published in The Econometrics Journal, 2, 1-28, 1999

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

Michael Lechner*
University of St. Gallen

Address for correspondence:
Michael Lechner
Swiss Institute for International Economics and Applied Economic Research (SIAW)
University of St. Gallen
Dufourstr. 48
CH-9000 St. Gallen
Switzerland
Email: Michael.Lechner@unisg.ch
WWW: http://www.siaw.unisg.ch /lechner

* Financial support from the Deutsche Forschungsgemeinschaft (DFG) and the Swiss National Science Foundation (SNF) is gratefully acknowledged. I thank the ZEW (Zentrum für Europäische Wirtschaftsforschung, Mannheim, Germany) and the Statistisches Bundesamt (Wiesbaden, Germany) 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. The paper has been presented at EC2 at Florence 1996, ESEM 1997 at Toulouse, and the Conference on the Application of Semiparametric Methods at Tilburg, 1997. I thank participants of all three conferences, as well as Bernd Fitzenberger, Gerd Ronning, and two anonymous referees of this journal, as well as Neil Shephard, for valuable comments and suggestions. All remaining errors are my own.

Abstract

This paper explores the potential of an approach suggested by Manski to obtain non-parametric bounds of treatment effects in evaluation studies without knowledge of the participation process. The practical concern is the estimation of 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 able to bound 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 extent 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, 1997, Lechner, 1996a, 1996b, 1999) give ambiguous answers on 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 where no social experiment is available. The fundamental problem is the necessity to infer enough from the labour market outcomes of those not participating in training about the outcomes of trainees that would have occurred 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).

If 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 and avoid selection bias. In practice, for many cases complete information on this mechanism is not available. Therefore more or less serious doubts about the validity of the chosen identifying assumptions plague almost every empirical evaluation study based on non-experimental data.

Hence in this paper the idea is to 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 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 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, 1993), and for treatment effects (Manski, 1990). Manski, Sandefur, McLanahan, and Powers (1992) provide an application of this methodology. In this paper, the econometric results are applied to the evaluation of the effects of vari-

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

ous types of training on individual income and employment probabilities. The estimation is based on a very large cross-section that constitutes approximately a 0.6% sample of the East German population in 1993. 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 respect. 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 emphasised: 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. 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 causality concept used in evaluation studies and discusses the identification problem of various treatment effects. Section 3 describes the sample used in the empirical part. Section 4 discusses the estimation and presents the result of the quantities that are identified and that are the ingredients of computing the bounds. Section 5 contains the derivation of bounds for treatment effects with minimal assumptions together with the empirical estimates of such bounds. Section 6 explores the potential for tightening the bounds by introducing additional restrictions. Section 7 draws conclusions. The Appendix contains technical details on the bounds as well as some extensions.

2 Causal effects of training

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 non-participation. 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 has an effect on 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 income for example, but whether the effect of training is higher income.[2] Given a large enough and sufficiently informative sample from the population, the question of association can easily be answered whereas the question of the causal relation raises serious identification issues.

The framework that will serve as a 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 the control state, i.e. no treatment).[3] 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 at the fundamental problem of causal analysis. The individual causal effect, for example defined as the difference of the two potential outcomes ($y_n^t - y_n^c$), 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 [$y_n = y_n^t s_n + y_n^c (1 - s_n)$] is not observable.

Note that at this stage it does not matter at all for the definition of causality whether the assignment to the different states is controlled by the econometrician, like in a social experiment, or whether it is just observed in the data. This literature however emphasises that *S* should be changeable – like an economic policy variable for example – and not be a fixed individual or personal attribute, like gender or race.

Another concept of causality introduced by Granger (1969) receives considerable attention in econometrics. Instead of focusing on the comparison of hypothetical outcomes, Granger proposes to uses the predictability of one variable by another to analyse causal relations. Whereas the former approach could be termed the experimental approach to causality, the latter may be termed the predictive approach to causality. Although Rosenbaum (1986) showed that both approaches coincide in a very specific setting, it appears that they are nevertheless fairly different concepts of causality. Reading the exchange between Granger and Holland contained in Holland (1986), many readers remain puzzled about the connection of those approaches. One simple but not really satisfying insight is that the hypothetical outcome approach is certainly easier to apply in a cross-section framework than when dynamics are important, whereas Granger's predictive approach is well suited for dynamic problems but may be much closer to statistical association instead of causation than the hypothetical outcome approach.

---

2 See Holland (1986) for an extensive discussion of concepts of causality in statistics, econometrics, and other fields.

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

It has been mentioned before that the individual causal effect defined as $(y_n^t - y_n^c)$ cannot be identified. Nevertheless, certain population averages of causal effects are in principle identifiable. Three such average causal effects of training for individuals with characteristic *x* are denoted by $\gamma^0(x)$, $\theta^0(x)$, $\xi^0(x)$, and defined in equations (1), (2), and (3):

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

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

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

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 non-participants with characteristics *x*. Whenever, there are more individual characteristics that could be used to describe the population, those different averages may not be equal.

