Discussion Paper

Discussion Paper No. 96-31

Nonparametric Bounds on Employment and Income Effects of Continuous Vocational Training in East Germany

Michael Lechner

ZEW

Zentrum für Europäische Wirtschaftsforschung GmbH

Labour Economics, Human Resources and Social Policy Series

Discussion Paper No. 96-31

Nonparametric Bounds on Employment and Income Effects of Continuous Vocational Training in East Germany

Michael Lechner

Nonparametric Bounds on Employment and Income Effects of Continuous Vocational Training in East Germany

Michael Lechner*

Universität Mannheim

October 1996

Comments welcome

Address for correspondence:
Michael Lechner
Fakultät für Volkswirtschaftslehre
Universität Mannheim
D-68131 Mannheim
Germany
Email: lechner@haavelmo.vwl.uni-mannheim.de
WWW: http://www.vwl.uni-mannheim.de/lehrst/ls_oek/NBL_1

***Acknowledgement**
Financial support from the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged. I thank the ZEW and the Statistisches Bundesamt for access to the ZEW-sample of the *Mikrozensus 1993*. Furthermore, I am grateful to the ZEW, in particular to Viktor Steiner and Francois Laisney, for giving me access to their computer facilities to perform all necessary computations. I thank Klaus Kornmesser for competent help with the data and Nadine Riede for carefully reading an earlier version of this paper. All remaining errors are my own.

Abstract

This paper explores the potential of an approach suggested by Manski of obtaining nonparametric bounds for treatment effects in evaluation studies without knowledge of the participation process. The practical concern is the effects of continuous vocational training in East Germany. The empirical application is based on a large cross-section that covers about 0.6% of the total population in 1993. The results are rather mixed. The large width of the intervals obtained emphasise the fundamental problem of all evaluation studies without good knowledge of the relationship between potential outcomes and the participation process. However, in some cases suitable exclusion restrictions are indeed capable of bounding the treatment effects strictly away from zero.

Keywords:

Nonparametric estimation of treatment effects, training evaluation, East German labour markets, *Mikrozensus*

JEL classification: C14, C49, J24, J31

1 Introduction

The effects of training on the individual labour market prospects of participants or prospective participants received considerable attention in the literature.[1] The case of East Germany after unification with West Germany is particularly interesting, because of a unique situation: Massive resources are used by the public sector and to some extend also by the private sector to retrain a substantial part of the East German labour force. The intention is to enable them to adjust quickly to the rules and technologies of western-type market economies and thereby to reduce unemployment substantially. Recent evaluation studies (e.g. Fitzenberger and Prey, 1996, Lechner, 1995, 1996a, 1996b) give ambiguous answers to the question whether these policies have been beneficial to the participants or not.

There is a long discussion in the econometric literature on how identification of causal effects in such training evaluation studies could be achieved in cases when no social experiment is available. The fundamental problem is the necessity of infering from the labour market outcomes of those not participating in training enough about the outcomes of trainees if they would not have participated in training. In an ideal experimental setting both these outcomes have the same mean. The mean of the outcomes of the non-training group (control group) is an unbiased estimate of the mean of the counterfactual outcome of the training group (treatment group).[2] Other moments can be estimated similarly.

When the assignment to the treatment and control group is not random, knowledge of the assignment mechanism is necessary to adjust these estimates of the mean accordingly. In practice, for many cases complete information on this mechanism is not available. Then there are several other ways to proceed. First, there are procedures that I will summarize under the heading of nonparametric identification.[3] The basic task is to find all such attributes of the individuals that could influence the assignments as well as the potential outcomes. These attributes should fulfil the requirement that they cannot be changed by the treatment status. For each value of these attributes, the empirical procedure could proceed as if the data were generated by a true experiment. Obviously, the major untestable assumption is that all attributes are really

[1] See Heckman (1994), Lynch (1994), and LaLonde (1995) to name only a few recent papers on that topic.

[2] The state of the discussion about whether it is advantageous or not to base evaluations on social experiments can be found in Burtless (1995) and Heckman and Smith (1995).

[3] Note that this and most of the following discussion is about identification of properties in the population. Assume for simplicity that an infinite large random sample from the target population is available to perform almost exact inference, once the identification issue is resolved.

included, and that they fulfil the requirement of *exogeneity* as defined above. This approach is first explicitly suggested by Rubin (1977) and is used in a modified way by Lechner (1995, 1996a, 1996b) for the evaluation of the effects of East German training. Another nonparametric approach based on instrumental variables (IV) might be applicable when it is not possible to observe all necessary attributes in the data set. It is discussed by Imbens and Angrist (1994), and Angrist, Imbens, and Rubin (1996). The two approaches have in common that they use information about the selection process itself to identify the treatment effects without imposing other distributional or functional form assumptions. Unfortunately, for the evaluation of training in East Germany, it appears fairly difficult to find instruments fulfilling the requirements necessary to apply the IV approach.

The second group of identification strategies uses latent variable models to describe the selection process. The interaction of the selection process with the process describing the outcome is typically modelled by assumptions about the joint distribution of respective error terms. It is in the nature of the subject of identification that most of these assumptions cannot be tested. Since assumptions about error structure typically cannot be derived in an exact way from fundamentals of the assignment and outcome processes, their validity is in many cases difficult to address. Semiparametric methods try to minimise the necessity for such assumptions, but some functional form assumptions typically remain (e.g. the survey by Powell, 1994, or the application by Werwatz, 1996). Another problem of these semiparametric limited dependent variable models is that many technical conditions - particularly when the outcomes are discrete - have to be imposed that are difficult to justify by economic reasoning.[4] Models stipulating the entire parametric error distribution (typically jointly normal or extreme value) are obviously even more restrictive (see Fitzenberger and Prey, 1996, for East Germany, and Card and Sullivan, 1988, for the US). The sensitivity of the estimates regarding these assumptions is widely acknowledged in the literature (e.g. Lalonde, 1986). However, there is less agreement on how the optimal type of evaluation is supposed to look like.[5]

Even in the case of the nonparametric identification in many cases there remain some doubts, as for example in Lechner (1996b), whether these untestable assumptions are true. Therefore, in this paper I will first see how much can be said about the causal effects without any assumption about the selection process. It will turn out, that indeed something can be said, which however is rather weak: There are bounds for the true effects, but they have a considerable

[4] Many of the procedures suggested by Heckman and Hotz (1989) can also be classified to this group.

[5] Chapter 1 in Bell, Orr, Blomquist, and Cain (1995) provides a more complete account of the development of the econometric evaluation literature.

width. This is a worst case scenario, because absolutely no restriction is imposed on the selection process. Furthermore, the treatment effects are allowed to vary entirely freely among different individuals. Therefore, in a second step this paper explores various ways to tighten the bounds, either by some assumption stipulating equality of the treatment effects for subgroups of individuals or by specific assumptions about the selection process into training. The latter assumptions are not model-based, but are for example postulating that only individuals with expected positive returns are selected into training. This part of the paper builds on seminal work by Manski for selection models (Manski 1989, 1993a, 1993b, 1995), and for treatment effects (Manski, 1990, 1995). Manski, Sandefur, McLanahan, and Powers (1992) provide an application of this methodology. That being the only application of this approach I am aware of. In this paper, the econometric results are applied to the evaluation of the effects of various types of training on individual earnings and employment probabilities. The estimation is based on a very large cross-section that constitutes approximately a 0.6% random sample of the East German population in 1993.[6] The sample is sufficiently large to allow the use of non-parametric estimation methods for the estimation of the bounds.

Since this paper appears to be the first application - with the exception of Manski et al. (1992)- of this general approach, one of its goals is to explore how far this approach can carry us in obtaining useful information for evaluation studies. Overall, the results appear to be rather mixed in this repect. One obvious result is that there is a huge price to pay for ignoring selection information in terms of width of the intervals for the treatment effects. Thus the problem of all evaluation studies is emphasized: Without good knowledge about the relationship between potential outcomes and the selection / assignment process, it is very difficult to bound the treatment effects strictly away from zero. Another result is that suitable chosen exclusion restrictions are indeed capable of bounding the treatment effects away from zero. However, typically they may be at least indirectly related to the selection process, and appear difficult to justify in the context of this paper.

The paper is organized as follows: The next section introduces the identification problem and discusses the derivation of the bounds for the treatment effect under different assumptions. Section 3 presents the sample used in the empirical part. Prior to the results that are contained in Section 5, Section 4 discusses several issues related to the estimation of the bounds. Section 6 concludes. Appendix A contains technical details on the bounds under different assumptions, whereas additional results are referred to Appendices B, C, and D.

[6] ZEW sample of the *microcensus* ("Mikrozensus").

2 Identification of the causal effects of training

2.1 Causality, potential outcomes, and identification

The empirical analysis tries to answer questions like "What is the average gain of training for a certain group of individuals, such as the training participants?" For example, for a training participant the relevant comparison is with the hypothetical or counterfactual outcome of nonparticipation. Therefore, the question refers to potential outcomes or states of the world, that never occur. The underlying notion of causality requires the researcher to determine whether participation or nonparticipation in training effects the respective outcomes, such as income or employment status. This is very different from asking whether there is an empirical association, typically related to some kind of correlation, between training and the outcome. Therefore, I do not try to answer the question whether training is associated with higher earnings for example, but whether the effect of training is higher earnings.[7] Given a large enough and sufficiently informative sample from the population, the answer to the question about association can easily be answered, whereas the question about the causal relation raises serious identification issues.

The framework that will serve as guideline for the empirical analysis is the potential-outcome approach to causality suggested by Rubin (1974). Y^t and Y^c denote the outcomes (t denotes treatment, i.e. training, c denotes control, i.e. no treatment).[8] Additionally, denote variables that are unaffected by treatments - called *attributes* by Holland (1986) - by X. Attributes are exogenous in the sense that their potential values for the different treatment states coincide ($X^t = X^c$). It remains to define a binary *assignment* indicator S, that determines whether unit n gets the treatment ($S = 1$) or not ($S = 0$). If the individual participates in training the actual (observable) outcome (Y) is Y^t, and Y^c otherwise. This notation points to the fundamental problem of causal analysis. The causal effect, for example defined as the difference of the two potential outcomes, can never be estimated, even with an infinite sample, because the *counterfactual* $(y_n^t, s_n = 0)$ or $(y_n^c, s_n = 1)$ to the observable outcome (y_n) is never observed.

Using the previous notation, different estimands of interest, which are average causal effects of training for individuals with characteristic x, are denoted by $\gamma^0(x)$, $\theta^0(x)$, and $\xi^0(x)$. They are defined in equations (1), (2), and (3):

[7] See Holland (1986) and Sobel (1994) for an extensive discussion of concepts of causality in statistics, econometrics, and other fields.

[8] As a notational convention capital letters denote random variables and small letters denote specific values of these variables.

$$\gamma^0(x) := E(Y^t - Y^c | X = x) = E(Y^t | X = x) - E(Y^c | X = x)\,, \tag{1}$$

$$\theta^0(x) := E(Y^t - Y^c | X = x, S = 1) = E(Y^t | X = x, S = 1) - E(Y^c | X = x, S = 1), \tag{2}$$

$$\xi^0(x) := E(Y^t - Y^c | X = x, S = 0) = E(Y^t | X = x, S = 0) - E(Y^c | X = x, S = 0). \tag{3}$$

The short hand notation $E(\cdot | X = x, S = s)$ denotes the mean in the population of all units with characteristics x that do (s=*1*) or do not (s=*0*) participate in training. The difference between the three treatment effects is that $\gamma^0(x)$ measures the expected treatment effect for an individual randomly drawn from the part of the population with characteristics x, $\theta^0(x)$ measures that effect for an individual drawn from the population of training participants with characteristics x, and $\xi^0(x)$ measures that effect for an individual drawn from the population of nonparticipants with characteristics x.

For the following analysis it is useful to rewrite equations (1), (2), and (3):

$$\begin{aligned}\gamma^0(x) &= [E(Y^t | X = x, S = 1) - E(Y^c | X = x, S = 1)]P(S = 1 | X = x) + \\ &\quad + [E(Y^t | X = x, S = 0) - E(Y^c | X = x, S = 0)][1 - P(S = 1 | X = x)] = \\ &= [g^t(x) - E(Y^c | X = x, S = 1)]p(x) + [E(Y^t | X = x, S = 0) - g^c(x)][1 - p(x)],\end{aligned} \tag{1'}$$

$$\theta^0(x) = g^t(x) - E(Y^c | X = x, S = 1), \tag{2'}$$

$$\xi^0(x) = E(Y^c | X = x, S = 0) - g^c(x). \tag{3'}$$

The question now is how these expressions can be identified from a large random sample of the population. The quantities $g^t(x) := E(Y^t | X = x, S = 1)$, $g^c(x) := E(Y^c | X = x, S = 0)]$ and $p(x) :=\; P(S = 1 | X = x)$ are not problematic, because their sample analogues are observed. However, the sample analogues of $E(Y^c | X = x, S = 1)$, i.e. the triplet (y_n^c, x, s_n) for observations with $(s_n = 1)$, and of $E(Y^t | X = x, S = 0)$, i.e. (y_n^t, x, s_n) for observations with $(s_n = 0)$, is not observable. Another useful way to rewrite equations (1), (2), and (3) is the following:

$$\gamma^0(x) = g^t(x) - g^c(x) + \lambda^\gamma(x),$$

$$\lambda^{\gamma} = g^c(x) - g^t(x) + E(Y^t|X=x) - E(Y^c|X=x) \tag{1''}$$

$$= [g^c(x) - E(Y^c|X=x,S=1)]p(x) - [g^t(x) + E(Y^t|X=x,S=0)][1-p(x)],$$

$$\theta^0(x) = g^t(x) - g^c(x) + \lambda^{\theta}(x),$$

$$\lambda^{\theta} = g^c(x) - E(Y^c|X=x,S=1) = \frac{g^c(x) - E(Y^c|X=x)}{p(x)}, \tag{2''}$$

$$\xi^0(x) = g^t(x) - g^c(x) + \lambda^{\xi}(x),$$

$$\lambda^{\xi} = E(Y^t|X=x,S=0) - g^t(x) = \frac{E(Y^t|X=x) - g^t(x)}{1-p(x)}. \tag{3''}$$

This representation emphasizes the *selection biases* occurring when only the observable sample quantities $g^t(x)$ and $g^c(x)$ are used for estimation, instead of the correct counterfactuals $E(Y^t|X=x,S=0)$ and $E(Y^c|X=x,S=1)$, respectively.

Without additional information consistent point estimation of $\gamma^0(x)$, $\theta^0(x)$, or $\xi^0(x)$ is not possible. Much of the literature on causal models in statistics and selectivity models in econometrics is devoted to finding reasonable identifying assumptions to predict, for example, the unobserved expected nontreatment outcomes of the treated population by using the observable nontreatment outcomes of the untreated population in different ways. For example, if there is random assignment to the training given the characteristics x, then the potential outcomes are independent from the assignment mechanism and $E(Y^c|X=x,S=1) = g^c(x)$, as well as $E(Y^t|X=x,S=0) = g^t(x)$ holds. For example Lechner (1995, 1996a, 1996b) argues that such an assumption is plausible in the training context considered and proceeds with nonparametric estimations of the respective quantities. Alternative identifying assumptions may be based on modelling at least a specific part of the relation between the potential outcomes and the assignment mechanism (e.g. Heckman and Hotz, 1989, Heckman and Robb, 1985, Fitzenberger and Prey, 1996). However, a general critique of these model-based approaches is that the results could be highly sensitive to the chosen assumptions (e.g. LaLonde, 1986).[9] It should be noted that these types of identification problems are closely related to the identification problems typically encountered in selection models (e.g. Maddala, 1983).

2.2 Bounding treatment and selection effects

In many cases it is reasonable to assume that the potential outcomes have a finite support. For these cases Manski explored in a series of papers the possibility of establishing bounds on parameters of interest mainly for selection models (Manski 1989, 1993a, 1993b), but also for models of potential outcomes (Manski, 1990). Manski, Sandefur, McLanahan, and Powers (1992) provide an application of this methodology to the analysis of the effects of family structure during adolescence on high school graduation.

In the following I restate the essentials of Manki's (1990) findings in the case of treatment evaluations. Manski (1990) is based on treatment effects defined by eq. (1). His analysis is extended to the case given in eq. (2) that has often more practical relevance for evaluation studies, because $\theta^0(x)$ is the quantity that is needed to assess the effects of training programs for the actual participants. Since the considerations for the treatment effects for the nontreated population based on eq. (3) are symmetric to the case given in eq. (2), it is not discussed explicitly in the following.

The outline for the remainder of this section is the following: After introducing necessary additional notation it is shown how the treatment effects can be bounded without additional information or assumptions. Although these bounds are informative, their width is generally too large to identify the signs of the treatment effects. Therefore, it is discussed how additional information (assumptions) can be introduced to tighten the bounds. However, none of these assumptions is related to the dependence of the potential outcomes and the selection process. This dependence is kept entirely unrestricted.

Denote by $L^t(x)$, $L^c(x)$ and $U^t(x)$, $U^c(x)$ the lower and upper bounds of the support of Y^t and Y^c for a given value of $X = x$. For example, if Y^t and Y^c measure the probability that a person is unemployed or not, these bounds are naturally given as 0 and 1. It is useful to define bounds for the two subpopulations of treated and nontreated individuals as well. The lower bounds

[9] This critique is particularly damaging, because only in very rare cases are all assumption chosen because of insights into fundamental relationships between assignment, outcomes and the available data. It is far more common that assumptions - such as joint normality of some error distributions or time constancy of error components - are selected to arrive at a computationally convenient estimator.

are denoted by $\ell^t(x,s)$, $\ell^c(x,s)$, and the upper bounds are denoted $u^t(x,s)$, $u^c(x,s)$.[10] It must be true that $L^t(x)=\inf\{\ell^t(x,0),\ell^t(x,1)\}$, $L^c(x)=\inf\{\ell^c(x,0),\ell^c(x,1)\}$, as well as $U^t(x)=\sup\{u^t(x,0), u^t(x,1)\}$, $U^c(x)=\sup\{u^c(x,0),u^c(x,1)\}$. In most applications the bounds in the two subpopulations coincide, but sometimes they may differ.[11] Without any loss of generality, the treatment effects are defined to elements of the following intervals: $\gamma^0(x)\in[B_\gamma^L(x),B_\gamma^U(x)]$, $\theta^0(x)\in[B_\theta^L(x),B_\theta^U(x)]$. The width of these intervals is defined as $W_\gamma(x)=B_\gamma^U(x)-B_\gamma^L(x)$, and $W_\theta(x)=B_\theta^U(x)-B_\theta^L(x)$, respectively. It is the purpose of this section to show that these bounds are non-trivial. The trivial case is when the sample does not provide any information about the size of the treatment effects. Then the bounds are given by $B_\gamma^L(x)=L^t(x)-U^c(x)$, $B_\gamma^U(x)=U^t(x)-L^c(x)$, and by $B_\theta^L(x)=\ell^t(x,1)-u^c(x,1)$, $B_\theta^U(x)=u^t(x,1)-\ell^c(x,1)$. Hence, the width is given by $W_\gamma(x)=[U^t(x)-L^t(x)]+[U^c(x)-L^c(x)]$, and $W_\theta(x)=[u^t(x,1)-\ell^t(x,1)]+[u^c(x,1)-\ell^c(x,1)]$.

