]p_{S(0)|S(1)}(1)+E\left[g\left(Y(0)\right)\middle|S(1)=1,S(0)=0\right]p_{1-S(0)|S(1)}(1)\\
&\overset{BR}{=}E\left[g\left(Y(0)\right)\middle|S(1)=1,S(0)=1\right]\frac{p_{S(0),S(1)}}{p_{S(1)}}+E\left[g\left(Y(0)\right)\middle|S(1)=1,S(0)=0\right]\frac{p_{1-S(0),S(1)}}{p_{S(1)}}\\
&\overset{A3a}{=}E\left[g\left(Y(0)\right)\middle|S(0)=1\right]\frac{p_{S(0)}}{p_{S(1)}}+E\left[g\left(Y(0)\right)\middle|S(1)=1,S(0)=0\right]\frac{p_{S(1)}-p_{S(0)}}{p_{S(1)}}. \qquad (2)
\end{aligned}
$$

The only unidentified element in (2) is $E\left[g\left(Y(0)\right)\middle|S(1)=1,S(0)=0\right]$ that is bounded by $\underline{b}_g$ and $\overline{b}_g$. Using those bounds, we obtain the result of Theorem 3a).

Part b) $P\left(S(1)\le S(0)\right)=1$.

Assumption 3b) excludes the existence of observations with $S(0)=0$ and $S(1)=1$. Thus, we get $p_{1-S(0),S(1)}=0$, $p_{S(0),S(1)}=p_{S(1)}$, $p_{S(0),1-S(1)}=p_{S(0)}-p_{S(1)}$, and $p_{1-S(0),1-S(1)}=1-p_{S(0)}$. These simplifications lead to the following equality: $E\left[g\left(Y(0)\right)\middle|S(1)=1\right]=$ $E\left[g\left(Y(0)\right)\middle|S(1)=1,S(0)=1\right]$.

The distribution of $g\left(Y(0)\right)|S(0)=1$ is identified by Assumption 1. The population defined by $S(0)=1$ is a mixture of the population $S(0)=1,S(1)=1$ with probability $\frac{p_{S(1)}}{p_{S(0)}}$ and of $S(0)=1,S(1)=0$ with probability $\frac{p_{S(0)}-p_{S(1)}}{p_{S(0)}}$. The upper bound will be attained when the

population $S(0)=1, S(1)=1$ represents the $\frac{p_{S(1)}}{p_{S(0)}}$ fraction of the population with $S(0)=1$ with the largest value of $g(Y(0))$. Similarly, the lower bound will be attained when the population $S(0)=1, S(1)=1$ represents the $\frac{p_{S(1)}}{p_{S(0)}}$ fraction of the population with $S(0)=1$ with the smallest value of $g(Y(0))$.

Proof of Theorem 4

Part a): The only unknown element in equation (2) implied by Assumptions 1 and 3-a) is $E\left[g(Y(0))\middle|S(1)=1, S(0)=0\right]$. By Assumption 2 and because $g(\cdot)$ is monotone increasing, the distribution of $g(Y(0))$ given $S(0)=0$ is stochastically dominated by the (identified) distribution of $g(Y(0))$ given $S(0)=1$. The upper bound is attained when these two distributions are identical. The distribution of $g(Y(0))$ given $S(0)=0$ is mixture of the sub-population $S(1)=0$ with probability $\frac{p_{1-S(1)}}{p_{1-S(0)}}$ and of the subpopulation $S(1)=1$ with probability $\frac{p_{S(1)}-p_{S(0)}}{p_{1-S(0)}}$. Therefore, the upper bound on $E\left[g(Y(0))\middle|S(1)=1, S(0)=0\right]$ is given by $\max\left|\frac{p_{S(1)}-p_{S(0)}}{p_{1-S(0)}}\right. E\left[g(Y(0))\middle|S(0)=1\right]$.

Part b) The proof is similar to part a), but with the stochastic dominance inverted.

Proof of Theorem 5

a) The only unknown element in equation (2) implied by Assumptions 1 and 3-a) is $E\left[g(Y(0))\middle|S(1)=1, S(0)=0\right]$. By Assumptions 3 and 4 and because $g(\cdot)$ is monotonic increasing,

$$E\left[g\left(Y(0)\right)\middle|S(1)=1,S(0)=0\right]\overset{A4}{\leq}E\left[g\left(Y(0)\right)\middle|S(1)=1,S(0)=1\right]\overset{A3a}{=}E\left[g\left(Y(0)\right)\middle|S(0)=1\right].(3)$$

Inserting (3) in (2) we get the result of Theorem 5-a).

b) Similar to part a).