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

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

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

$$\xi^0(x) = E(Y^t \mid X = x, S = 0) - g^c(x)\,. \qquad (3')$$

A general question is how these expressions can be identified from a large random sample of the population.[4] The quantities $g^t(x) := E(Y^t | X = x, S = 1)$, $g^c(x) := E(Y^c \mid X = x, S = 0)$ 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. Hence, 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

[4] See the following section for a description of the actual sample used in this paper.

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 (like in a social experiment where only the characteristics x influence the selection probability), 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. To make such an assumption plausible 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 estimation could proceed as if the data were generated by a true experiment. Obviously, the major untestable assumption is that all attributes are really 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 for example by Lechner (1999) for the evaluation of the effects of East German training.

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, 1997). 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).[5] 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).

In practise there may be many cases where none of the possible identifying strategies appears to be plausible or feasible. Under these circumstances no point estimates of the various average causal effects could be obtained. In the remainder part of the paper we will discuss what could still be said without such identifying assumptions in our specific application. However, before ways to obtain bounds of these unidentified quantities are discussed, it is useful to see what kind of data is available to conduct the empirical analysis.

3 Data

3.1 The German microcensus

The data used for the empirical part of the paper is the German microcensus. The *microcensus* is an important component of the official German statistics system. The federal statistical

[5] 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.

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 non-response. It is a representative sample containing information on socio-demographics, as well as on variables related to the individual labour market situation. Most of the information refers 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.

Some general remarks about the data 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 somewhat different. Clearly, the German microcensus is in no way informative enough to control for selection issues (and hence to obtain consistent point of the treatment effects). Since the issue here is not necessarily to obtain point estimates but only to bound the effects, this problem may not be so acute. Quite to the contrary, when estimating and interpreting the bounds it is very important to get a very precise estimate, otherwise they might become too large to contain useful information about the effects. The second reason why a very large sample is important for this exercise is that the nonparametric nature of the bounds requires nonparametric methods to estimate them (see below). It is well known that nonparametric methods require a substantial sample size to perform acceptably.

Due to data confidentiality reasons originating in German federal laws, the original sample of the German *microcensus* 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*.

While this study was done, only the 1991 and 1993 surveys were 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.

3.2 Sample selection, definition of training, and descriptive statistics

The empirical analysis is based on prime-age individuals not yet subject to early retirements. Therefore, the sample is restricted to individuals aged between 20 and 54. Since the training for newly immigrated foreigners (as the major share of foreigners in East Germany is) has fairly different objectives than training of 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.[6]

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. 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 (OFFJ), and other off-the-job-training (OFFJO). Only completed training is considered. Individuals still participating in some sort of training are not considered. ONJ is close to employer-related training that is analysed with GSOEP data for example in Lechner (1996b). Unfortunately, it is not known whether the individuals obtained benefits from the labour office during the training. Hence the distinction between general off-the-job training (Lechner, 1999) and public-sector sponsored training (Lechner, 1996a) is impossible with this sample. However, it is reasonable to conjecture that most of the public-sector sponsored training is indeed included in OFFJ because it is usually done in these school-like training institutions.

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

Table 1: Descriptive statistics

	no training *control*	on-the-job *ONJ*	off-the-job *OFFJ*	other off-the-job *OFFJO*
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
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.***

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 considerably shorter than for OFFJ and OFFJO (medians about 6 months). The larger proportion of females in OFFJ and the fact that OFFJ participants experience higher unemployment than the rest also points at a close relationship between OFFJ and public-sector-sponsored training. Not surprisingly, the post-training unemployment rate for ONJ is rather low.[7] 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 income.

3.3 Targets of the evaluations and conditioning variables

Choices have to be made regarding the targets of the evaluations. Following the training literature, I chose unemployment[8] and net income to measure the effects of training. The only information available on these variables refers to the date of the interview. Hence, the distance to the end of the training differs from one individual to another. 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 computation of 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 of the treatment effects (see below). 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. Here however the problem arises that there is no upper bound for the income variable.[9] However, it will be seen in the next section that all what is needed is not an upper bound for the support of the random variable, but of its expectation. The assumption that the respective expectations are no larger than DM 8000 is used in the following application. This is a very conservative number given that only about 0.1% of the sample is observed with an income of DM 7500 or more.

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

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

[9] It will become clear in the next section that this is really a problem for estimating the bounds.

The second possibly problematic feature of income as defined by the *microcensus* is that it 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 are 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 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. The definition of a conditional treatment effect makes only 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.[10] 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.

4 Estimation method and results for the identified quantities

4.1 Estimation method

Estimation of the bounds is done in a stepwise procedure. In this section only the very first step, ie. estimation of the identified quantities is discussed. Since these quantities are conditional expectations ($E(Y^t|X=x,S=1)$ $[=g^t(x)]$, $E(Y^c|X=x,S=0)$ $[=g^c(x)]$, and $E(S|X=x)$ $[=p(x)]$), they are estimated in a straightforward manner by 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: On the one hand, one could use a parametric model to estimate these conditional expectations as in standard linear or binary choice regression. This however is not attractive, because it is somewhat contrary to fundamental ideas of estimating treatment effects without incurring inconsistencies by incorrect assumptions of functional forms. Hence, nonparametric estimation of the bounds appears to be imperative. On the other hand, some smoothing methods (kernels, nearest neighbour, series estimation, ...) could be used in principle to compute the bounds to improve efficiency. They are however not attractive here because of the discreteness of all X-variables.