However, it is obvious that these bounds are wider than necessary. Those quantities in eq. (1′) and (2′) that do have sample counterparts can be treated as known for the purpose of the discussion of identification, because they can be consistently estimated from the sample.[12] Therefore, the following bounds are obtained for $\gamma^0(x)$: $B_\gamma^L(x)=[g^t(x)-u^c(x,1)]p(x)+[\ell^t(x,0)-g^c(x)]$

[10] To be exact, it is sufficient that $E(Y^t|X=x,S=0)$ and / or $E(Y^c|X=x,S=1)$ have finite supports. However, typically their bounds are unknown. Nevertheless, by the definition of an expectation it is always true that their bounds have to be within the interval given by the support of the respective conditional distribution functions. Hence, the bounds may contain areas for which the pdf is almost indistinguishable from zero, typically expected to be far out in the tails of a distribution (for example the right tail of the earnings distribution). If tail behaviour is sufficiently regular, these areas would however have a negligible impact on the expectation. In that sense as well, the bounds are very conservative.

[11] An example for different bounds is the following: An unemployed person may lose the right to receive future unemployment benefits if she does not participate in training offered to her by the labour office. Hence, the lower bound of her earnings would be the level of the unemployment benefits, whereas her nontraining earnings may be bounded from below by zero or some other benefit level that should usually be much lower than unemployment benefits (e.g. social assistance in the German case). Note however that this describes a difference between the bounds for the two different potential outcomes. Different bounds for the subgroup of participants can occur if they are always chosen in a specific way, i.e. if the potential outcome is very much related to the selection process. This however requires knowledge of the selection process.

[12] The width of the bounds may increase to a certain extent when sampling uncertainty is accounted for.

$[1-p(x)]$, $B_\gamma^U(x)=[g^t(x)-\ell^c(x,1)]p(x)+[u^t(x,0)-g^c(x)][1-p(x)]$. The now reduced width is given by $W_\gamma(x)=[u^c(x,1)-\ell^c(x,1)]p(x)+[u^t(x,0)-\ell^t(x,0)]$ $[1-p(x)]$. The respective bounds for $\theta^0(x)$ are given by $B_\theta^L(x)=g^t(x)-u^c(x,1)$ and $B_\theta^U(x)=g^t(x)-\ell^c(x,1)$. The width simplifies to $W_\theta(x)=[u^c(x,1)-\ell^c(x,1)]$. To appreciate the reduction of the widths assume that the different lower, respectively upper bounds of the support of the outcome variables are the same. In this case the width of the interval for both average causal effects is reduced by 50%. In the previous example with 0/1 outcomes, the width in the no data case is 2 (interval [-1,1]), now it is 1. This reduction comes without any assumptions about the selection process. Note however, that at least in the case of equal support for all outcomes, it is not possible to sign the treatment effects, because 0 is always included in the interval. The bounds for two causal effects differ mainly in the sense that to bound $\gamma^0(x)$ information about the treated ($g^t(x)$) and the untreated population ($g^c(x)$) as well as the conditional treatment probabilities $p(x)$ is needed. The bounds for $\theta^0(x)$ are simpler to compute, because they depend solely on $g^t(x)$. No information about the controls or the participation probabilities is required. In other words, such information is not informative without additional assumptions.

2.3 Shrinking the intervals

To tighten the bounds Manski (1990) suggested to use the availability of the characteristics X to introduce additional assumptions that could be plausible in some circumstances. These so-called *level-set restrictions* stipulate that either the treatment effect or some expectations of the potential outcomes are constant in a subspace (χ^0) of the total space of characteristics ($\chi, \chi^0 \subseteq \chi$).[13]

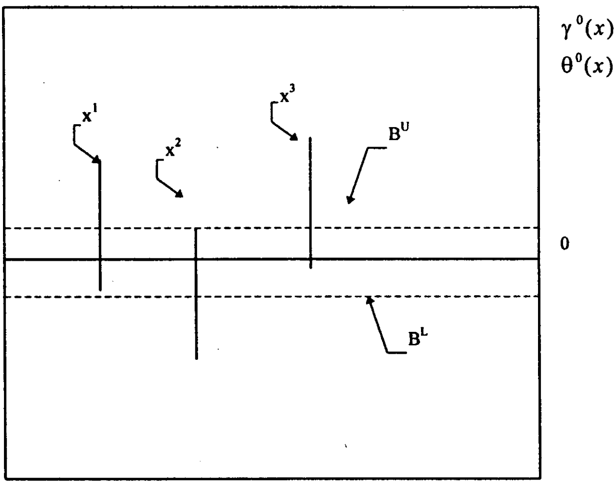
Let us be begin with a level-set restriction on the treatment effect.[14] Figure 1 shows graphically how a reduction of the widths is achieved. For illustrative purposes it is assumed that x is one-dimensional and that the treatment effect is identical for three different values of x. Clearly, the treatment effect must be included in all three intervals. The larger the space χ^0 and the more variable

[13] There is an issue of what should be the primitives for imposing assumptions. It is correctly observed by Manski (1989) that statisticians tend to introduce assumptions for moments of the outcome distributions conditional on selection ($S=s$). Then they derive the properties of the unconditional moments. Econometricians tend to make assumptions directly about moments of the unconditional distribution typically by the use of latent variable models. In this case the moments of the conditional distributions are the derived quantities.

[14] The level-set restrictions used in this section should be more precisely called local exclusion restrictions.

the bounds are within the space, the larger is the gain in width reduction. However, at least for the case of equal bounds on the outcomes, the reduction of width will be insufficient to exclude zero treatment effects. These level-set restrictions are untestable. The exact formulas for the bounds and the widths can be found in Tables A.1 and A.2 of Appendix A. Popular examples of such an assumption are models with an additive treatment effect, such as: $E(Y|X=x)= E(Y|X=x,S=s)= f(x)+\alpha s$. α denotes the treatment effect assumed to be constant in the population. More general models that allow α to vary with x - for example by including interaction terms of s and x - are also included as special cases in this discussion of level-set restrictions.[15]

Figure 1: Level-set restrictions on the treatment effects



Note: $\{x^1,x^2,x^3\} \subseteq \chi^0$.

A tighter assumption is to assume that the expected values of one particularly chosen potential outcome or both potential outcomes are constant in some regions of the X-space ($\chi^{0,t},\chi^{0,c} \subseteq \chi$). Let us first consider the case for $\gamma^0(x)$ as discussed in Manski (1990). Assume that $E(Y^t|X=x)$ is constant in $\chi^{0,t}$ and $E(Y^c|X=x)$ is constant in $\chi^{0,c}$. Then the reasoning from the section of the constant treatment effects implies that the interval on each expected potential outcome within the sets is the intersection of the intervals for each value of $x \in \chi^{0,t}$, respectively $x \in \chi^{0,c}$. As before, the exact formulas for the bounds and

[15] Note that such an exclusion restriction is one of the necessary conditions for consistent IV estimation of local average causal effects (i.e. Imbens and Angrist, 1994).

the widths are given in Table A.1. Under this assumption the treatment effect is constant on $\chi^0, \chi^0 = \chi^{0,t} \cap \chi^{0,c}$. For this restriction to bite the sets $\chi^{0,t}$ and $\chi^{0,c}$ must overlap. For this set the bounds are at least as tight as those derived under the assumption of a locally constant treatment effect before. Therefore, there is an increased likelihood to identify the sign of the effect. In the case of bounding $\theta^0(x)$ it is sufficient to assume that $E(Y^c | X = x)$ is constant in $\chi^{0,c}$ ($=\chi^0$), because $\theta^0(x)$ does not depend on $E(Y^t | X = x, S = 0)$. An alternative assumption is provided by assuming that $E(Y^c | X = x, S = 1)$ is constant in $\chi^{0,c}$. Using the reasoning applied above this implies that its upper bound is $\inf_{x \in \chi^{0,c}} u^c(x,1)$ and its lower bound is $\sup_{x \in \chi^{0,c}} \ell^c(x,1)$. This is used to derive the respective quantities given in Table A.2. This set of assumptions does not generally include the previous ones, hence there is no guarantee that the bounds will actually shrink. Quite to the contrary, if the bounds do not vary with x, then the bounds are identical to the bounds without any assumptions. In the same manner as for $\theta^0(x)$ these bounds can be derived for $\gamma^0(x)$ as well. The results given in Table A.1 are based on the case when a symmetric condition is imposed on $E(Y^t | X = x, S = 0)$ as well. The intermediate case of a level-set restriction of only one of these expectations is however straightforward to deduce from the results given.

Another restriction that is explored in the application is the assumption that all individuals selected in training have non-negative expected values for $\theta^0(x)$. Clearly, this tightens the lower bounds of the selection effects. At first sight it seems that a minimum condition of training participation should be that at least the expected returns should be positive for participants. Note however that this implies that before the training the amount of information available is considerable. It requires also that no systematic surprises (such as a suddenly appearing stigma for training) occur that might invalidate the previous considerations. An additional restriction applied to the data is that all selection effects may vary with x, but they are the same for participants and non-participants in a given subset of the X-space. Note that both sets of restrictions do not generally exclude a zero treatment effect. Finally, note that in an obvious way all these restrictions can be imposed simultaneously.

Other restrictions proposed by Manski do not appear to be attractive in this context. However for the sake of completeness, some of them are briefly discussed in Appendix A. Table A.3 in this appendix contains the respective expressions for the selection effects obtained under the assumptions discussed above.

3 Data

3.1 General considerations

Before the actual data base used in the empirical part of the paper is introduced, some general remarks about the type of data required are in order. In a typical training evaluation, samples that are as informative as possible about variables that explain the selection process into training are desirable. Typically, they contain panel data that allow to control for the important pre-training labour market history. Since constructing such data bases is costly, the number of variables is usually inversely related to the number of sample units in the sample. In other words, such samples reduce the bias but the price to pay for the econometrician is in terms of sampling variance. Usually, this is a reasonable price to pay, to avoid wrong conclusions from the analysis. Here, the situation is different. First, note that having more variables explaining selection does not generally shrink the bounds without additional assumptions, such as local exclusion restrictions. Second, as will be explained below, nonparametric methods should be used to estimate the bounds. It is now common knowledge that nonparametric methods require a substantial amount of data to perform acceptably. Therefore, an almost ideal data source for this exercise is the largest cross-section available in Germany, the *microcensus*.

3.2 The microcensus

The *microcensus* is an important component of the official German statistics system. The federal statistical office (*Statistisches Bundesamt*) collects the data of the *microcensus* by interviewing about 1% of the German population each year. The data collection is regulated by federal law. For most of the questions answering is compulsory. Hence the *microcensus* is a mixture of official (register) data and survey data with partial nonresponse. It contains information on socio-demographics, as well as on variables related to the individual labour market situation. Most of the information referes to the week of the interview, but there are also some retrospective questions. Although there is an overlap in the populations interviewed each year, the single cross-sections cannot be related to each other to form a panel.

Due to data confidentiality reasons originating in German federal laws, the original sample is not available to researchers outside the *Statistisches Bundesamt*. The ZEW however obtained an anonymized 70% random sample of the original data from the *Statistische Bundesamt*. The empirical part of this study is based on that particular sample. For the year 1993 it constitutes a 0.626% random sample of the population in the Federal Republic of Germany. See Statistisches Bundesamt (1994) for the questionnaire used and Pfeiffer and Brade (1995) for another empirical work with the ZEW-file of the *microcensus*.

When this study was done, only the 1991 and 1993 surveys are available at the ZEW. The information about training is based on retrospective questions on whether the individual participated in continuous vocational training in the last two years. The 1991 survey also covers training that started before unification. Since the goal of this paper is to investigate the effects of post-unification training, only the 1993 survey is used.

The empirical analysis is based on prime-age individuals not yet subject to early retirements. Therefore, the sample is restricted to individuals between 20 and 54 years old. Since the training for newly immigrated foreigners (as is the major share of foreigners in East Germany) has fairly different objectives than training for the rest of the labour force, the sample is restricted to individuals with German nationality living in East Germany including East-Berlin, but excluding West-Berlin. For obvious reasons valid information on training as well as on schooling is required.[16]

The information about continuous vocational training in the *microcensus* is different from the information contained in many other surveys such as the German socio-economic panel (GSOEP). The retrospective question about the incidence of training during the last two years differentiates training only by the duration (4 categories) and the location where training takes place.[17] Here, the second distinction is used to define three broad categories of training: On-the-job-training (ONJ), off-the-job-training in school-type institutions for training and retraining (OFJ), and other off-the-job-training (OFJO). Only completed training is considered. Observations still participating in some sort of training are deleted. ONJ is close to employer-related training that is analyzed with GSOEP data for example in Lechner (1996b). Unfortunately it is not known whether the individuals obtained benefits from the labour office during training. Hence the distinction between general off-the-job training (Lechner, 1995) and public-sector sponsored training (Lechner, 1996a) is not possible with this sample. However, it might be reasonable to conjecture that most of the public-sector sponsored training is indeed included in OFJ, because it is usually done in these school-like training institutions.

Table 1 gives more details on the definitions and variables used in the empirical analysis. The descriptive statistics for the three training groups and the group of individuals without training in the last two years suggest substantial differences between the groups. First, the duration of ONJ (median about 1 month) is

---

[16] It is not compulsory to answer these questions.

[17] Question 1993 survey: "Seit 1991: Haben Sie eine berufliche Fortbildung, Umschulung oder sonstige zusätzliche praktische Berufsausbildung erhalten? **Ja,** ⊗ am Arbeitsplatz; ⊗ bei der Industrie- und Handelskammer, usw.; ⊗ in besonderen Fortbildungs- und Umschulungsstätten; ⊗ an einer berufsbildenden Schule / Hochschule; ⊗ durch Fernunterricht; ⊗ auf andere Art. ⊗ **Nein.**"

considerably smaller than for OFJ and OFJO (medians about 6 months). The larger proportion of females in OFJ and the fact that OFJ participants experience higher unemployment than the rest, also points to a close relationship between OFJ and public-sector-sponsored training. Not surprisingly, the post-training unemployment rate for ONJ is rather low.[18] Note also that the usual pattern regarding education can be observed: The share of the highest schooling degree is substantially higher for training participants than for non-participants. This is also reflected in the distribution of earnings.

Table 1: Descriptive statistics

	no training	on-the-job	off-the-job	other off-the-job
	control	*ONJ*	*OFJ*	*OFJO*
Variable		mean (std) or share in %		
*Duration of training *)*				
less than 1 month	-	47	19	22
less than 6 month	-	81	50	47
less than 1 year	-	92	74	65
less than 2 years	-	98	94	84
Age	37.9 (10.0)	37.4 (9.4)	36.6 (9.2)	35.6 (9.2)
Gender: female	48	48	62	53
Federal states (Länder)				
Berlin (East)	8	13	11	10
Brandenburg	17	15	17	16
Mecklenburg-Vorpommern	11	11	12	14
Sachsen	31	30	27	25
Sachsen-Anhalt	17	17	17	17
Thüringen	16	14	16	17
Years of schooling (highest degree)				
12	13	24	23	31
10	63	65	66	62
8 or no degree	23	11	11	7

Table 1 to be continued...

[18] These empirical facts have also been found in Lechner (1996a) and Lechner (1996b) for public-sector-sponsored training and employer-related training.

Table 1 : Descriptive statistics : continued

	no training	on-the-job	off-the-job	other off-the-job
	control	*ONJ*	*OFJ*	*OFJO*
Variable	mean (std) or share in %			
Unemployed	14.8	5.4	25.6	12.9
*Net monthly income **)*				
0	1.8	0.1	1.0	0.5
less than 300	2.9	0.3	1.8	1.4
less than 600	8.3	1.8	6.2	4.9
less than 1000	23.1	8.0	26.5	17.9
less than 1400	41.6	19.3	44.6	32.8
less than 1800	64.1	41.9	62.0	50.3
less than 2200	81.9	67.7	78.9	68.5
less than 2500	90.4	82.2	88.4	80.5
less than 3000	95.5	91.6	94.6	89.3
less than 3500	97.8	95.6	97.1	93.7
less than 4000	98.7	97.7	98.3	96.2
less than 4500	99.11	98.9	98.9	97.3
less than 5000	99.46	99.37	99.49	98.2
less than 5500	99.62	99.73	99.72	98.5
less than 6000	99.74	99.87	99.79	99.14
less than 6500	99.81	99.87	99.88	99.49
less than 7000	99.83	99.97	99.95	99.49
less than 7500	99.89	100.00	99.98	99.54
Observations (total: 40793)	31477 (77%)	3003 (7%)	4341 (11%)	1972 (5%)

Note: *No training*: no continuous vocational training in last two years; *on-the-job*: vocational training at the workplace or within the firm; *off-the-job*: off-the-job continuous vocational training in learning institutions ("Fortbildungs- und Umschulungsstätten"); *other off-the-job*: other off-the-job training in a chamber of industry and commerce ("Industrie- und Handelskammer", 38%), a vocational school or university (19%), as distance teaching ("Fernunterricht", 10%), or other (33%). *) Data contain some zero values due to reporting errors. **) Includes benefits, and income from sources other than employment (e.g. returns from holding assets or other economic activities). Individuals who are still in training during the interview are deleted from the sample.

4 Estimation of the bounds on the causal effects of training

4.1 Empirical operationalisation

This section contains several issues concerning the empirical application that need to be addressed before the estimation method and the results are presented.

The first choices are to be made regarding the targets of the evaluations. Following the training literature I chose unemployment[19] and net income to measure the effects of training. The only information on these variables refers to the date of the interview. Hence, the distance to the end of training is different for different individuals. Nevertheless, it is a valid measure for the expected effect of completed pre-interview training on the outcome for the week before the interview takes place. The unemployment variable is coded as 0/1. There are special features of the income variable in the *microcensus:* First, it is measured in 18 categories (see Table 1) with an open upper bound. However, for the treatment effects one is interested in the expectation of the underlying continuous variable. Although in this case consistent point estimates are only possible with stringent distributional assumptions that are ad-hoc approximations, the additional uncertainty due to discrete observability can be naturally incorporated in the framework used here. Instead of estimating $g^t(x)$ and $g^c(x)$ one can still estimate lower and upper bounds of these quantities and use them appropriately to compute the bounds for the treatment effects. Thus, these bounds do not only indicate the uncertainty about the selection process, but also the uncertainty due to the limited information on the income variable. The income variable is unbounded (the upper is open). However, as has been mentioned before what is needed is not an upper bound for the support of the random variable, but of its expectation. The assumption that is used in the following application is that the respective expectations are no larger than DM 8000. This appears to be a very conservative number, given that only about 0.1% of the sample is observed with an income of DM 7500 or more. Second, income as defined by the *microcensus* is different from earnings, because it also contains all sorts of unearned income, such as unemployment and other benefits, income from financial assets, and so on. Treatment effects on income instead of earnings is still a valid and sensible concept. However, in case the other income components are not influenced by training participation, the differences in potential incomes are equal to the differences in the potential earnings. Another problem is the lack of information about the income of assisting family members of self-employed persons and about all self-employed individuals in agriculture. Hence, these observations are deleted from the

[19] The recoded variable *not unemployed* (1-unemployment) is used in the actual evaluations.

sample. Additionally, individuals not responding to that question are deleted as well.