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

Although the sample means are asymptotically normally distributed, it will be shown in the next section that there is no such simple asymptotic distribution theory for the estimated bounds. Therefore, a bootstrap estimation of the sampling distribution of the respective estimators is one possible solution. To use a unified way to report results in this paper, section 4.3 reports the results for the identified quantities not based on first order asymptotics, but also based on results of the following bootstrap procedure that follows the approach by Manski et al. (1992): (i) Draw 500 bootstrap samples from the *microcensus* (samples of same size as original sample; observations drawn with replacement); (ii) compute all relevant means for each bootstrap sample; (iii) report the 5% and the 95% quantile of the bootstrap distribution of the estimates as an estimate for centred approximate 90% centred confidence intervals.[11] For the income variable the 5%-quantile of the bootstrap distribution based on the means of the lower bounds of the income categories and the 95% quantile of the bootstrap distribution based on the means of the upper bounds are reported.[12]

4.2 Choice of treatment effects and conditioning sets

As the results may differ for different treatments 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 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. I am concentrating on pair-wise comparisons of the different groups conditional on being in one or the other group or conditional on being in either of the two groups under consideration.[13] The results presented in the paper cover an even more special case, namely the effects of on-the-job training vs. no training for the participants in on-the-job training. Admittedly, the focus on on-the-job training is rather arbitrary. However, it is 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, 1999) are reasonably well identified, whereas the identification of on-the-job (employer-related) training appeared to be much more doubtful because of a lack of sufficient information about the respective workplaces and employers (Lechner, 1996b).[14]

---

[11] 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.

[12] 95% bootstrap intervals are computed as well for all variables, 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 available on request from the author.

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

[14] The corresponding results for off-the-job training versus no training, and off-the-job training vs. on-the-job training can be downloaded from my website.

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

4.3 Results for $p(x)$, $g^t(x)$, and $g^c(x)$

Table 2 contains the estimates of the participation probabilities [$p(x)$] of on-the-job training versus no training at all conditional on schooling and federal states ("Länder") for men and women separately. Let us first consider the heterogeneity in the data. Heterogeneity is significant with respect to schooling.[15] 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.

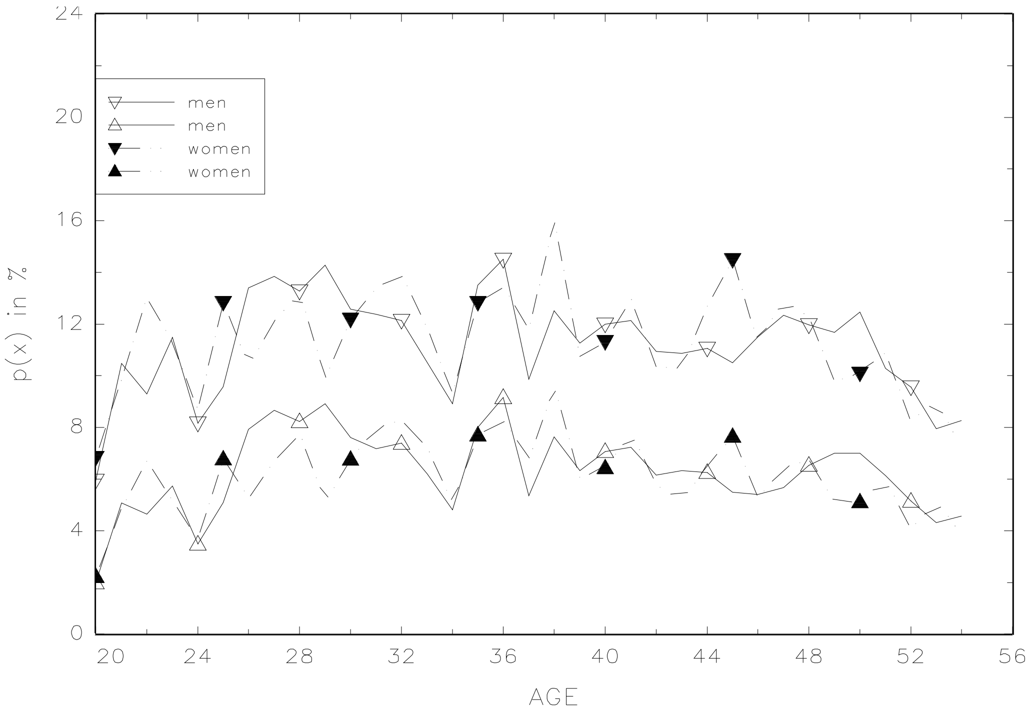
Table 2: Probabilities for on-the-job training versus no training in %

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.