Other issues concern the selection of conditioning variables to be included in X. Clearly, the definition of a conditional treatment effect only makes sense if the observation of one or the other potential outcome does not change the realisation x, that is there are no *potential characteristics* for the same individual.[20] Time constant variables and variables prior to the beginning of the selection process should always fulfil this criterion. The latter are not available in this section, hence we concentrate on age, schooling, sex, and federal state (*Bundesland*). The treatment effects are a priori expected to vary for different groups defined by these variables.

Finally, there is the issue of imposing restrictions to shrink the bounds. It does not appear to be justified by economic theories or other considerations to impose level-set restrictions on expectations of the outcome variables or the treatment effects on sets composed of different federal states, schooling or gender. Hence, level-set restrictions using variations within age groups of width of five years are explored (conditional on other components of X). They are termed *rolling (or moving) level-set restrictions* (RLS) in the following, because they are computed for each group separately.[21] RLSs are imposed on the treatment effects as well as on $E(Y^c|age)$. The underlying idea is that the exact age should not matter as long as two individuals belong to the same narrowly defined age group (± 2 years of age). Additionally, the effects of selecting only individuals with expected nonnegative gains from the treatment as well as the effects of the restriction that the treatment effects are the same for the treated and nontreated groups are explored. Finally, the different restrictions are combined to additionally tighten the bounds.

4.2 Estimation method used

Estimation of the bounds is done in a stepwise procedure. In the first step the conditional expectations $E(Y^t|X=x,S=1)$ $[=g^t(x)]$, $E(Y^c|X=x,S=0)$ $[=g^c(x)]$, and $E(S|X=x)$ $[=p(x)]$ are estimated using their sample analogues, i.e. by their weighted sample means (with weights provided in the *microcensus*) within the cells defined by the different values of x. This is a feasible approach, because the sample is sufficiently large so that there are 'enough' observations within the single cells. There are two alternatives to this simple approach: One the one hand, one could use a parametric model (as in

[20] This is a strict form of exogeneity (in regression language).

[21] This is logically inconsistent, because such an intersection or overlap of χ^0-regions - in a rigorous sense - implies that the level-set restrictions are valid for all ages. Since this is obviously too restrictive, the chosen approach provides a flexible alternative.

standard linear or binary choice regression) to estimate these conditional expectations. This however is not attractive, because it is somewhat contrary to fundamental ideas of estimating treatment effects without incurring inconsistencies by incorrect assumptions of some model. Hence, nonparametric estimation of the bounds appears to be imperative. On the other hand, some smoothing method (kernels, nearest neighbour, series estimation, ...) could be used in principle to compute the bounds to improve efficiency. This is however not attractive because of the discreteness of all X-variables.

Although the sample means are asymptotically normally distributed, to my knowledge there is no distribution theory for the quantities given in Tables A.1, A.2, and A.3. Therefore, I follow the approach by Manski et al. (1992) and implement a simple bootstrap procedure: (i) draw 500 bootstrap samples from the *microcensus;* (ii) compute for each bootstrap sample all relevant bounds; (iii) for the lower bounds report the 5% quantile of the bootstrap distribution of the estimates; for the upper bounds report the 95% quantile.[22] Note that this is still an approximation that appears very conservative because due to sample uncertainty the bounds appear to be wider than they could possibly be. It is left for future work to improve this procedure, that is to take into account that estimates of the upper and lower bounds are correlated.

5 Results

As the results may differ for different treatment and selection effects evaluated in different cells of the X-space, and there does not appear to be a suitable summary measure such as coefficients in parametric models, some choices have to be made to avoid getting lost in a bulk of results. The main body of the paper focuses on a particular case, additional interesting results are referred to three appendices. Results not presented in the paper are available on request from the author.

The first choice is on which treatment effects should be focused on. Obviously, three different treatment effects for four types of treatment status (that is 18 effects) are too much for presentation. Here, I concentrate on pairwise comparisons of the four different groups conditional on being in one or the other group or conditional on being in either of the two groups under con-

[22] Note that the procedure in Manski et al. (1992) is slightly more elaborate. They try to estimate the relevant joint distributions and conditional distribution of the random variables, and then draw the bootstrap observations from these estimates. Among other problems, this procedure is not attractive here, because of the larger X-space used at least for some estimates. See also Hall (1994) for the properties of bootstrap estimators.

sideration.[23] It appears most interesting to compare on-the-job and off-the-job training with the case of no training at all, as well as to compare the effects of on-the-job versus off-the-job training directly. For the first two comparisons the treatment effects for the population of respective training participants appears to be most interesting: this is the appropriate measure of the effectiveness of different types of training. However, in a direct comparison the training effect unconditional on training status appears to be the most relevant concept: it gives the expected effect for an individual randomly drawn from the population that participates in either type of training. Again for reasons of space, selection effects are not presented. In the main body of the paper results are only presented for the case of on-the-job training versus no training. Most of the results are for men only. The corresponding results for women are in Appendix D. The results for off-the-job training versus no training are found in Appendix B, and the results for the direct comparison of the two types of training are contained in Appendix C. Admittedly, the focus on on-the-job training for men is rather arbitrary. However, focusing more on on-the-job training is also motivated by previous results of the author. There it appeared that the effects of public-sector-sponsored training (Lechner, 1996a) as well as off-the job training (Lechner, 1995) are identified, whereas the identification of on-the-job (employer-related) training appeared to be more doubtful (Lechner, 1996b).

Another issue relates to the choice of the conditioning set. Using all available cross-products is prohibited by considerations of space and in some case results in cell sizes that are too small for reliable estimation of the bounds. Hence, in the paper I concentrate on effects jointly conditional on gender and one other factor. The latter is either schooling, region, or age.

The following tables and figures show approximate 90% centered confidence intervals. The bounds of these intervals are computed by using the 5% quantile of the bootstrap distribution of the lower bound of the repective causal effect and the 95% quantile of the distribution of the upper bound.[24]

5.1 Identified quantities ($p(x)$, $g^t(x)$, $g^c(x)$)

Table 1 contains the estimates of the participation probabilities [$p(x)$] of on-the-job training versus no training at all conditional on the level of schooling and federal states ("Länder") for men and women separately. Let us first consider the heterogeneity in the data. Heterogeneity is significant with respect

[23] Conditioning on being in the third or fourth group is not interesting, because this is like the no data situation.

[24] 95% bootstrap intervals are computed as well, but merely 500 bootstrap replications are probably not enough to get precise estimates of the 2.5% and 97.5% quantiles needed. These results are also available from the author on request.

to schooling.[25] The higher the level of schooling, the higher is the participation rate. Regarding the states there is a pronounced difference between East-Berlin and the remaining states with East-Berliners having a higher participation probability. The participation probabilities do not appear to differ systematically with age. Figure 1 shows a similar analysis for age (in years). Despite a somewhat lower rate for the youngest age group, systematic differences are hard to detect. Similarly, no significant differences for men compared to women conditional on age or schooling or states can be observed. The latter however is not true when off-the-job instead of on-the-job training is considered (see Appendix B). Here, participation probabilities for women are significantly higher than for men. This is compatible with a higher share of unemployed women, who thus have easier access to public-sector sponsored training programs. The same regional differences also show uü when on-the-job and off-the-job training are compared directly (Appendix C).

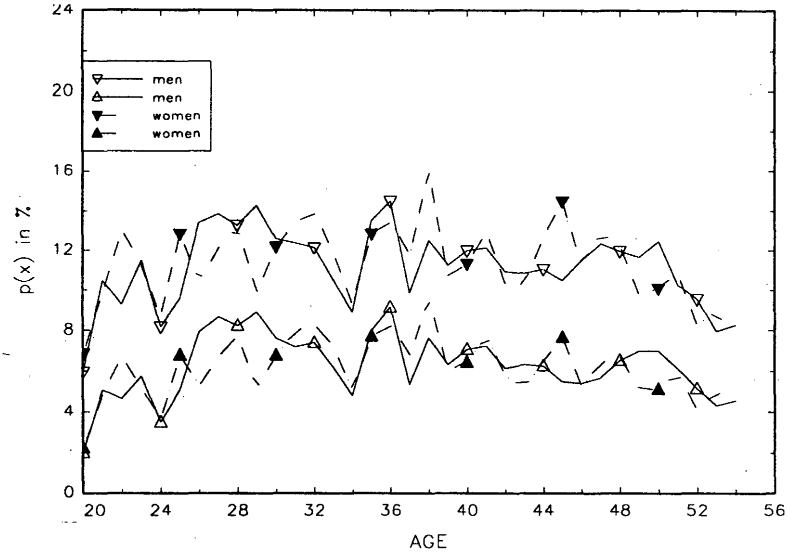
Table 2: Probabilities for on-the-job training versus no training in %

X-variables	Men		Women	
Years of schooling (highest degree)				
12	12.9	15.5	13.6	16.5
10	8.3	9.4	8.5	9.6
8 or no degree	4.2	5.5	3.3	4.7
Federal states (Länder)				
Berlin (East)	11.0	14.4	12.6	16.6
Brandenburg	6.6	8.5	6.5	8.4
Mecklenburg-Vorpommern	7.7	10.1	7.7	10.3
Sachsen	7.8	9.3	7.5	8.9
Sachsen-Anhalt	7.9	10.0	8.1	10.1
Thüringen	6.8	9.0	6.1	8.3

Note: Table shows 5% and 95% quantile of respective bootstrap distributions.

[25] This feature is already observed for example by Lechner (1996b) for another data set.

Figure 1: Probabilities for on-the-job training versus no training in % conditional on age



Note: Figure shows 5% and 95% quantile of respective bootstrap distributions.

Table 3 and Figures 2 and 3 give the confidence intervals for the means of $g^t(x)$ and $g^c(x)$. The upper part of Table 3 as well as Figure 2 focus on the indicator variable *not being unemployed.* These probabilities increase with schooling, but show only limited regional variation. They are somewhat larger for the treated group than for the control group. The latter is also true for the estimates conditional on age. The spikes in the estimate of $g^t(x)$ suggest that the number of bootstrap replications (or the sample size) is not large enough to trace out the distribution of the mean of an indicator variable with a true mean very close to 1.

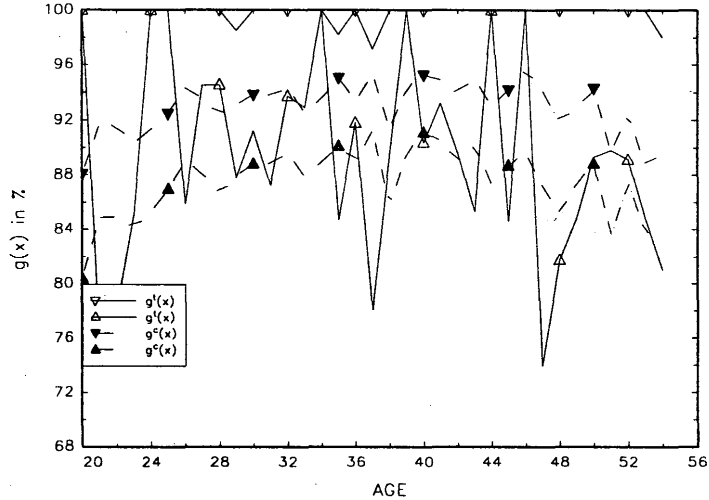
Table 3: Estimates of $g^t(x)$ and $g^c(x)$ conditional on schooling and federal state for on-the-job training versus no training

	Probabilities of not being unemployed in %				Income in DM			
X-variables	$g^t(x)$		$g^c(x)$		$g^t(x)$		$g^c(x)$	
Years of schooling (highest degree)								
12	95.5	98.8	94.1	95.9	2282	2919	2155	2683
10	94.1	96.7	91.0	92.0	1760	2251	1593	2015
8 or no degree	88.4	96.0	83.5	85.8	1501	2076	1353	1787
Federal states (Länder)								
Berlin (East)	93.6	98.9	90.3	93.3	2217	2890	1946	2473
Brandenburg	92.3	97.6	89.0	91.2	1812	2420	1568	2025
Mecklenburg-Vorpommern	90.0	97.4	84.5	87.6	1653	2294	1509	1979
Sachsen	94.4	97.9	90.9	92.4	1841	2387	1597	2045
Sachsen-Anhalt	90.9	96.9	88.8	91.1	1670	2260	1544	1998
Thüringen	94.3	99.1	89.5	91.8	1804	2423	1585	2046

Note: Table shows 5% and 95% quantile of respective bootstrap distributions. Most of the width of the income variable is due to its grouped character (see Table 1). Men only.

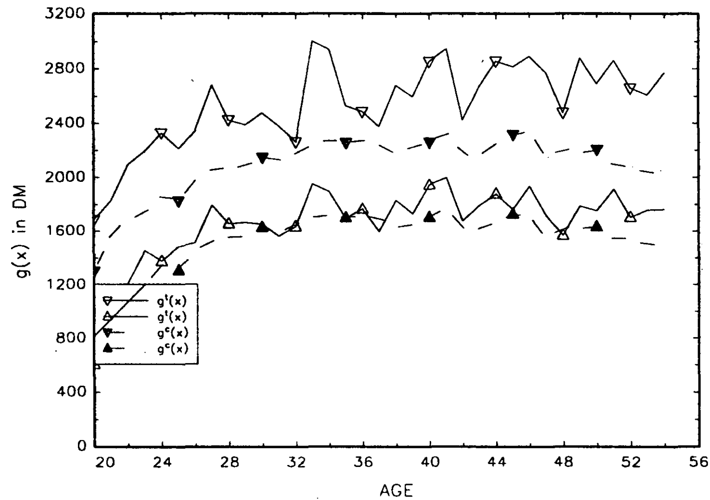
The results for women (Table D.1 and Figure D.1 in Appendix D) look remarkably similar to those for men. Considering off-the-job training (Table B.2 and Figure B.2 in Appendix B) a similar heterogeneity pattern is found, but the estimates for the treated group are generally lower than for the control group. Thus, no such spiky behaviour of the estimates conditional on age is observed.

Figure 2: Estimates of $g^{t}(x)$ and $g^{c}(x)$ conditional on age for on-the-job training versus no training: probabilities of not being unemployed in %



Note: Figure shows 5% and 95% quantiles of respective bootstrap distributions. Men only.

Figure 3: Estimates of $g^{t}(x)$ and $g^{c}(x)$ conditional on age for on-the-job training versus no training: income in DM



Note: Figure shows 5% and 95% quantile of respective bootstrap distributions. Most of the width is due to the grouped nature of the income variable (see Table 1). Men only.

The lower part of Table 3 and Figure 3 focus on *income*, which is a grouped variable.[26] Table 3 shows that income increases with schooling and that there is some regional variation as well, with people living in the former capital and administrative centre of the GDR having a higher mean income. Income increases with age for those younger than about 30. The differences between mean incomes for the treated population appear to be higher than for the controls. Interestingly, this difference is much more pronounced for women than for men. The income data for off-the-job training do not reveal such differences between training and control group.

Another important point is: how well are the respective quantities determined. Although due to the smaller cell size - the probabilities are estimated for every age group separately - the widths of the confidence intervals appearing in the age plot are somewhat larger than for the other characteristics, generally the estimates seem well determined.

5.2 Bounds

Let us start by considering the bounds on the treatment effects given only the sample estimates described in the previous section. They are contained in Table 4 under the column heading *none*, as well as in Figures 4 and 5 labelled *no assumpt.*. For the indicator variable *not being unemployed* the bounds have width 1. For the income variable the width is 8000. When interpreting the bounds presented, one should have in mind that sampling uncertainty makes the location but not the width of these intervals uncertain. The results show that the bounds are rather one-sided. Technically, this is a result of the estimated $g^t(x)$ being close to the upper bound of the support of the indicator variable, whereas for *income* the estimated $g^t(x)$ is closer to the lower bound. Note that no probability statements can be made about single points in the interval (with the exemption of points very close to the bounds of the interval that are influenced by sampling error), therefore the fact that the interval for men with 12 years of schooling is [-4.5, 98.8] does by no means imply that a positive treatment effect is more likely for this group. It is very well possible that the true expected treatment effect is -3.0, for example. The remaining entries in Table 4 as well as in Figures 4 and 5 are obtained by assuming that members of the treated and untreated population defined by x have the same treatment effect. This assumption, that underlies many regression approaches to evaluation problems,

[26] See Table 1 for the widths of the different groups. As has been described above, the loss of information due to grouping is incorporated in the estimates. Hence, a significant part of the interval given is not due to sampling error but is solely due to grouping. As with the uncertainty due to selection, this part of the interval does not shrink when the sample size increases.

appears to be very powerful.[27] It reduces the width dramatically, but as mentioned above, this assumption is not sufficient to identify the sign of the treatment effect.