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

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



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

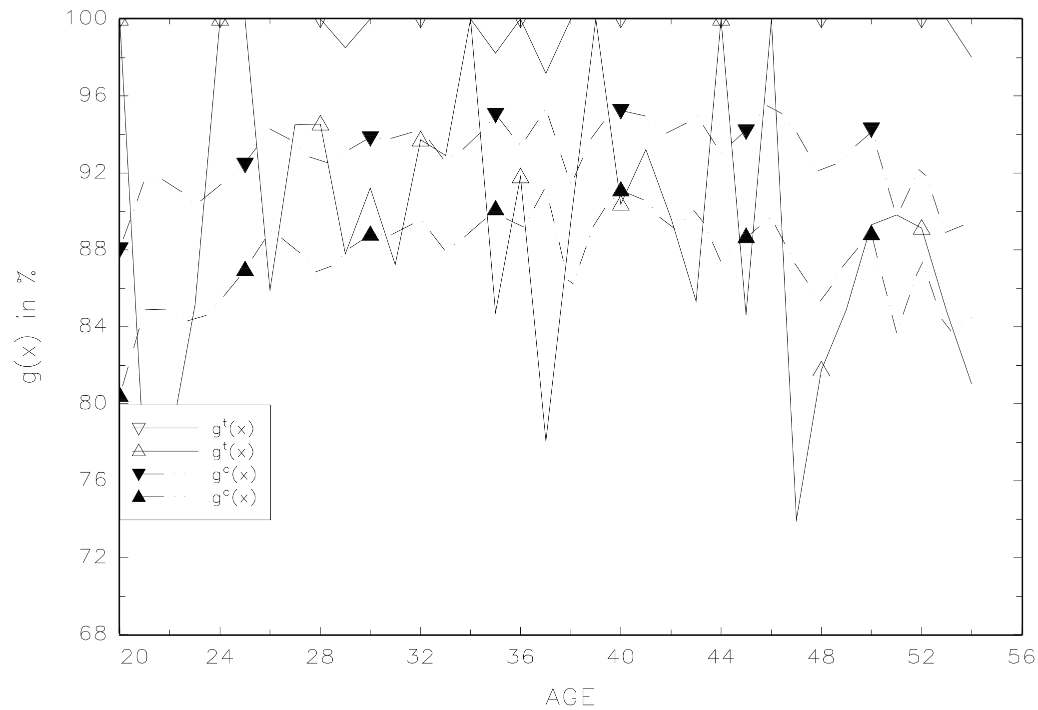
Table 3: Estimates of $g^t(x)$ and $g^c(x)$ conditional on schooling and federal state for on-the-job training versus no training

	Probability 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.

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. The results for women look remarkably similar to those for men.[16]

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



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

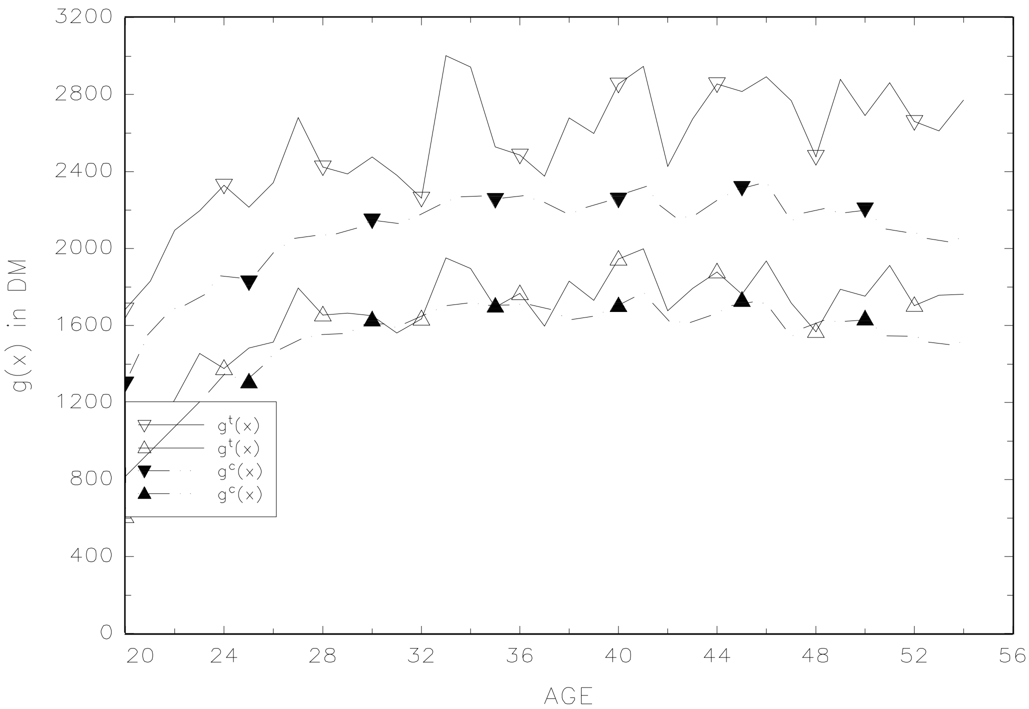
The lower part of Table 3 and Figure 3 focus on *income*, which is a grouped variable.[17] Table 3 shows that income increases with schooling and that there is some regional variation as well, with people living in the former capital and administrative centre of the GDR having a higher mean income. Income increases with age for those younger than about 30. The differences between mean incomes of the treated population appear to be higher than of the controls. Interestingly, this difference is much more pronounced for women than for men.

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.

[16] All results for women that are not presented in the paper can be downloaded from my website.

[17] 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.

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.***

5 Bounding treatment effects

5.1 Theoretical considerations

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, 1993), but also for models of potential outcomes (Manski, 1990). Manski et al. (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 equation (1), but his analysis is easily extended to the case given in equation (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.[18] Since the considerations for the

[18] Note that the fact that there are more than two treatments to which an individual can be assigned to is ignored in the definition of the treatment effects. Although this is not a problem for $\theta^0(x)$, it is a problem for the interpretation of $\gamma^0(x)$. As defined above, it refers to a person with characteristics *x* randomly drawn from the population of persons with characteristics *x*. However, as used here in these pair-wise comparisons, the

treatment effects for the nontreated population based on equation (3) are symmetric to the case given in equation (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, L^c and U^t, U^c the lower and upper bounds of the support of Y^t and Y^c.[19] 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. Without any loss of generality, the treatment effects are defined as 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 $B_\gamma^L(x) = B_\theta^L(x) = L^t - U^c$, $B_\gamma^U(x) = B_\theta^U(x) = U^t - L^c$. Hence, the width is given by $W_\gamma(x) = W_\theta(x) = \{U^t - L^t\} + \{U^c - L^c\}$.

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.[20] Therefore, the following bounds are obtained for $\gamma^0(x)$: $B_\gamma^L(x) = \{g^t(x) - U^c\} p(x) + \{L^t - g^c(x)\} \{1 - p(x)\}$, $B_\gamma^U(x) = \{g^t(x) - L^c\} p(x) + \{U^t - g^c(x)\}\{1 - p(x)\}$. The now reduced width is given by $W_\gamma(x) = U^c - L^c$. The respective bounds for $\theta^0(x)$ are given by $B_\theta^L(x) = g^t(x) - U^c$ and $B_\theta^U(x) = g^t(x) - L^c$. Again, the width simplifies to $W_\theta(x) = U^c - L^c$. To appreciate the reduction of the widths assume that the lower, respectively upper bounds of the support of the different 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

population from which this person is randomly drawn consists only of the participants in the particular treatment plus persons choosing the particular alternative considered (here: participants in on-the-job-training plus persons without continuous training in the last two years). Deriving the bounds of $\gamma^0(x)$ for the total population would involve also choice probabilities for the two other types of treatments (training). Although interesting, this is not persued here any further because of lack of space. Therefore, the results section focuses on $\theta^0(x)$ only. I thank a referee of this journal for bringing this issue to my attention.

19 In the Appendix, a more general case in which the supports are allowed to vary with *X* and *S* is considered.

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

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.

One very simple way to make the information about the controls useful, is to assume that conditional on *X*, control and treatment population have the same expected effect of participating in training.[21] It is not clear whether in the case of ONJ this assumption is tenable. It implies that firms chose their trainees without effectively selecting those who can use the training more effectively than others. Although this is probably not true in general, it is possible when firms do not have enough information or, when acquiring that information is too costly, or else when there are other rules for obtaining ONJ (e.g. work contract could give employees a right to obtain specific training in relation age). However, in this data set there is no way to find out why the specific participant obtained ONJ.

With this restriction of same treatment effects for treated and controls, the interval for the expected treatment effect is the intersection of the previously defined intervals for $\theta^0(x)$ and $\xi^0(x)$. Therefore, this restriction can shrink the bounds, but cannot exclude a zero treatment effect.[22]

5.2 Estimation methods

In the previous section (4.1) it is noted that $p(x)$, $g^t(x)$, and $g^c(x)$ are estimated for each bootstrap sample by the appropriate weighted sample means in the *X*-cells. These estimates are now used to construct the appropriate intervals and the corresponding bounds for each bootstrap sample.[23] The numbers reported are the 5%-quantile of the bootstrap distribution of the lower bound and the 95%-distribution of the bootstrap distribution of the upper bound. Since the upper and lower bounds are based on the same estimated intervals, the reported boundaries are correlated. The corresponding probability statement is that the probability that the true interval will not be within these reported boundaries is approximately 10%. Note that this does not imply a similar probability statement for the expected treatment effects. Indeed the probability that the true treatment effect is outside the reported boundaries must be (much)

---

21 Note that this is not the case of random assignment like in a (ideal) social experiment, because from random assignment follows the more powerful property that not only the expected treatment effects are the same, but that the expected conditional outcomes are the same. The former is insufficient to identify the treatment effect, whereas the latter identifies the treatment effect.

22 The typical focus on excluding zero effects could be a too extreme view because if there are costs of training it would be very informative to know that the effect of training is no larger than a small positive amount determined by the costs. This would be enough to conclude that training is ineffective. However, since there is no information about costs in this sample, this approach is not persued any further.

23 For each bound the income variable is used such as to make conservative statements.

lower. Furthermore, note also that from the reported boundaries we cannot infer a probability statements about the length of the intervals (as in the example above, in many cases they are even nonstochastic and thus smaller).