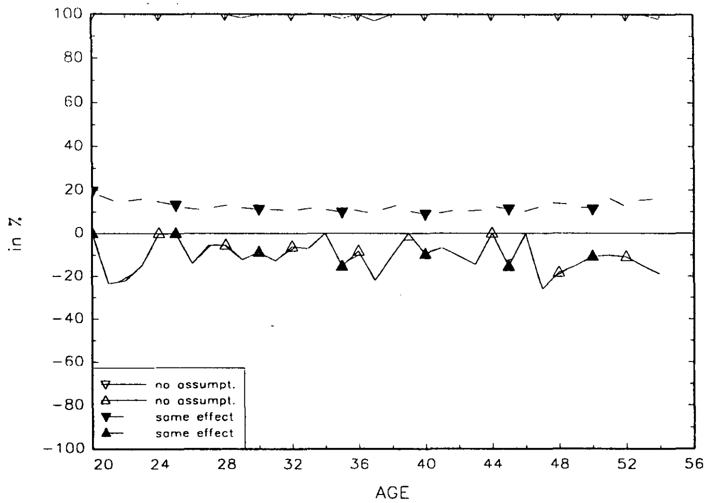
Table 4: Bounds for the treatment effects conditional on training participation for on-the-job training versus no training: no restrictions, same expected treatment effects for treated and controls

	Probabilities of not being unemployed in %				Income in DM			
Restrictions X-variables	none		same effect for treated and controls		none		same effect for treated and control	
Federal states (Länder)								
Berlin (East)	-6.4	98.9	-6.4	9.6	-5782	2890	-2473	2890
Brandenburg	-7.7	97.6	-7.7	11.0	-6187	2420	-2026	2420
Mecklenburg-Vorpommern	-10.0	97.4	-10.0	15.5	-6346	2294	-1981	2294
Sachsen	-5.6	97.9	-5.6	9.1	-6158	2387	-2046	2387
Sachsen-Anhalt	-9.1	96.9	-9.1	11.1	-6329	2260	-1999	2260
Thüringen	-5.7	99.1	-5.7	10.5	-6195	2423	-2046	2423
Years of schooling (highest degree)								
12	-4.5	98.8	-4.5	5.9	-5717	2919	-2686	2919
10	-5.9	96.7	-5.9	9.0	-6239	2251	-2016	2251
8 or no degree	-11.6	96.0	-11.6	16.5	-6498	2076	-1787	2076

Note: Sampling uncertainty due to the estimation of $g^t(x)$ and $g^c(x)$ is accounted for by showing the 5% and 95% quantiles of the bootstrap sampling distribution of the lower repectively upper bounds of the intervals. Men only.

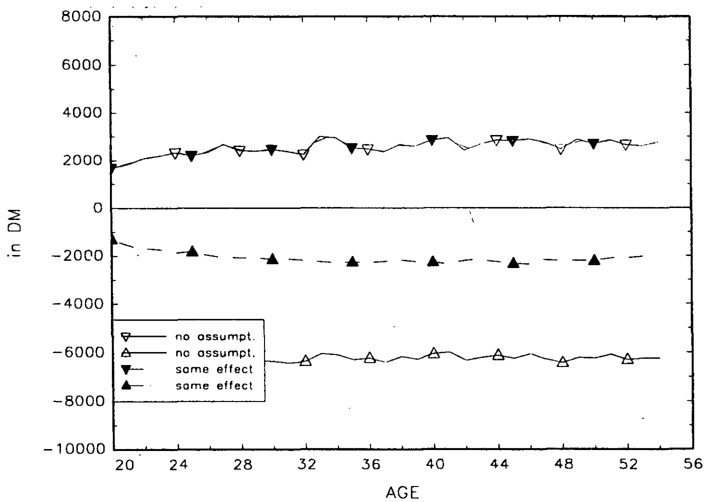
[27] Note that this assumption implies that the selection process into training does not depend on the realised average returns from training. This could be the case, either because the selection does account for the returns, and / or because the estimated returns prior to training are independent of the realised returns (due to unforeseen changes in the economy or insufficient or wrong information about the potential participant).

Figure 4: Bounds for the treatment effects conditional on age and training participation in % (not being unemployed) for on-the-job training versus no training: no restrictions, same expected treatment effects for treated and controls



Note: Men only.

Figure 5: Bounds for the treatment effects conditional on age and training participation in DM (income) for on-the-job training versus no training: no restriction, restriction of same expected treatment effects for treated and controls

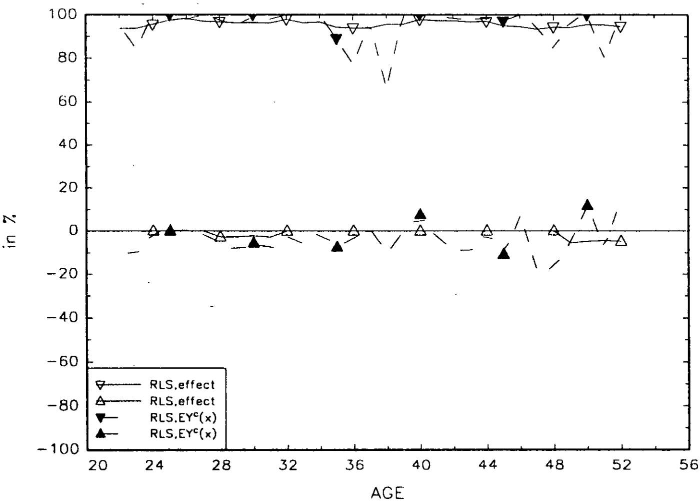


Note: Men only.

Figures 6 and 7 present the results when imposing rolling level-set restriction within narrow age groups (±2 years) on the treatment effects or on $E(Y^c|X=x)$. Note that this expectation is not conditional on treatment status (see discussion of these restrictions above). The corresponding lines in these figures are denoted as *RLS, effect*, and *RLS, EYc(x)*, respectively. Figure 4 already shows that there is little variation of the treatment effect for different ages. Hence, the reductions of width by imposing rolling level-set restrictions on the treatment effect are rather small. They do not exclude zero effects (this is true, independent of the data). The results for these exclusion restrictions imposed on $E(Y^c|X=x)$ look very similar. However, there is a small, but very important difference for the probability of not being unemployed (Figure 6): For some ages, the intervals lie entirely on the positive part of the support of the treatment effect. Thus, for men of ages 40 and 50, there is a positive effect of training if this exclusion restriction is true. Appendix D (Figure D.4.a) shows that the same is true for women of ages 24, 30, and 50. Similarly for off-the-job-training, positive effects appear for men of ages 37 and 50, and for women of age 50 (Figure B.4). Note that this result depends solely on the assumption that, for example $E(Y^c|men, age=48)$ = $E(Y^c|m,49)$ = $E(Y^c|m,50)$ = $E(Y^c|m,51)$ = $E(Y^c|m,52)$ (example for the effect for 50 year old men). Still, this remains a rather technical condition and it appears to be difficult to decide whether this (untestable) condition is true.[28] Additionally, it is unclear why there are only positive effects for specific ages that look rather arbitrary. Note however that such an exclusion restriction still appears to be easier to interpret than the typical distributional assumption often used in parametric treatment (selection) models.

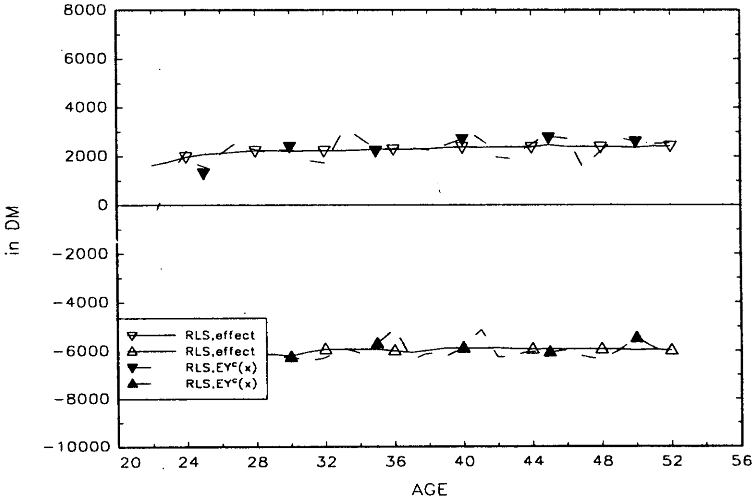
[28] The only evidence against this assumption would be a resulting empty set for the treatment effect. Given the large size of the intervals, this test has extremely low power in the current setting.

Figure 6: Bounds for the treatment effects conditional on age and training participation in % (not being unemployed) for on-the-job training versus no training: rolling level-set restriction within narrow age groups for treatment effect or $E(Y^c|X=x)$



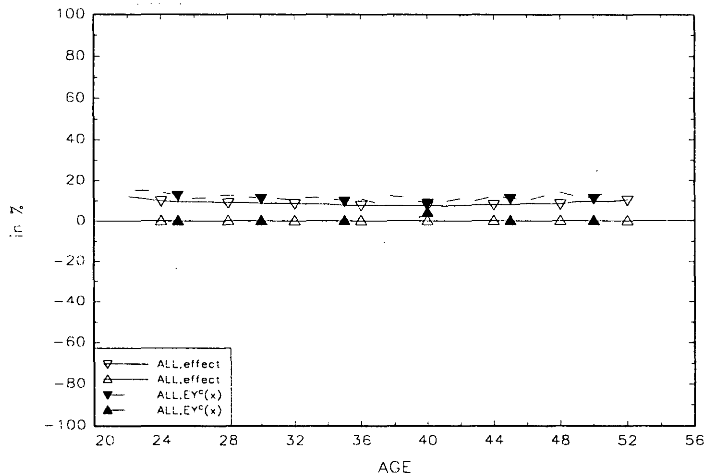
Note: Men only. Level-set restriction is for ±2 years.

Figure 7: Bounds for the treatment effects conditional on age and training participation in DM (income) for on-the-job training versus no training: rolling level-set restriction within narrow age groups for treatment effect or $E(Y^c|X=x)$



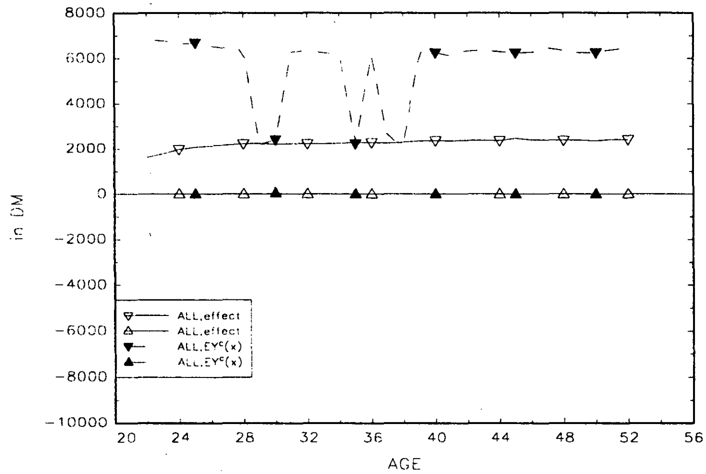
Note: Men only. Level-set restriction is for ±2 years.

Figure 8: Bounds for the treatment effects conditional on age and training participation in % (not being unemployed) for on-the-job training versus no training: combining several restrictions



Note: Men only.

Figure 9: Bounds for the treatment effects conditional on age and training participation in DM (income) for on-the-job training versus no training: combining several restrictions



Note: Men only.

The results presented in Figures 8 and 9 are based on the combination of the two different level-set restrictions with the assumption of having the same treatment effect for the treated and the control population as well as with the assumption that the selection into treatment happens only for groups of individuals with a nonnegative expected outcome. As before, only the level-set restriction on $E(Y^c|X=x)$ is potentially powerful enough to exclude zero treatment effects. The combination of the assumptions does shrink the bounds rather drastically. For the probability of not being unemployed (Figure 8), they collapse for some ages to almost a single point.[29] For the income variable the width of about DM 2000 is still substantial (Figure 9). For the case of off-the-job training there are also positive income effects for age 26 (Figure B.5). The width of the interval of the effect on income is also smaller than in the case of on-the-job training. Combing these assumptions leads to positive effects, particularly for the probability of not being unemployed, for on-the-job training when directly compared to off-the-job training (Figure C.5).

6 Conclusion and outlook

This paper is one of the very few applications of the approach suggested by Manski in various papers to find nonparametric bounds on treatment effects. One of its aims is to explore the potential of this approach for evaluation studies. For the particular case under consideration, that is continuous vocational training in East Germany, the results appear to be rather mixed in this respect. The first finding concerns the width of the intervals and emphasises the fundamental problem of all evaluation studies: Without good knowledge of the relationship between potential outcomes and the selection / assignment process, it is very difficult to bound the treatment effects strictly away from zero. However, in some cases suitable exclusion restrictions are indeed capable of bounding the treatment effects away from zero. Unfortunately, they appear difficult to justify in the context of this paper. Furthermore, they still allow a wide range of possible values for the treatment effects.

That implies that the use of bounds as tests for predictions of parametric selection models will result in a test of very low power. However, they are certainly a useful tool for judging the impact of certain assumptions, such as exclusion restrictions for attributes or treatment status, on the width of the intervals.

Future research might be directed to the issue of providing some distribution theory, perhaps by modified bootstrap procedures, for the estimated locations and widths (in case of exclusion restrictions) of the intervals. Furthermore,

[29] The fact that there are some effects that appear as positive in Figure 6 do not appear to be positive in Figure 8 is the result of taking account of sampling error.

additional restrictions should be explored that could be used to shrink the intervals. Ideally a smooth process of imposing restrictions would start with the no-information case and end with point estimates of different selection models.

References

Angrist, J.D., G.W. Imbens and D.B. Rubin (1996): "Identification of Causal Effects Using Instrumental Variables", *Journal of the American Statistical Association*, 91, 444-472, with discussion by J.J. Heckman, R.A. Moffitt, J.M. Robins and S. Greenland, and R.P. Rosenbaum.

Bell, S.H., L.L. Orr, J.D. Blomquist, and G.G. Cain (1995): *Program Applicants as a Comparison Group in Evaluating Training Programs*, Upjohn: Kalamazoo.

Burtless, G. (1995): "The Case for Randomized Field Trials in Economic and Policy Research", *Journal of Economic Perspectives*, 9, 63-84.

Card, D. and D. Sullivan (1988): "Measuring the Effect of Subsidized Training Programs on Movements in and out of Employment", *Econometrica*, 56, 497-530.

Fitzenberger, B. and H. Prey (1996): "Training in East Germany: An Evaluation of the Effects on Employment and Earnings", *Unpublished manuscript*, University of Konstanz.

Heckman, J.J. and J.A. Smith (1995): "Assessing the Case for Social Experiments", *Journal of Economic Perspectives*, 9, 85-110.

Heckman, J. J. (1994): "Is Job Training Oversold?", *The Public Interest*, 115, 91-115.

Heckman, J.J. and V.J. Hotz (1989): "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training", *Journal of the American Statistical Association*, 84, 862-880 (includes comments by Holland and Moffitt and a rejoinder by Heckman and Hotz).

Hall, P. (1994): "Methodology and Theory for the Bootstrap", in: Engle, R.F. and D.L. McFadden, Hrsg., *Handbook of Econometrics, Vol. 4*, 2342-2381, Amsterdam: North-Holland.

Holland, P.W. (1986): "Statistics and Causal Inference", *Journal of the American Statistical Society*, 81, 945-970 (includes comments by Cox, Granger, Glymour, Rubin and a rejoinder by Holland).

Imbens, G.W. and J.D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, 62, 446-475.

LaLonde, R.J. (1986): "Evaluating the Econometric Evaluations of Training Programs with Experimental Data", *American Economic Review*, 76, 604-620.

LaLonde, R.J. (1995): "The Promise of Public Sector-Sponsored Training Programs", *Jounal of Economic Perspectives*, 9, 149-168.

Lechner, M. (1995): "Effects of Continuous Off-the-job Training in East Germany after Unification", *Discussion Paper*, # 95-27, Zentrum für Europäische Wirtschaftsforschung, Mannheim.

Lechner, M. (1996a): "An Evaluation of Public Sector Sponsored Continuous Vocational Training Programs in East Germany", *Beiträge zur angewandten Wirtschaftsforschung*, # 539-96, University of Mannheim Discussion Papers.

Lechner, M. (1996b): " The Effects of Enterprise-related Continuous Vocational Training in East Germany on Individual Employment and Earnings", *Beiträge zur angewandten Wirtschaftsforschung*, # 542-96, University of Mannheim Discussion Papers.

Lynch, L.M. (1994): *Training and the Private Sector - International Comparisons*, Chicago: University of Chicago Press.

Manski, C.F. (1989): "The Anatomy of the Selection Problem", *Journal of Human Ressources*, 24, 343-360.

Manski, C.F. (1990): "Nonparametric Bounds on Treatment Effects", *American Economic Review, Papers and Proceedings*, 80, 319-323.

Manski, C.F. (1993a): "The Selection Problem", in: C.A. Sims, *Sixth World Congress of the Econometric Society in Barcelona*, Vol. I, 143-170, Cambridge: Cambridge University Press.

Manski, C.F. (1993b): "The Selection Problem in Econometrics and Statistics", in Maddala, G.S., Rao, C.R. and Vinod, H.D. (eds.), *Handbook of Statistics, Vol. 11: Econometrics*, Ch. 3, Amsterdam: North-Holland.

Manski, C.F. (1995): *Identification Problems in the Social Sciences*, Cambridge (MA): Harvard University Press.

Manski, C.F., G.D. Sandefur, S. McLanahan, and D. Powers (1992): "Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation", *Journal of the American Statistical Association*, 87, 25-37.

Rubin, D.B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66, 688-701.

Rubin, D.B. (1977): "Assignment of Treatment Group on the Basis of a Covariate", *Journal of Educational Statistics*, 2, 1-26.

Powell, J.L. (1994): "Estimation of Semiparametric Models", in Engle, R.F. and D.L. McFadden, Hrsg., *Handbook of Econometrics, Vol. 4*, 2444-2521, Amsterdam: North-Holland.

Sobel, M.E. (1994): "Causal Inference in the Social and Behavioral Sciences", in: G. Arminger, C.C. Clogg and M.E. Sobel (eds.): *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, New York: Plenum Press.

Statistisches Bundesamt (1994): *Bevölkerung und Erwerbstätigkeit* (Fachserie 1), *Stand und Entwicklung der Erwerbstätigkeit 1993* (Reihe 4.1.1), Stuttgart: Metzler-Poeschel.

Pfeiffer, F. and J. Brade (1995): "Weiterbildung, Arbeitszeit und Lohneinkommen", *ZEW-discussion paper*, # 95-14.

Werwatz, A. (1996): "How firm-specific is German apprenticeship training", *SFB 373 Discussion Paper*, # 96/12, Humboldt University Berlin.