5.3 Results

Obviously, three different treatment effects for all the X-cells again are too many to be presented. Focusing on the treatment effects for the population of respective training participants appears to be most interesting, because this is the appropriate measure of the effectiveness of the performed training (i.e. the effect of the treatment on the treated). Again for reasons of space, most of the results for the presented bounds are for men only. The focus on men is due to the fact that in contrast to the other types of training, men are the quantitatively more import group for East German ONJ.[24]

Table 4: Bounds of 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

	Probability 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	
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
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

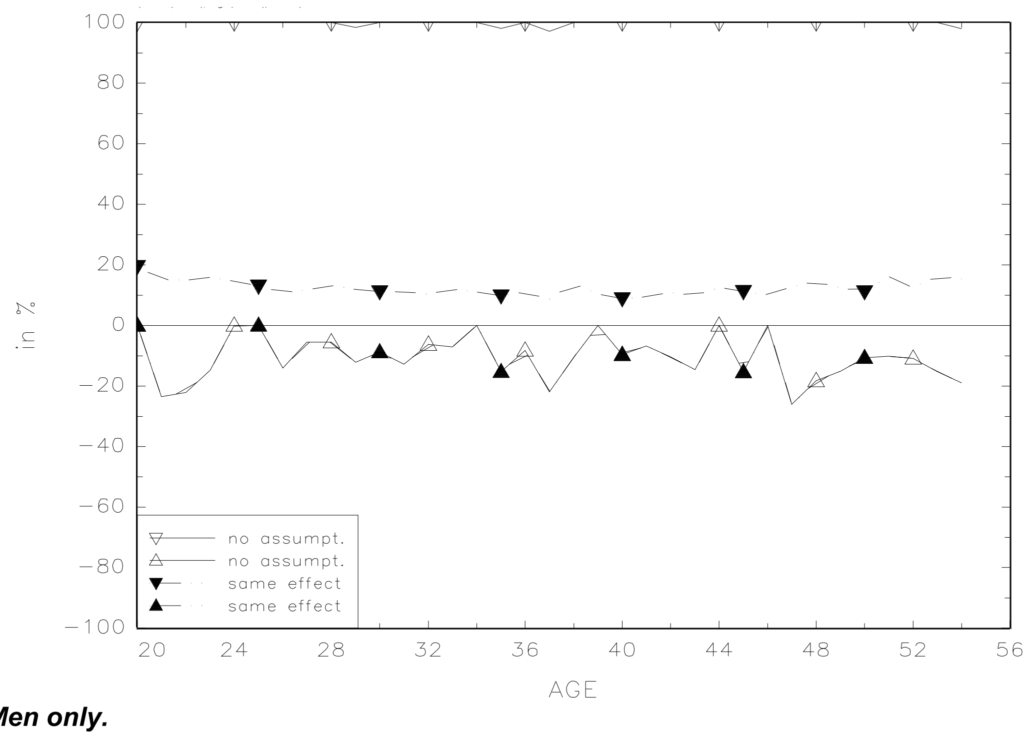
Note: Sampling uncertainty due to the estimation of $g^t(x)$ and $g^c(x)$ is accounted for by showing the 5% and 95% quantiles of the bootstrap sampling distribution of the lower respectively upper bounds of the intervals. Men only.

Let us start by considering the bounds on the treatment effects given only the sample estimates without any assumption. 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

[24] The corresponding results for women can be downloaded from my website.

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, appears to be very powerful.[25] It reduces the width dramatically, but as mentioned above, this assumption is not sufficient to identify the sign of the treatment effect.

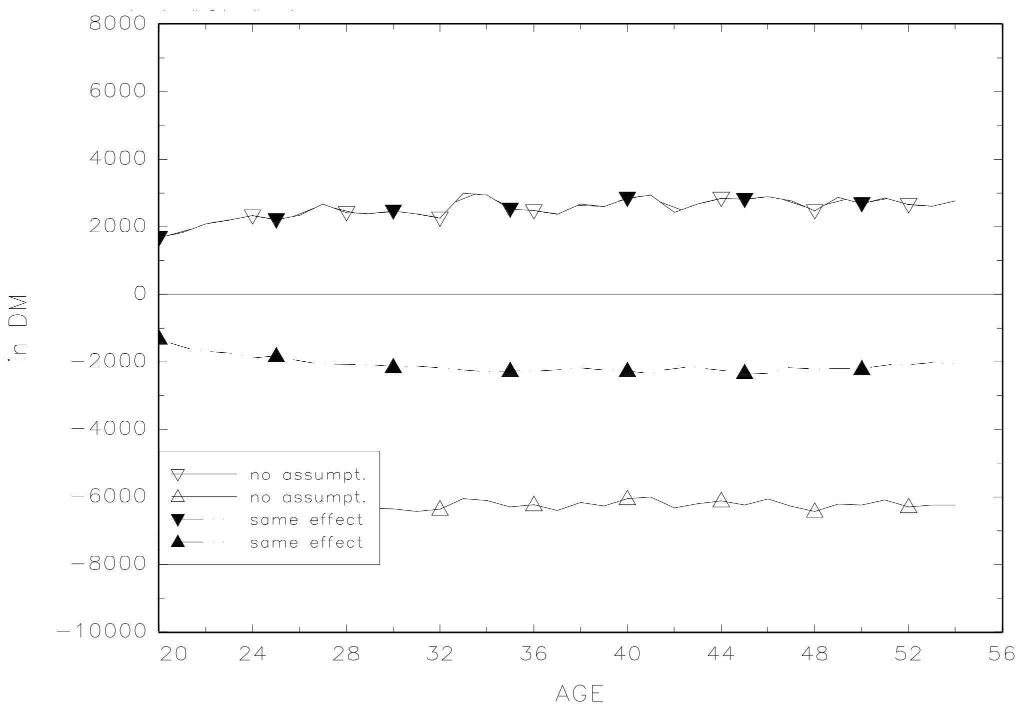
Figure 4: Bounds of the treatment effects conditional on age and training participation in % (not being unemployed) for on-the-job training versus no training: no restrictions, same expected treatment effects for treated and controls



Note: Men only.