Appendix A: Econometrics

Table A.1.a: Bounds for $\gamma^0(x)$

	$B_\gamma^L(x)$	$B_\gamma^U(x)$
No data	$L^t(x)-U^c(x)$	$U^t(x)-L^c(x)$
No assumptions	$[g^t(x)-u^c(x,1)]p(x)+$ $[\ell^t(x,0)-g^c(x)][1-p(x)]$	$[g^t(x)-\ell^c(x,1)]p(x)+$ $[u^t(x,0)-g^c(x)][1-p(x)]$
Local exclusion for $\gamma^0(x)$, $\forall x \in \chi^0$	$\sup\limits_{x\in\chi^0}\{[g^t(x)-u^c(x,1)]p(x)+$ $[\ell^t(x,0)-g^c(x)][1-p(x)]\}$	$\inf\limits_{x\in\chi^0}\{[g^t(x)-\ell^c(x,1)]p(x)+$ $[u^t(x,0)-g^c(x)][1-p(x)]\}$
Local exclusion for $E(Y^t\|X=x)$, $\forall x\in\chi^{0,t}$, $E(Y^c\|X=x)$, $\forall x\in\chi^{0,c}$	$\sup\limits_{x\in\chi^{0,t}}\{g^t(x)p(x)+\ell^t(x,0)[1-p(x)]\}-$ $\inf\limits_{x\in\chi^{0,c}}\{g^c(x)[1-p(x)]+u^c(x,1)]p(x)\}$	$\inf\limits_{x\in\chi^{0,t}}\{g^t(x)p(x)+u^t(x,0)[1-p(x)]\}-$ $\sup\limits_{x\in\chi^{0,c}}\{g^c(x)[1-p(x)]-\ell^c(x,1)]p(x)\}$
Local exclusion for $E(Y^c\|X=x,S=1)$, $\forall x\in\chi^{0,c}$, and $E(Y^t\|X=x,S=0)$, $\forall x\in\chi^{0,t}$	$[g^t(x)-\inf\limits_{x\in\chi^{0,c}}u^c(x,1)]p(x)+$ $[\sup\limits_{x\in\chi^{0,t}}\ell^t(x,0)-g^c(x)][1-p(x)]$	$[g^t(x)-\sup\limits_{x\in\chi^{c,0}}\ell^c(x,1)]p(x)+$ $[\inf\limits_{x\in\chi^{0,t}}u^t(x,0)-g^c(x)][1-p(x)]$
Selection is such that $E(Y^t\|X=x,S=1)-E(Y^c\|X=x,S=1)\geq 0$	$[\ell^t(x,0)-g^c(x)][1-p(x)]$	$[g^t(x)-\ell^c(x,1)]p(x)+$ $[u^t(x,0)-g^c(x)][1-p(x)]$
$\gamma^0(x)=\theta^0(x)=\xi^0(x)$, $\forall x\in\chi^0$	$\max[B_\theta^L(x),B_\xi^L(x)]$	$\min[B_\theta^U(x),B_\xi^U(x)]$

Note: $B_\xi^L(x)$ and $B_\xi^U(x)$ denote the lower and the upper bounds of the treatment effects for the nontreated (*S=0*).

Table A.1.b: Interval widths for $\gamma^0(x)$

	$W_\gamma(x)$
No data	$U^t(x) - L^t(x) + U^c(x) - L^c(x)$
No assumptions	$[u^c(x,1) - \ell^c(x,1)]p(x) + [u^t(x,0) - \ell^t(x,0)][1-p(x)]$
Local exclusion for $\gamma^0(x)$, $\forall x \in \chi^0$	$\inf_{x\in\chi^0}\{[g^t(x) - \ell^c(x,1)]p(x) + [u^t(x,0) - g^c(x)][1-p(x)]\} -$ $\sup_{x\in\chi^0}\{[g^t(x) - u^c(x,1)]p(x) + [\ell^t(x,0) - g^c(x)][1-p(x)]\}$
Local exclusion for $E(Y^t\|X=x)$, $\forall x \in \chi^{0,t}$, $E(Y^c\|X=x)$, $\forall x \in \chi^{0,c}$	$\inf_{x\in\chi^{0,t}}\{g^t(x)p(x) + u^t(x,0)[1-p(x)]\} - \sup_{x\in\chi^{0,c}}\{g^c(x)[1-p(x)] + \ell^c(x,1)]p(x)\} -$ $\sup_{x\in\chi^{0,t}}\{g^t(x)p(x) + \ell^t(x,0)[1-p(x)]\} + \inf_{x\in\chi^{0,c}}\{g^c(x)[1-p(x)] + u^c(x,1)]p(x)\}$
Local exclusion for $E(Y^c\|X=x,S=1)$, $\forall x \in \chi^{0,c}$, and $E(Y^t\|X=x,S=0)$, $\forall x \in \chi^{0,t}$	$[\inf_{x\in\chi^{0,c}} u^c(x,1) - \sup_{x\in\chi^{0,c}} \ell^c(x,1)]p(x) + [\inf_{x\in\chi^{0,t}} u^t(x,0) - \sup_{x\in\chi^{0,t}} \ell^t(x,0)][1-p(x)]$
Selection is such that $E(Y^t\|X=x,S=1) - E(Y^c\|X=x,S=1) \geq 0$	$[g^t(x) - \ell^c(x,1)]p(x) + [u^t(x,0) - \ell^t(x,0)][1-p(x)]$
$\gamma^0(x) = \theta^0(x) = \xi^0(x)$, $\forall x \in \chi^0$	$\min[B_\theta^U(x), B_\xi^U(x)] - \max[B_\theta^L(x), B_\xi^L(x)]$

Table A.2.a: Bounds for $\theta^0(x)$

	$B_\theta^L(x)$	$B_\theta^U(x)$
No data	$\ell^t(x,1) - u^c(x,1)$	$u^t(x,1) - \ell^c(x,1)$
No assumptions	$g^t(x) - u^c(x,1)$	$g^t(x) - \ell^c(x,1)$
Local exclusion for $\theta^0(x)$, $\forall x \in \chi^0$	$\sup_{x\in\chi^0}\{g^t(x) - u^c(x,1)\}$	$\inf_{x\in\chi^0}\{g^t(x) - \ell^c(x,1)\}$
Local exclusion for $E(Y^c\|X=x)$, $\forall x \in \chi^0$ *)	$g^t(x) - g^c(x) + \dfrac{g^c(x) - \inf_{x\in\chi^0} v^u(x)}{p(x)}$	$g^t(x) - g^c(x) + \dfrac{g^c(x) - \sup_{x\in\chi^0} v^\ell(x)}{p(x)}$
Local exclusion $E(Y^c\|X=x,S=1)$, $\forall x \in \chi^{0,c}$	$g^t(x) - \inf_{x\in\chi^{0,c}} u^c(x,1)$	$g^t(x) - \sup_{x\in\chi^{0,c}} \ell^c(x,1)$
Selection is such that $E(Y^t\|X=x,S=1) - E(Y^c\|X=x,S=1) \geq 0$	0	$g^t(x) - \ell^c(x,1)$

Note: *) $v^\ell(x) = \ell^c(x,1)p(x) + g^c(x)[1-p(x)]$, $v^u(x) = u^c(x,1)p(x) + g^c(x)[1-p(x)]$. Obtained from $E(Y^c|X=x) = E(Y^c|X=x,S=1)p(x) + g^c(x)[1-p(x)]$ and equation (2").

Table A.2.b: Interval widths for $\theta^0(x)$

	$W_\theta(x)$
No data	$u^t(x,1)-\ell^t(x,1)+u^c(x,1)-\ell^c(x,1)$
No assumptions	$u^c(x,1)-\ell^c(x,1)$
Local exclusion for $\theta^0(x)$, $\forall x \in \chi^0$	$\inf\limits_{x\in\chi^0}\{g^t(x)-\ell^c(x,1)\}-\sup\limits_{x\in\chi^0}\{g^t(x)-u^c(x,1)\}$
Local exclusion for $E(Y^c \mid X=x)$, $\forall x \in \chi^0$	$\dfrac{\inf\limits_{x\in\chi^0} v^u(x)-\sup\limits_{x\in\chi^0} v^\ell(x)}{p(x)}$
Local exclusion for $E(Y^c \mid X=x, S=1)$, $\forall x \in \chi^{0,c}$	$\inf\limits_{x\in\chi^{0,c}} u^c(x,1)-\sup\limits_{x\in\chi^{0,c}} \ell^c(x,1)$
Selection is such that $E(Y^t \mid X=x,S=1)-E(Y^c \mid X=x,S=1)\geq 0$	$g^t(x)-\ell^c(x,1)$

Note: See note on Table A.2.a.

Combining the cases given in Tables A.1.a and A.2.a with the assumption that treatment effects are the same in the treated and control population, i.e. $\theta^0(x) = \gamma^0(x)$, $\forall x \in \chi^0$ is straightforward: The lower bound is given by $\sup\limits_{x\in\chi^0}\{\theta^0(x), \gamma^0(x)\}$ and the upper bound is equal to $\inf\limits_{x\in\chi^0}\{\theta^0(x), \gamma^0(x)\}$.

Table A.3: Bounds for the selection effect $\lambda^\theta(x)$

	$B^L_{\lambda^\theta}(x)$	$B^U_{\lambda^\theta}(x)$
No assumptions	$g^c(x)-u^c(x,1)$	$g^c(x)-\ell^c(x,1)$
Local exclusion for $\theta^0(x)$, $\forall x \in \chi^0$	$\sup\limits_{x\in\chi^0}\{g^t(x)-u^c(x,1)\}+g^c(x)-g^t(x)$	$\inf\limits_{x\in\chi^0}\{g^t(x)-\ell^c(x,1)\}+g^c(x)-g^t(x)$
Local exclusion for $E(Y^c \mid X=x)$, $\forall x \in \chi^{0\,*)}$	$\dfrac{g^c(x)-\inf\limits_{x\in\chi^0} v^u(x)}{p(x)}$	$\dfrac{g^c(x)-\sup\limits_{x\in\chi^0} v^\ell(x)}{p(x)}$
Local exclusion for $E(Y^c \mid X=x,S=1)$, $\forall x \in \chi^{c,0}$	$g^c(x)-\inf\limits_{x\in\chi^{0,c}} u^c(x,1)$	$g^c(x)-\sup\limits_{x\in\chi^{0,c}} \ell^c(x,1)$
Selection is such that $E(Y^t \mid X=x,S=1)$ - $E(Y^c \mid X=x,S=1)\geq 0$	$g^c(x)-g^t(x)$	$g^c(x)-\ell^c(x,1)$

Note: The width of the interval of the selection effect is the same as for the respective treatment effect given in Table A.2.a.
See also notes of Table A.2.

Manski (1990) introduced an additional restriction which assumes that only individuals with a nonnegative effect (all of them) are selected. However, in a social context this is hardly plausible, because it may very well require too many ressources and too much information (about the future!) for those who select participants. Additionally, it a priori restricts $\theta^0(x)$ (but not $\gamma^0(x)$) to be nonnegative. This does not appear to be a credible strategy when evaluating social programs. Several other restrictions appear in the different papers by Manski. Those most closely related to our problem are discussed in Manski (1993a, p. 163, 164). However, the assumption of *ordered outcomes* a priori assumes that outcomes when treated are never less than outcomes when not treated. Obviously, such an assumption is not attractive in the context of this paper. The assumption of ordered outcomes $P(y^t = y^c + \alpha(x) | X = x) = 1$ also appears to be too restrictive in this context. One of the reasons is that the shift is not in expectation, but with probability one. Assuming instead that $g^t(x) = E(Y^c | X = x, S = 1) + \alpha(x)$, and that $E(Y^c | X = x, S = 1)$ as well as $\alpha(x)$ are constant for at least two different values of x (level-set restriction), then $\theta^0(x)$ is identified provided $g^t(x)$ varies. It is however not plausible that $E(Y^c | X = x, S = 1)$ should be constant in some region of the X-space, while $E(Y^t | X = x, S = 1)$ [$= g^t(x)$] is assumed to vary exactly in the same region.

Appendix B: Results for off-the-job training versus no training

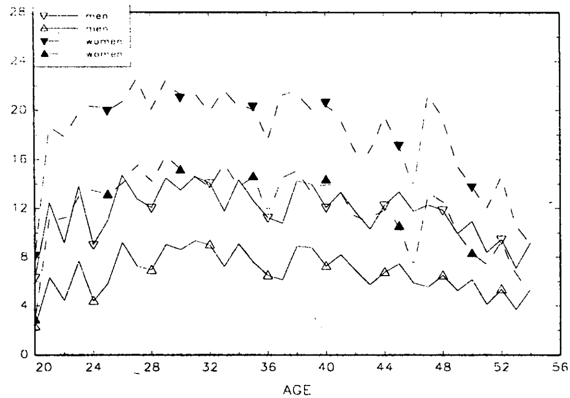
B.1 Identified quantities ($p(x)$, $g^t(x)$, $g^c(x)$)

Table B.1: Probabilities for off-the-job training versus no training in %

X-variables	Men		Women	
Federal states (Länder)				
Berlin (East)	10.9	14.1	15.8	19.7
Brandenburg	7.8	9.8	13.6	16.1
Mecklenburg-Vorpommern	9.0	11.7	15.0	18.3
Sachsen	7.6	8.8	12.6	14.5
Sachsen-Anhalt	8.6	10.8	13.2	15.9
Thüringen	7.7	9.7	14.7	17.2
Years of schooling (highest degree)				
12	16.2	19.0	20.2	23.4
10	8.5	9.5	15.4	16.8
8 or no degree	3.4	4.6	7.2	8.9

Note: Table shows 5% and 95% quantiles of respective bootstrap distributions.

Figure B.1: Probabilities for off-the-job training versus no training in %



Note: Figure shows 5% and 95% quantile of respective bootstrap distributions.

Table B.2: Estimates of $g^{t}(x)$ and $g^{c}(x)$ conditional on schooling and federal state for off-the-job training versus no training

	Men				Women			
X-variables	$g^{t}(x)$		$g^{c}(x)$		$g^{t}(x)$		$g^{c}(x)$	
	Probabilities of not being unemployed in %							
Years of schooling (highest degree)								
12	81.4	88.0	94.1	95.9	81.6	88.0	91.9	94.3
10	81.5	86.0	91.0	92.0	64.3	68.7	79.7	81.3
8 or no degree	60.4	76.2	83.5	85.8	52.2	63.4	68.6	71.8
Federal states (Länder)								
Berlin (East)	76.2	87.4	90.3	93.3	73.6	83.4	84.0	88.0
Brandenburg	74.1	83.4	89.0	91.2	62.2	71.2	77.6	80.6
Mecklenburg-Vorpommern	73.8	84.2	84.5	87.6	61.2	71.3	75.1	79.1
Sachsen	80.9	87.8	90.9	92.4	64.4	71.3	78.4	80.6
Sachsen-Anhalt	78.9	86.9	88.8	91.1	60.3	69.7	78.0	81.1
Thüringen	82.1	90.8	89.5	91.8	69.6	77.3	77.0	80.3
	Income in DM							
Years of schooling (highest degree)								
12	2148	2764	2155	2683	1640	2202	1672	2156
10	1529	2027	1593	2015	1106	1546	1133	1535
8 or no degree	1115	1739	1353	1787	895	1404	900	1307
Federal states (Länder)								
Berlin (East)	1924	2600	1947	2473	1491	2082	1439	1922
Brandenburg	1616	2242	1568	2025	1205	1747	1154	1586
Mecklenburg-Vorpommern	1548	2204	1509	1979	1107	1634	1104	1555
Sachsen	1625	2197	1597	2045	1153	1639	1089	1500
Sachsen-Anhalt	1618	2245	1544	1998	1111	1633	1118	1551
Thüringen	1535	2159	1585	2046	1014	1515	1069	1501

Note: Table shows 5% and 95% quantiles of respective bootstrap distributions. Most of the width of the income variable is due to the grouped nature of it (see Table 1).

Figure B.2: Estimates of $g^t(x)$ and $g^c(x)$ conditional on age for off-the-job training versus no training

Figure B.2.a: Probabilities of not being unemployed in % for men

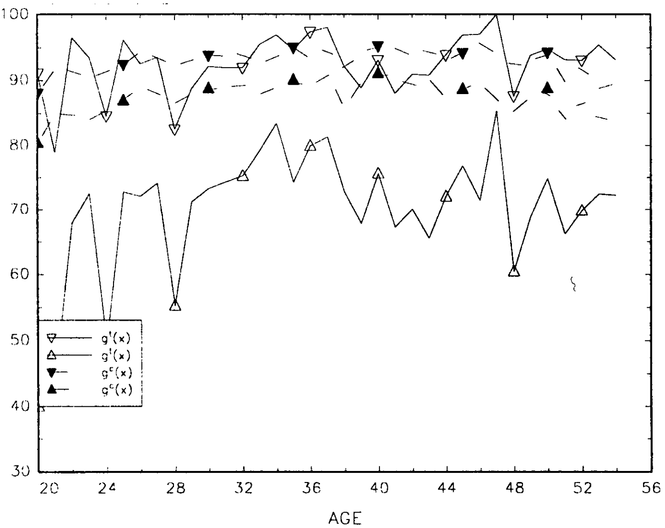


Figure B.2.b: Probabilities of not being unemployed in % for women

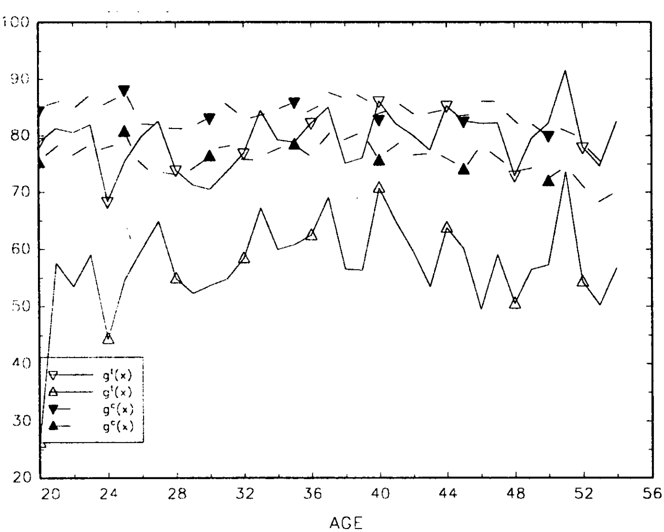


Figure B.2.c: Income in DM for men

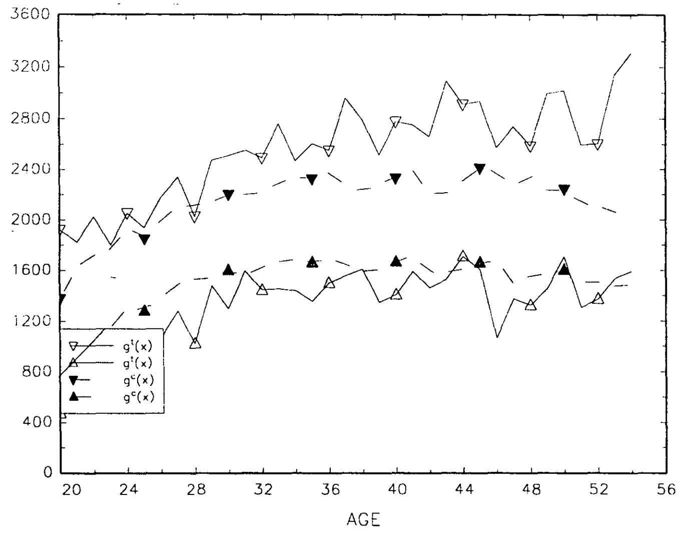
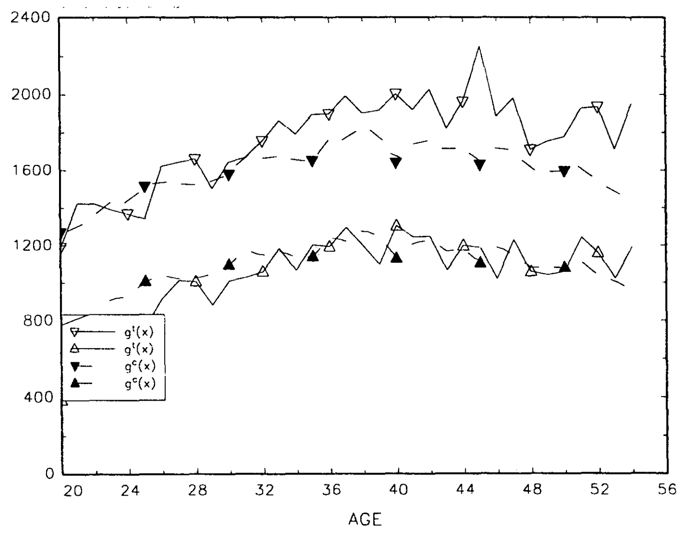


Figure B.2.d: Income in DM for women



Note: Figures shows 5% and 95% quantile of respective bootstrap distributions.

B.2 Bounds

Table B.1: Bounds for the treatment effects conditional on training participation for off-the-job training versus no training

	Men				Women			
Restrictions	none		same effect for		none		same effect for	
X-variables			treated and controls				treated and controls	
	Probabilities of not being unemployed in %							
Years of schooling (highest degree)								
12	-18.6	88.0	-18.6	5.9	-18.4	88.0	-18.4	8.1
10	-18.5	86.0	-18.5	9.0	-35.7	68.7	-35.7	20.3
8 or no degree	-39.6	76.2	-39.6	16.5	-47.8	63.4	-47.8	31.2
Federal states (Länder)								
Berlin (East)	-23.8	87.4	-23.8	9.6	-26.4	83.4	-26.4	16.0
Brandenburg	-25.9	83.4	-25.9	11.0	-37.8	71.2	-37.8	22.4
Mecklenburg-Vorpommern	-26.2	84.2	-26.2	15.5	-38.8	71.3	-38.8	24.8
Sachsen	-19.1	87.8	-19.1	9.1	-35.6	71.3	-35.6	21.6
Sachsen-Anhalt	-21.1	86.9	-21.1	11.1	-39.7	69.7	-39.7	22.0
Thüringen	-17.9	90.8	-17.9	10.5	-30.4	77.3	-30.4	23.0
	Income in DM							
Years of schooling (highest degree)								
12	-5852	2764	-2686	2764	-6359	2202	-2158	2202
10	-6470	2027	-2016	2027	-6893	1546	-1535	1546
8 or no degree	-6884	1739	-1787	1739	-7104	1404	-1308	1404
Federal states (Länder)								
Berlin (East)	-6076	2600	-2473	2600	-6509	2082	-1924	2082
Brandenburg	-6383	2242	-2026	2242	-6794	1747	-1587	1747
Mecklenburg-Vorpommern	-6451	2204	-1981	2204	-6893	1635	-1556	1635
Sachsen	-6364	2197	-2046	2197	-6846	1639	-1500	1639
Sachsen-Anhalt	-6381	2245	-1999	2245	-6889	1633	-1552	1633
Thüringen	-6464	2159	-2046	2159	-6985	1515	-1501	1515

Note: Sampling uncertainty due to the estimation of $g^t(x)$ and $g^c(x)$ is accounted for by showing the 5% and 95% quantiles of the bootstrap distribution of the lower repectively upper bounds of the intervals.

Figure B.3: Bounds for the treatment effects conditional on age and training participation for off-the-job training versus no training: no restriction, restriction of same expected treatment effects for treated and controls

Figure B.3.a: Probabilities of not being unemployed in %-points for men

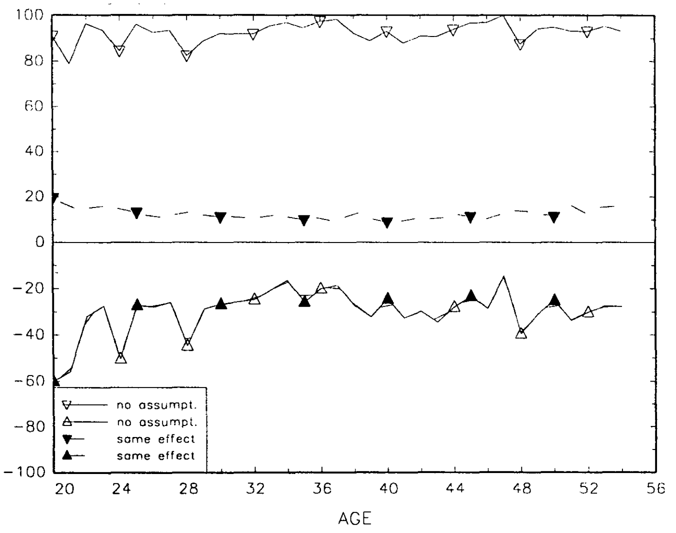


Figure B.3.b: Probabilities of not being unemployed in %-points for women

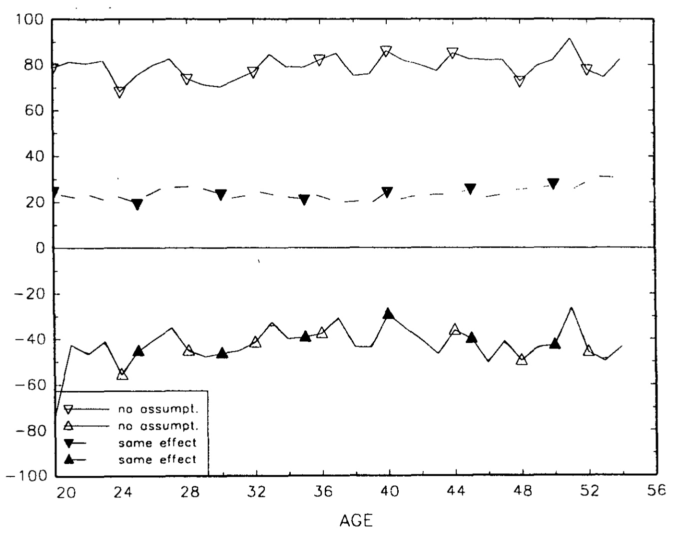


Figure B.3.c: Income in DM for men

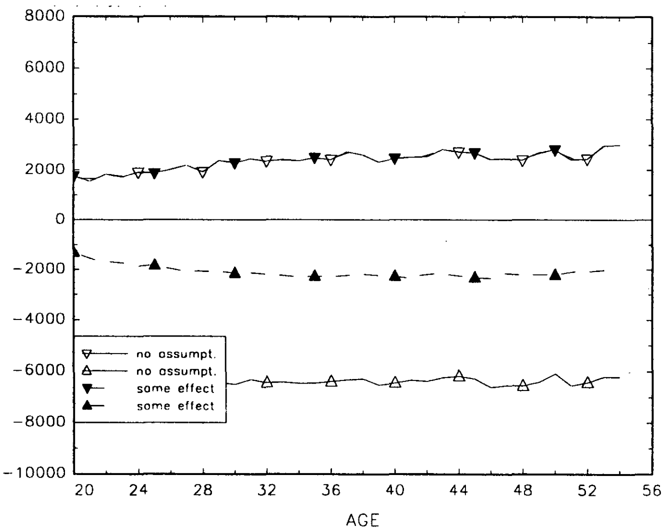


Figure B.3.d: Income in DM for women

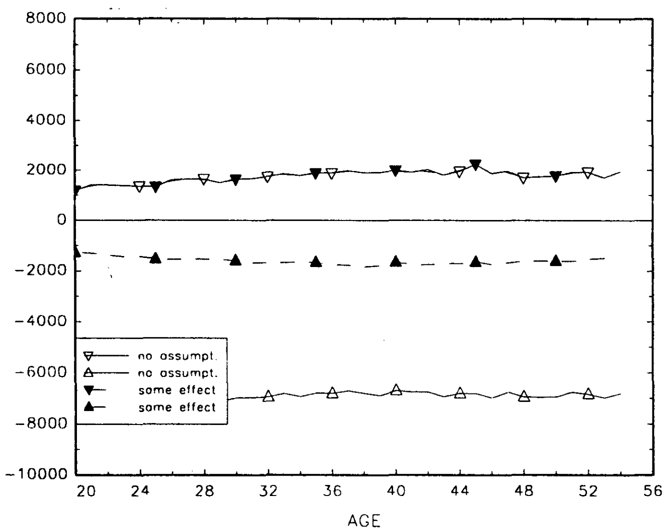


Figure B.4: Bounds for the treatment effects conditional on age and training participation for off-the-job training versus no training: rolling level-set restriction within narrow age groups for treatment effect or $E(Y^c | X = x)$

Figure B.4.a: Probabilities of not being unemployed in %-points for men

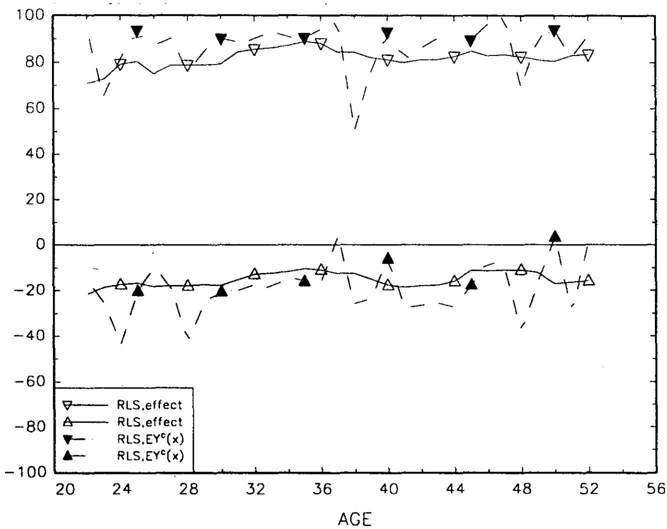


Figure B.4.b: Probabilities of not being unemployed in %-points for women

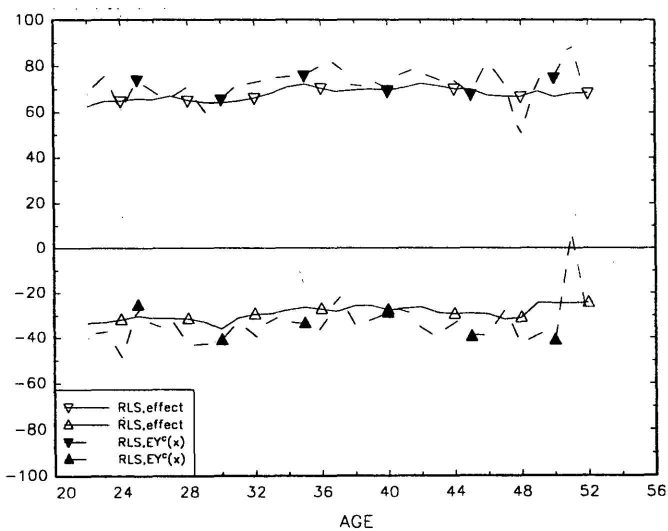


Figure B.4.c: Income in DM for men

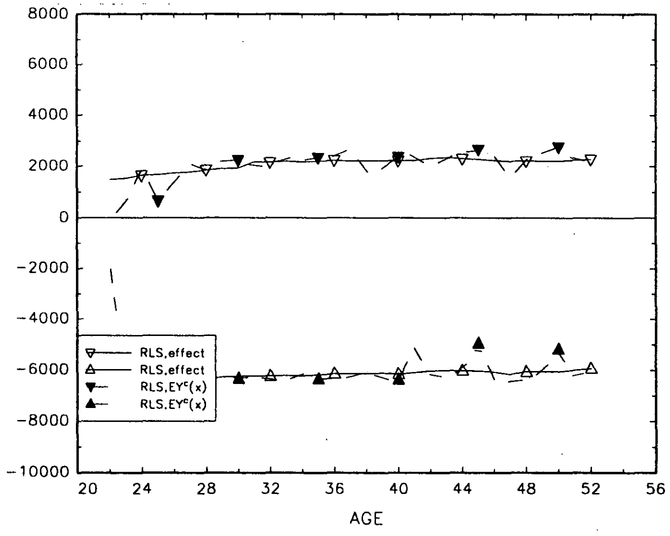
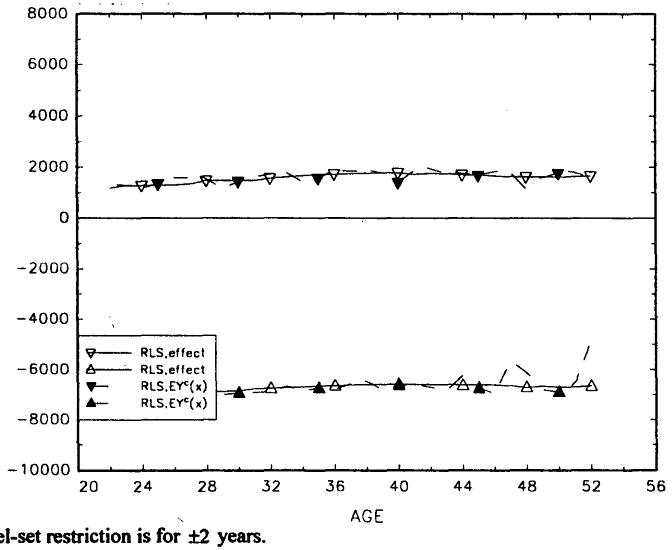


Figure B.4.d: Income in DM for women



Note: Level-set restriction is for ±2 years.

Figure B.5: Bounds for the treatment effects conditional on age and training participation for off-the-job training versus no training: combining several restrictions

Figure B.5.a: Probabilities of not being unemployed in %-points for men

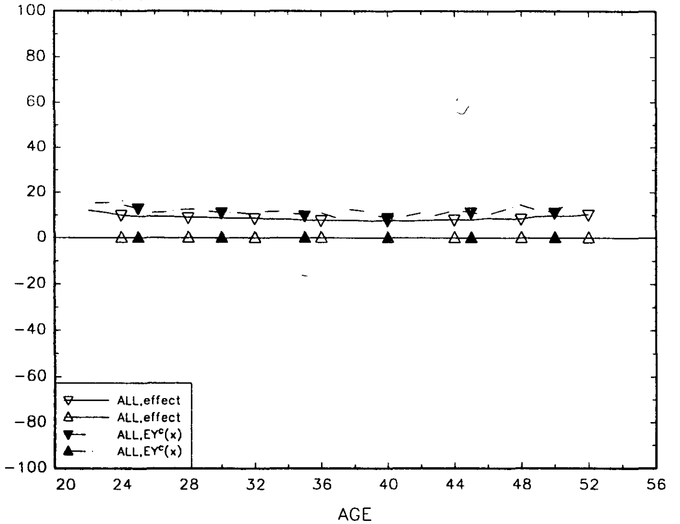


Figure B.5.b: Probabilities of not being unemployed in %-points for women

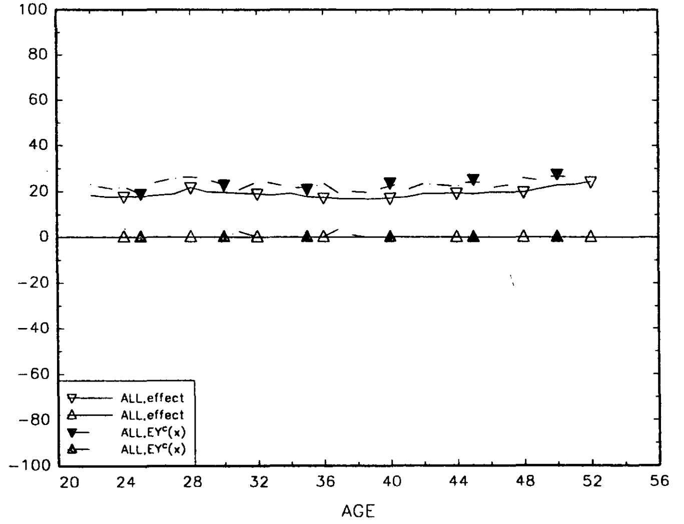


Figure B.5.c: Income in DM for men

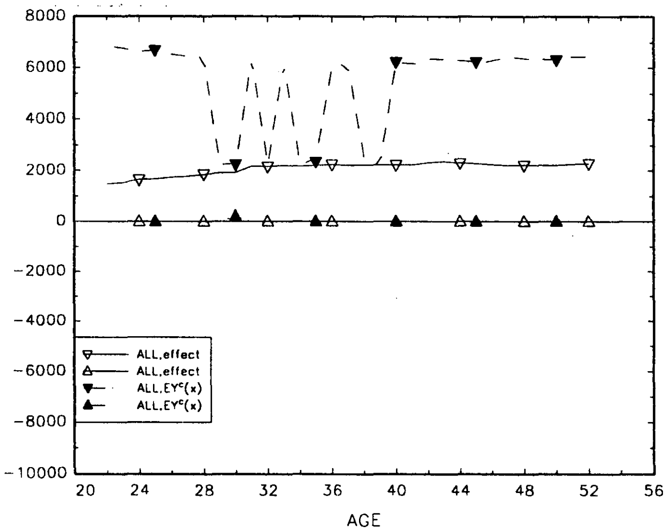
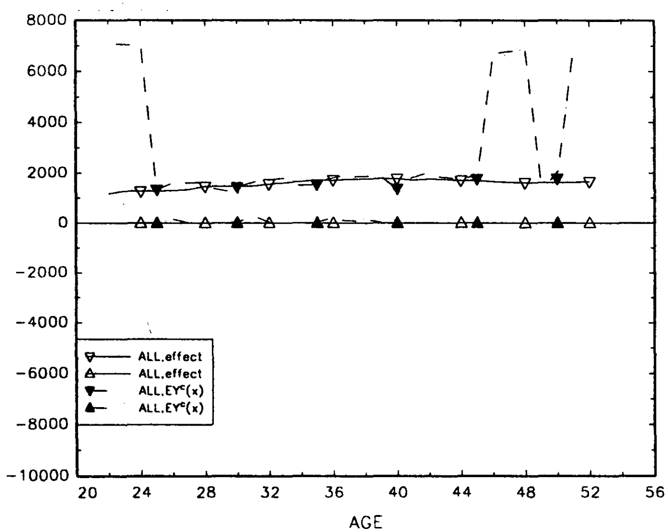


Figure B.5.d: Income in DM for women



Appendix C: Results for on-the-job-training versus off-the-job-training

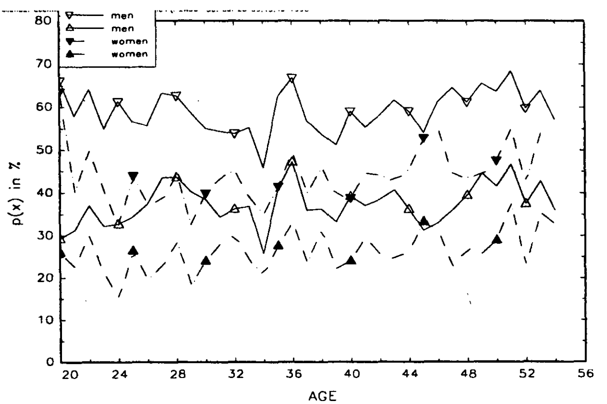
C.1 Identified quantities ($p(x)$, $g^{l}(x)$, $g^{c}(x)$)

Table C.1: Probabilities for on-the-job versus off-the-job training in %

X-variables	Men		Women	
Federal states (Länder)				
Berlin (East)	44.9	55.4	39.6	48.7
Brandenburg	41.5	50.3	27.7	34.8
Mecklenburg-Vorpommern	40.6	51.4	29.0	37.3
Sachsen	48.1	54.3	33.4	39.0
Sachsen-Anhalt	43.3	52.0	33.5	40.6
Thüringen	42.6	51.5	25.5	32.8
Years of schooling (highest degree)				
12	40.4	47.3	35.4	42.2
10	47.2	51.6	32.4	36.1
8 or no degree	49.5	60.2	28.1	36.7

Note: Table shows 5% and 95% quantiles of respective bootstrap distributions.