[25] 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 5: Bounds of 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.

6 Further shrinkage of the intervals

6.1 Level set restrictions

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 ($\chi^0 \subseteq \chi$).[26]

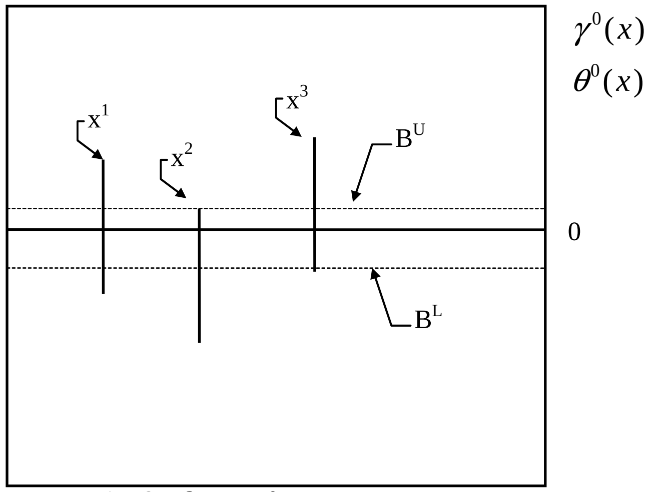
Let us be begin with a level-set restriction on the treatment effect.[27] Figure 6 shows graphically how a reduction of the widths is achieved. For illustrative purposes, it is assumed that *x* is one-dimensional and that the treatment effect is identical for three different values of *x*. Clearly, the true treatment effect must be included in all three intervals. The larger the space χ^0 and the more variable the bounds are within the space, the larger is the gain in width

[26] 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.

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

reduction. However, at least for the case of equal bounds on both outcomes, the reduction of width will be insufficient to exclude zero treatment effects. These level-set restrictions are untestable. The plausibility of them will depend on the particular application.[28] The exact formulas for the bounds and the widths can be found in the Tables of the Appendix. Popular examples of such an assumption are models with an additive treatment effect, such as: $E(Y|X=x,S=s)= f(x)+\alpha s$. 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.

Figure 6: 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 discussion of locally 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 of the bounds and the widths are given in the Appendix. Under this assumption the treatment effect is constant on χ^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 the set χ^0, the bounds are at least as tight as those derived under the assumption of a locally constant treatment effect. Therefore, it is more likely to identify the sign of the effect. In the

[28] See the next section for a the empirical implementation and a discussion whether this implementation might be realistic for the case of ONJ.

case of bounding $\theta^0(x)$ it is sufficient to assume that $E(Y^c|X=x)$ is constant in $\chi^{0,c}$ (=χ^0) because $\theta^0(x)$ does not depend on $E(Y^t|X=x,S=0)$.[29]

Another restriction that is explored in the application is the assumption that all individuals selected into training have non-negative expected values for $\theta^0(x)$. Clearly, this tightens the lower bounds.

6.2 Implementation of the level set restrictions

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.[30] RLSs are imposed on the treatment effects as well as on $E(Y^c|age)$. The underlying idea in this application is that the exact age should not matter as long as two individuals belong to the same narrowly defined age group ($\pm$ 2 years of age).

Additionally, it is assumed that only individuals with expected nonnegative gains from the treatment participate. At first sight it seems that a minimum condition of training participation should be positive expected returns for participants. This implies however that, prior to training, the amount of information available is considerable because in reality the true expected effect of training is unobservable even for the firm and the individual. For ONJ it requires the firm to know enough about the participant (and the future in general) to be able to estimate his future productivity with and without ONJ accurately enough. For example, systematic surprises occurring to all participants could invalidate these pre-training considerations. Such a surprise might be a negative productivity (or positive technology) shock that makes the new skills redundant (quite plausible for East Germany after unification). On the other hand, this restriction could be weaker than necessary, because firms (and participants) will not only require a positive return from training but also a return that is larger than their costs.

Finally, the different restrictions are combined to additionally tighten the bounds.