Figure C.1: Probabilities for on-the-job versus off-the-job training in %



Note: Figure shows 5% and 95% quantiles of respective bootstrap distributions.

Table C.2: Estimates of $g^{t}(x)$ and $g^{c}(x)$ conditional on schooling and federal state for on-the-job versus off-the-job training

	Men				Women			
X-variables	$g^{t}(x)$		$g^{c}(x)$		$g^{t}(x)$		$g^{c}(x)$	
	Probabilities of not being unemployed in %							
Years of schooling (highest degree)								
12	95.5	98.8	81.4	88.0	94.8	98.5	81.6	88.0
10	94.1	96.7	81.5	86.0	91.1	94.4	64.3	68.7
8 or no degree	88.4	96.0	60.4	76.2	85.0	94.7	52.2	63.4
Federal states (Länder)								
Berlin (East)	93.6	98.9	76.2	87.4	95.5	99.5	73.6	83.4
Brandenburg	92.3	97.6	74.1	83.4	90.5	97.2	62.2	71.2
Mecklenburg-Vorpommern	90.0	97.4	73.8	84.2	93.5	98.8	61.2	71.3
Sachsen	94.4	97.9	80.9	87.8	88.9	94.2	64.4	71.3
Sachsen-Anhalt	90.9	96.9	78.9	86.9	84.8	92.7	60.3	69.7
Thüringen	94.3	99.1	82.1	90.8	91.5	97.9	69.6	77.3
	Income in DM							
Years of schooling (highest degree)								
12	2282	2919	2148	2764	1914	2507	895	2202
10	1760	2251	1529	2027	1509	1987	1106	1546
8 or no degree	1501	2076	1115	1739	1274	1836	1640	1404
Federal states (Länder)								
Berlin (East)	2217	2890	1924	2600	1900	2501	1491	2082
Brandenburg	1812	2420	1616	2242	1553	2138	1205	1747
Mecklenburg-Vorpommern	1653	2294	1548	2204	1457	2044	1107	1635
Sachsen	1841	2387	1635	2197	1503	2029	1153	1639
Sachsen-Anhalt	1670	2260	1618	2245	1413	1973	1111	1633
Thüringen	1804	2423	1535	2159	1489	2107	1014	1515

Note: Table shows 5% and 95% quantiles of respective bootstrap distributions.

Figure C.2: Estimates of $g^t(x)$ and $g^c(x)$ conditional on age for on-the-job versus off-the-job training

Figure C.2.a: Probabilities of not being unemployed in %-points for men

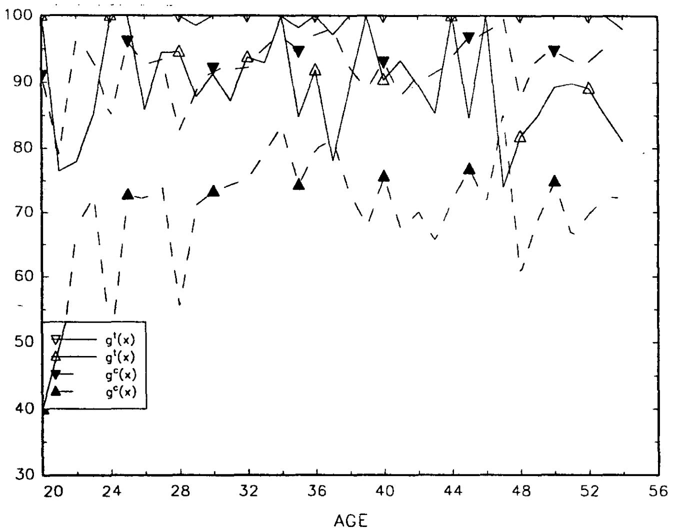


Figure C.2.b: Probabilities of not being unemployed in %-points for women

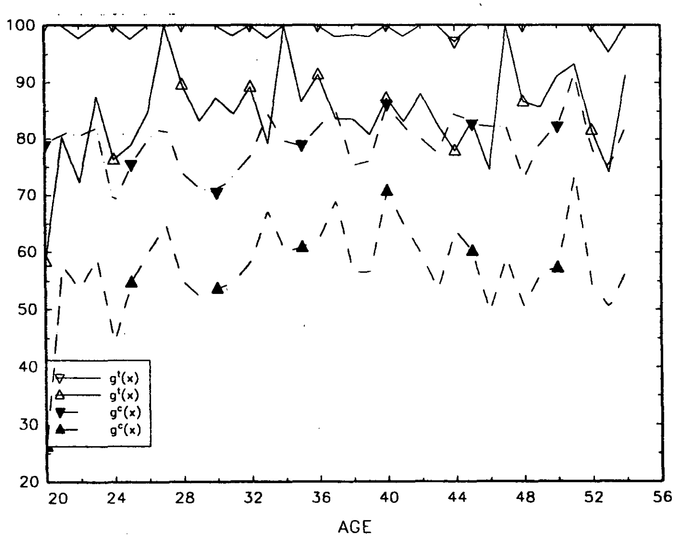


Figure C.2.c: Income in DM for men

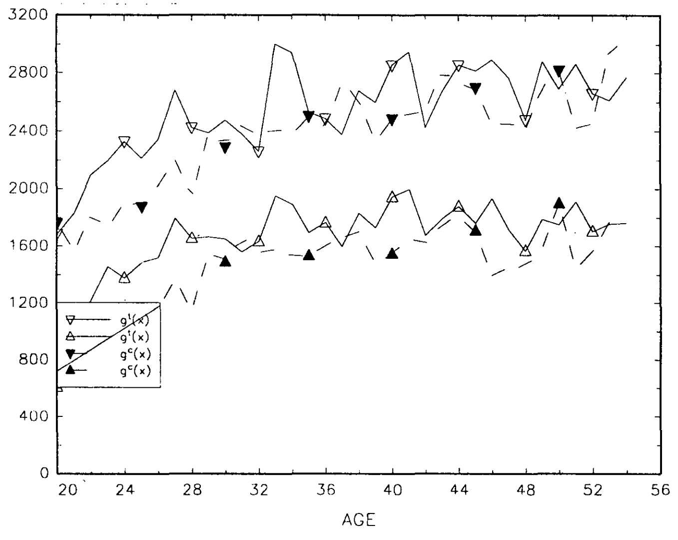
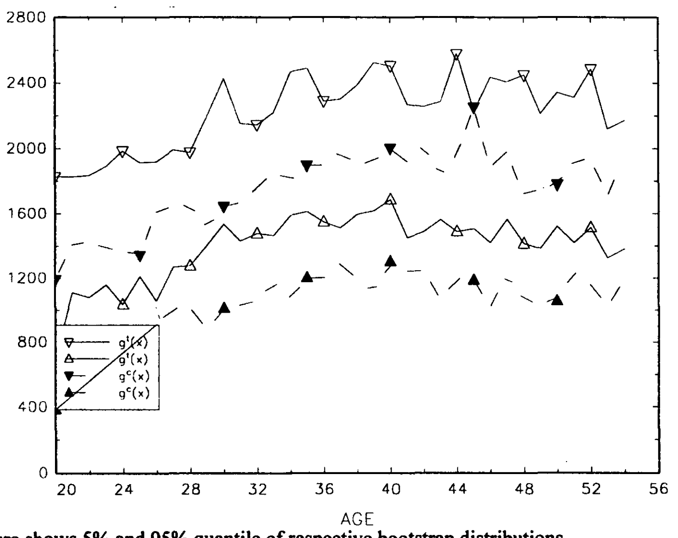


Figure C.2.d: Income in DM for women



Note: Figure shows 5% and 95% quantile of respective bootstrap distributions.

C.2 Bounds

Table C.2: Bounds for the treatment effects conditonal on schooling and federal states for on-the-job training versus off-the-job training

	Men				Women			
Restrictions X-variables	none		same effect for treated and controls		none		same effect for treated and controls	
	Probabilities of not being unemployed in %							
Years of schooling (highest degree)								
12	-52,4	54,2	-4.5	18.4	-57,1	61,8	-5.2	18.1
10	-46,8	57,7	-5.9	18.4	-47,8	55,5	-8.9	35.6
8 or no degree	-40,3	70,2	-11.6	38.9	-47,1	50,3	-15.0	47.7
Federal states (Länder)								
Berlin (East)	-47,4	62,8	-6.4	23.6	-49,5	59,6	-4.5	26.1
Brandenburg	-49,1	59,7	-7.7	25.8	-51,6	56,3	-9.5	37.6
Mecklenburg-Vorpommern	-50,5	59,2	-10.0	26.2	-50,1	58,7	-6.5	38.8
Sachsen	-46,4	60,0	-5.6	19.1	-49,4	56,5	-11.1	35.5
Sachsen-Anhalt	-50,2	58,1	-9.1	21.1	-48,5	58,8	-15.2	39.5
Thüringen	-52,2	57,1	-5.7	17.6	-57,9	50,6	-8.5	30.3
	Income in DM							
Years of schooling (highest degree)								
12	-4077	4583	-2770	2076	-6366	2205	-2204	2507
10	-4168	4428	-2028	2251	-6899	1552	-1547	1987
8 or no degree	-4524	4411	-1739	2919	-7112	1410	-1409	1836
Federal states (Länder)								
Berlin (East)	-4263	4536	-2612	2890	-3951	4824	-2082	2501
Brandenburg	-4147	4632	-2249	2490	-3325	5436	-1748	2138
Mecklenburg-Vorpommern	-4216	4655	-2206	2294	-3387	5421	-1636	2044
Sachsen	-4287	4393	-2199	2387	-3490	5189	-1641	2029
Sachsen-Anhalt	-4285	4490	-2246	2260	-3579	5177	-1638	1973
Thüringen	-4155	4638	-2160	2423	-3086	5656	-1515	2107

Note: Sampling uncertainty due to the estimation of $g^t(x)$ and $g^c(x)$ is accounted for by showing the 5% and 95% quantiles of the bootstrap distribution of the lower repectively upper bounds of the intervals. These are expected treatment effects for individuals randomly drawn from the population in both types of training (γ).

Figure C.3: Bounds for the treatment effects conditional on age for on-the-job training versus off-the-job training: no restriction, restriction of same expected treatment effects for treated and controls

Figure C.3.a: Probabilities of not being unemployed in %-points for men

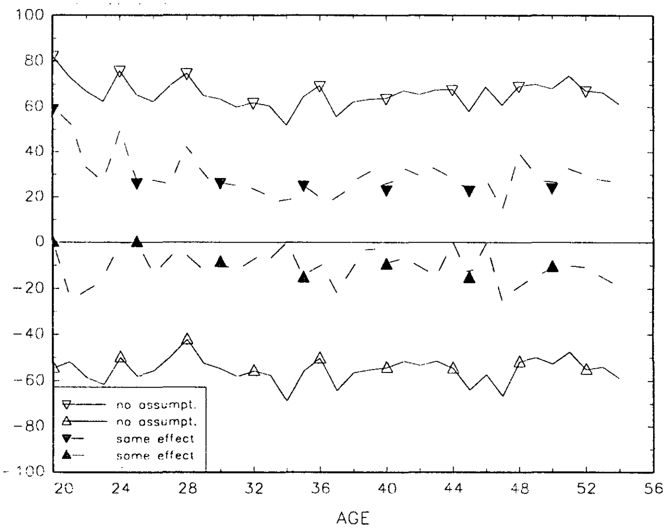


Figure C.3.b: Probabilities of not being unemployed in %-points for women

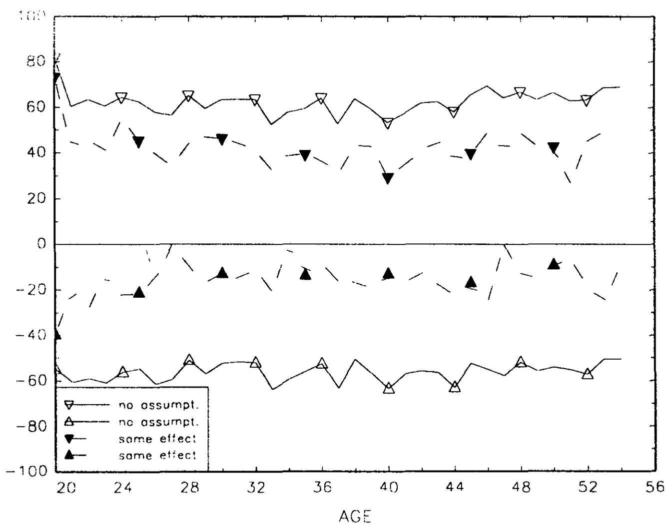


Figure C.3.c: Income in DM for men

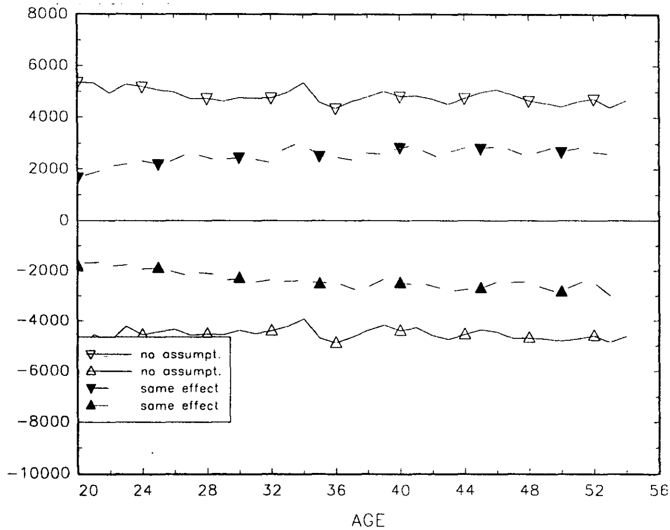
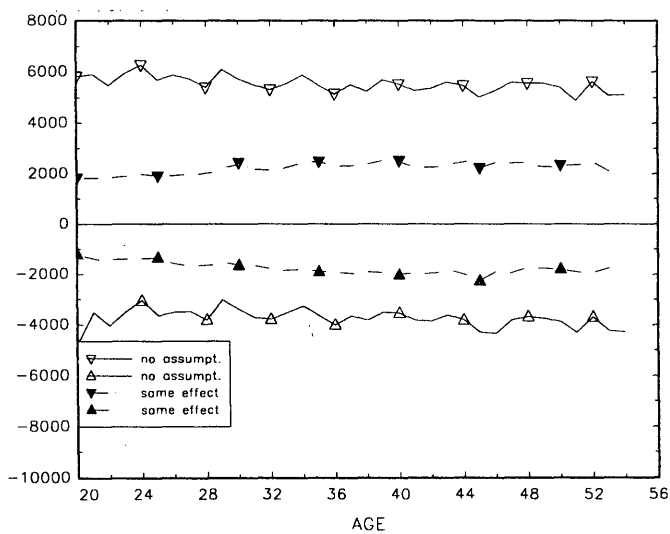


Figure C.3.d: Income in DM for women



Note: These are expected treatment effects for individuals randomly drawn from the population in both types of training (γ).

Figure C.4: Bounds for the treatment effects conditional on age for on-the-job training versus off-the-job training: rolling level-set restriction within narrow age groups for treatment effect or $E(Y^c|X=x)$

Figure C.4.a: Probabilities of not being unemployed in %-points for men

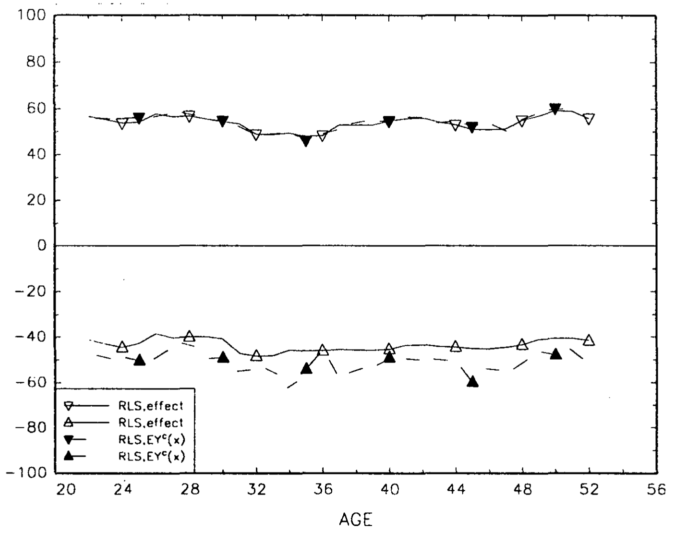


Figure C.4.b: Probabilities of not being unemployed in %-points for women

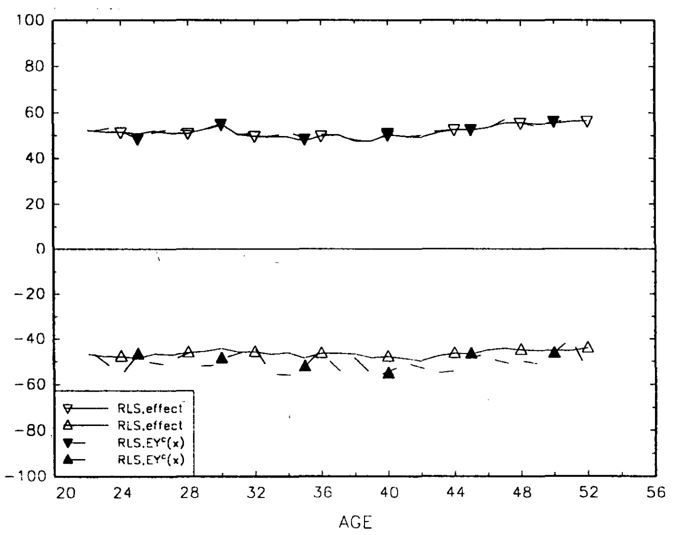


Figure C.4.c: Income in DM for men

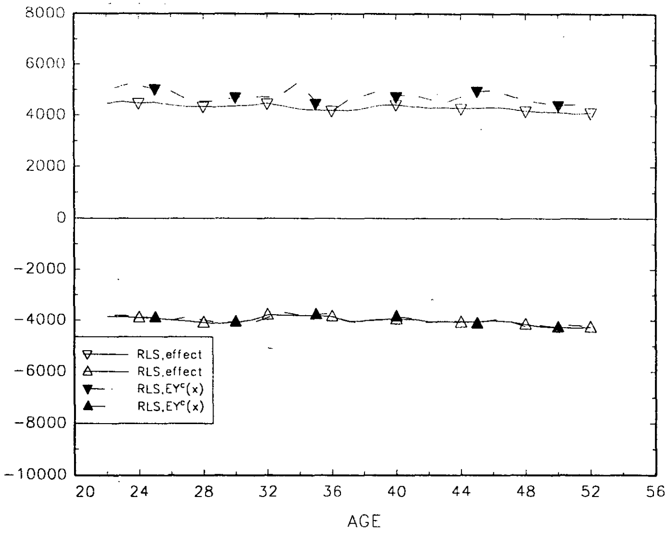
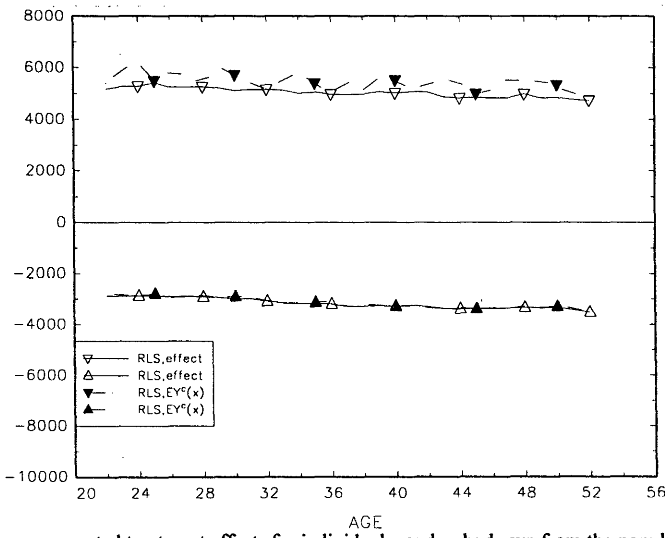


Figure C.4.d: Income in DM for women



Note: These are expected treatment effects for individuals randomly drawn from the population in both types of training (γ).Level-set restriction is for ±2 years.

Figure C.5: Bounds for the treatment effects conditional on age for on-the-job training versus off-the-job training: combining several restrictions

Figure C.5.a: Probabilities of not being unemployed in %-points for men

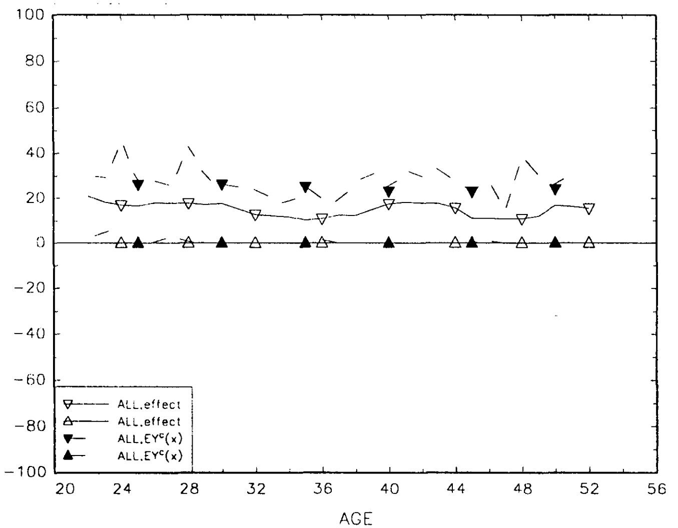


Figure C.5.b: Probabilities of not being unemployed in %-points for women

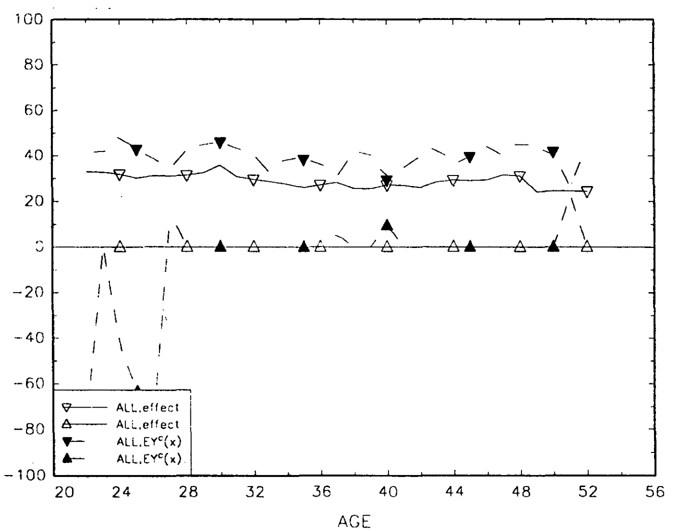


Figure C.5.c: Income in DM for men

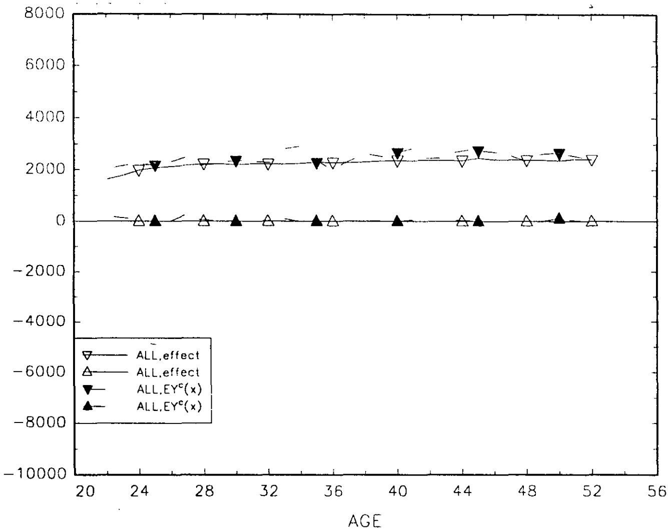
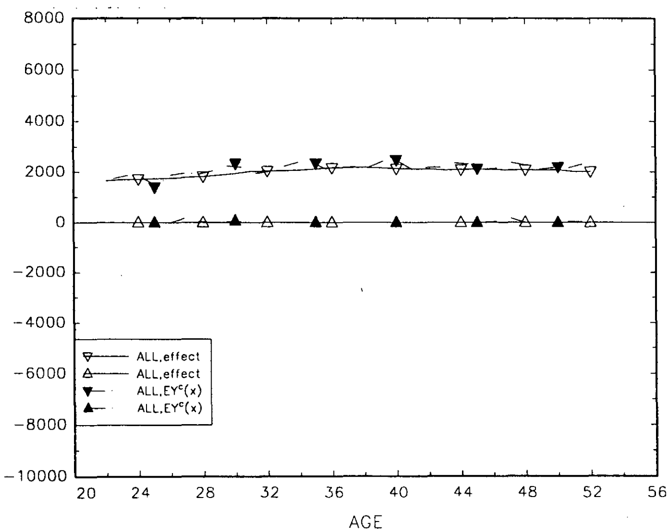


Figure C.5.d: Income in DM for women



Note: These are expected treatment effects for individuals randomly drawn from the population in both types of training (γ).

Appendix D: Additional results for on-the-job-training versus no training (women only)

D.1 Identified quantities ($p(x)$, $g^t(x)$, $g^c(x)$)

Table D.1: Estimates of $g^t(x)$ and $g^c(x)$ for on-the-job training versus no training conditional on schooling and federal state

	Probabilities of not being unemployed in %				Income in DM			
X-variables	$g^t(x)$		$g^c(x)$		$g^t(x)$		$g^c(x)$	
Years of schooling (highest degree)								
12	94.8	98.5	91.9	94.3	1914	2507	1671	2155
10	91.1	94.4	79.7	81.3	1509	1987	1133	1535
8 or no degree	85.0	94.7	68.6	71.8	1274	1836	900	1307
Federal states (Länder)								
Berlin (East)	95.5	99.5	84.0	88.0	1900	2501	1439	1922
Brandenburg	90.5	97.2	77.6	80.6	1553	2138	1153	1586
Mecklenburg-Vorpommern	93.5	98.8	75.1	79.1	1457	2044	1104	1555
Sachsen	88.9	94.2	78.4	80.6	1503	2029	1089	1500
Sachsen-Anhalt	84.8	92.7	78.0	81.1	1413	1973	1118	1552
Thüringen	91.5	97.9	77.0	80.3	1489	2107	1069	1501

Note: Table shows 5% and 95% quantile of respective bootstrap distributions. Most of the width of the income variable is due to the grouped nature of the income variable (see Table 1). Women only.

Figure D.1: Estimates of $g^t(x)$ and $g^c(x)$ conditional on age for on-the-job training versus no training

Figure D.1.a: Probabilities of not being unemployed in %-points

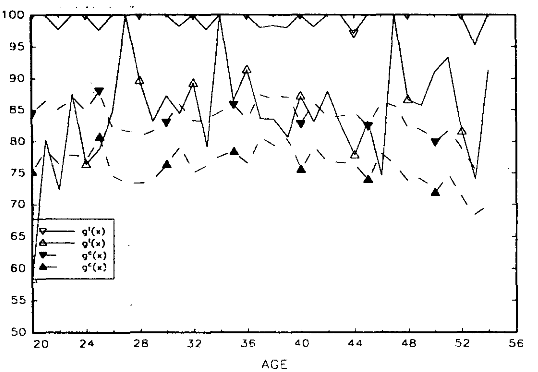
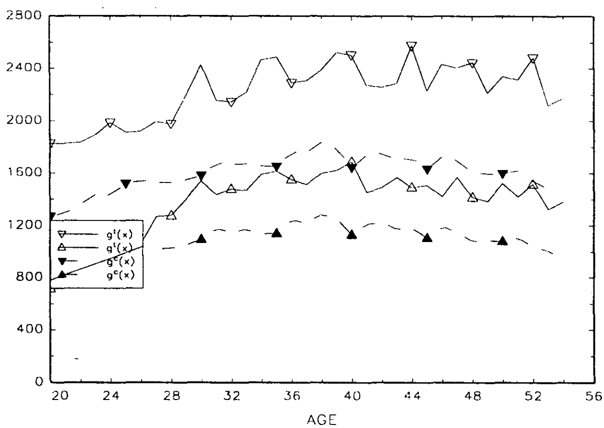


Figure D.1.b: Income in DM



Note: Figures show 5% and 95% quantile of respective bootstrap distributions. Women only.

D.2 Bounds

Table D.2: Bounds for the treatment effects conditional on training participation for on-the-job training versus no training

	Probabilities of not being unemployed in %				Income in DM			
Restrictions X-variables	none		same effect for treated and controls		none		same effect for treated and controls	
Federal states (Länder)								
Berlin (East)	-4.5	99.5	-4.5	16.0	-6099	2501	-1924	2501
Brandenburg	-9.5	97.2	-9.5	22.4	-6447	2138	-1587	2138
Mecklenburg-Vorpommern	-6.5	98.8	-6.5	24.8	-6542	2044	-1556	2044
Sachsen	-11.1	94.2	-11.1	21.6	-6496	2029	-1500	2029
Sachsen-Anhalt	-15.2	92.7	-15.2	22.0	-6586	1973	-1552	1973
Thüringen	-8.5	97.9	-8.5	23.0	-6510	2107	-1501	2107
Years of schooling (highest degree)								
12	-5.2	98.5	-5.2	8.1	-6085	2507	-2158	2507
10	-8.9	94.4	-8.9	20.3	-6490	1987	-1535	1987
8 or no degree	-15.0	94.7	-15.0	31.2	-6725	1836	-1308	1836

Note: Sampling uncertainty due to the estimation of $g^t(x)$ and $g^c(x)$ is accounted for by showing the 5% and 95% quantiles of the bootstrap sampling distribution of the lower repectively upper bounds of the intervals. Women only.

Figure D.3: Bounds for the treatment effects conditional on age and training participation for on-the-job training versus no training: no restriction, restriction of same expected treatment effects for treated and controls

Figure D.3.a: Probabilities of not being unemployed in %-points

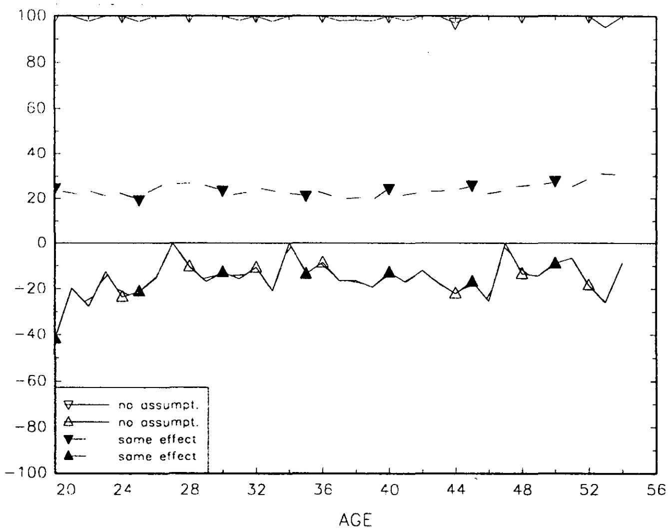


Figure D.3.b: Income in DM

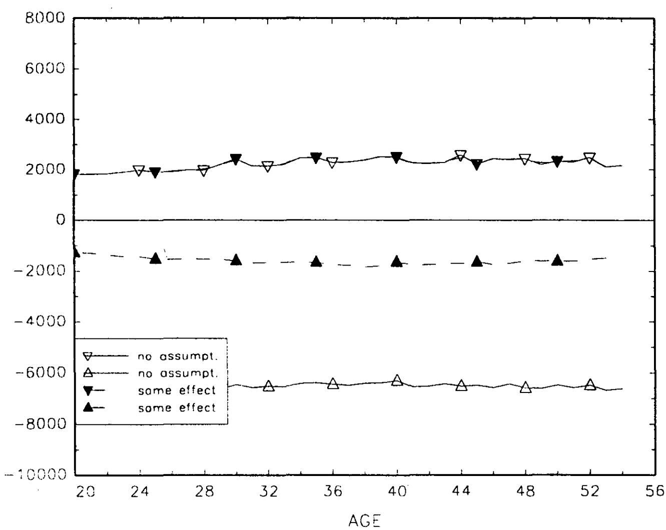


Figure D.4: Bounds for the treatment effects conditional on age and training participation for on-the-job training versus no training: rolling level-set restriction within narrow age groups for treatment effect or $E(Y^c|X=x)$

Figure D.4.a: Probabilities of not being unemployed in %-points

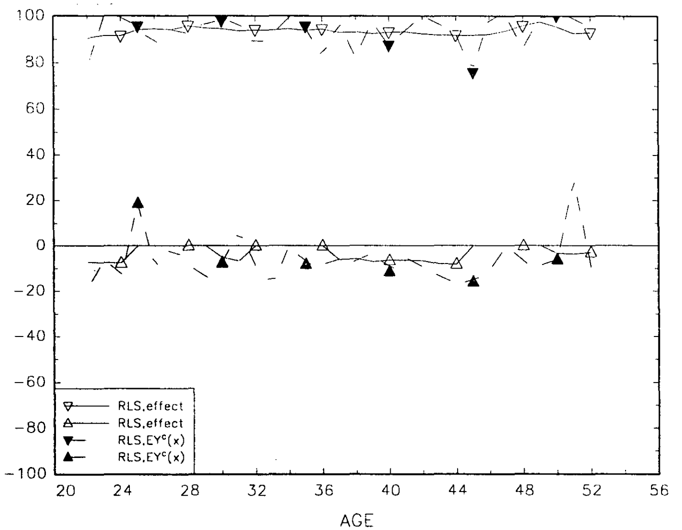
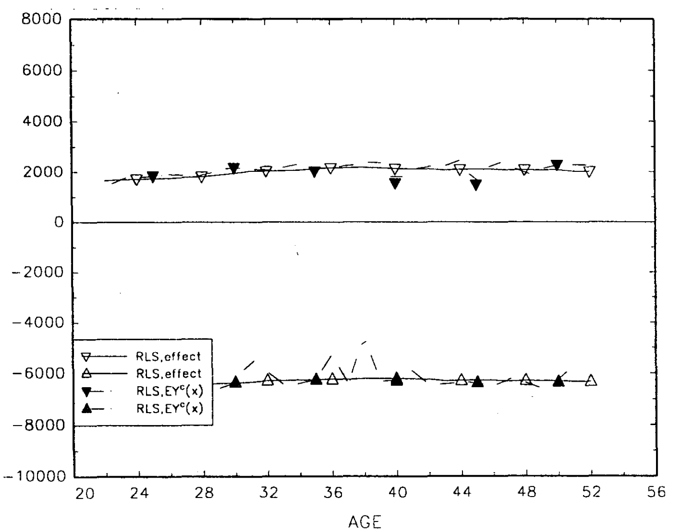


Figure D.4.b: Income in DM



Note: Level-set restriction is for ±2 years.

Figure D.5: Bounds for the treatment effects conditional on age and training participation for on-the-job training versus no training: combining several restrictions

Figure D.5.a: Probabilities of not being unemployed in %-points

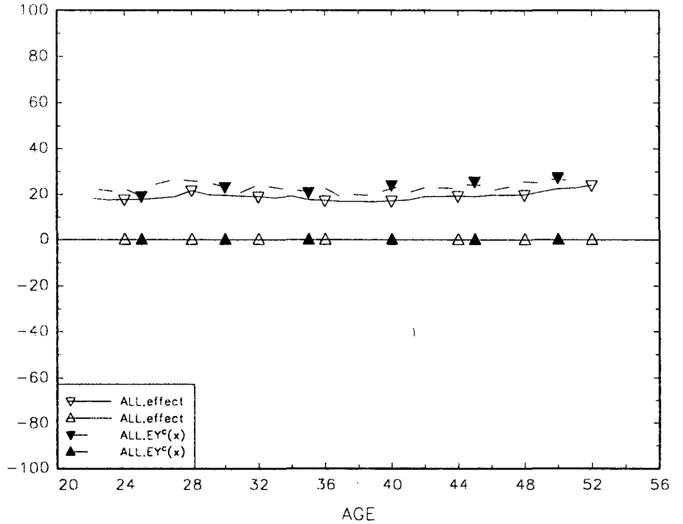


Figure D.5.b: Income in DM