6.3 Results

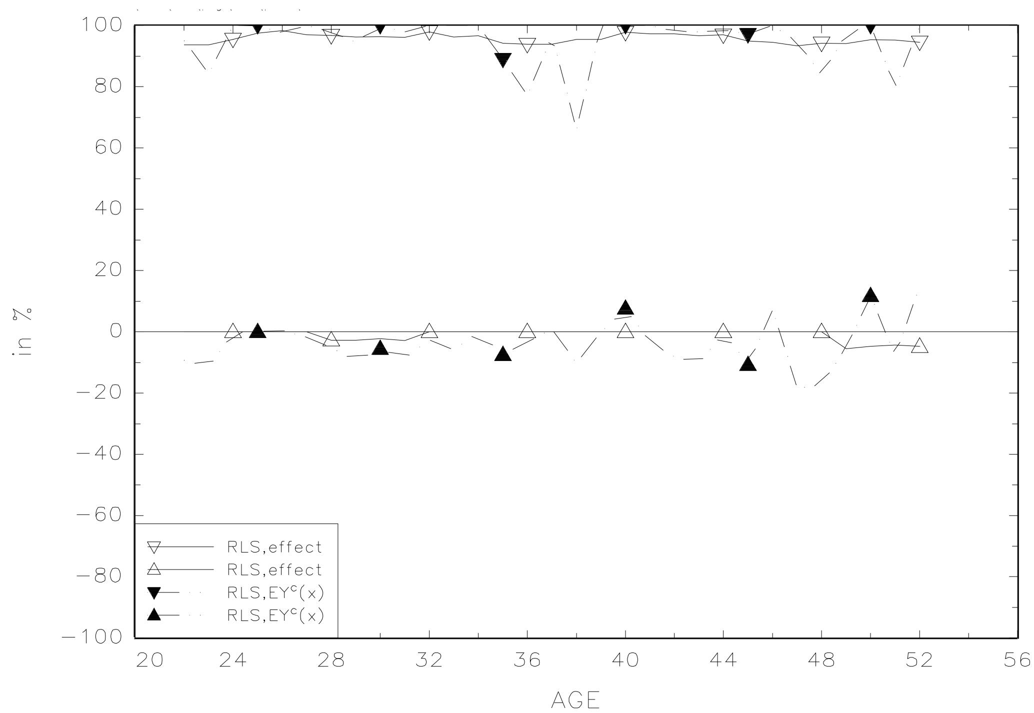
Figures 7 and 8 present the results when imposing rolling level-set restriction within narrow age groups (± 2 years) on the treatment effects or on $E(Y^c|X=x)$. The corresponding lines

29 Additional restrictions that are however not used in the empirical parts are discussed in the Appendix.

30 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.

in these figures are denoted as *RLS, effect*, and *RLS*, $EY^c(x)$, respectively. Figure 5 already showed 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 effects 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 7): For some ages, the intervals lie entirely on the positive part of the support of the treatment effect. Thus, for men aged 40 and 50, there is a positive effect of training if this exclusion restriction is true. The same is true for women aged 24, 30, and 50. Note that this result depends solely on the assumption that, for example $E(Y^c | men, age = 38)$ = $E(Y^c | m,39)$ = $E(Y^c | m,40)$ = $E(Y^c | m,41)$ = $E(Y^c | m,42)$ (example for the effect for men aged 40). Still, this remains a rather technical condition, and it appears to be difficult to decide whether this (untestable) condition, though plausible, is true.[31] Additionally, it is unclear why there are only positive effects for specific ages that look rather arbitrary.

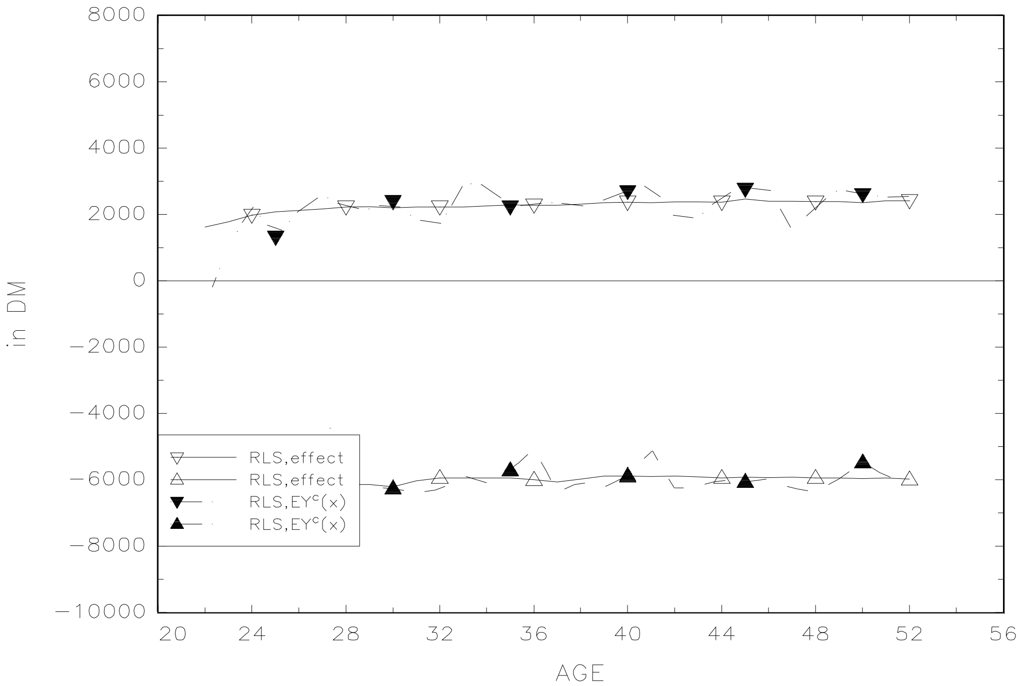
Figure 7: Bounds of the treatment effects conditional on age and training participation in %-points (not being unemployed) for on-the-job training versus no training: rolling level-set restriction within narrow age groups for treatment effect or $E(Y^c | X = x)$



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

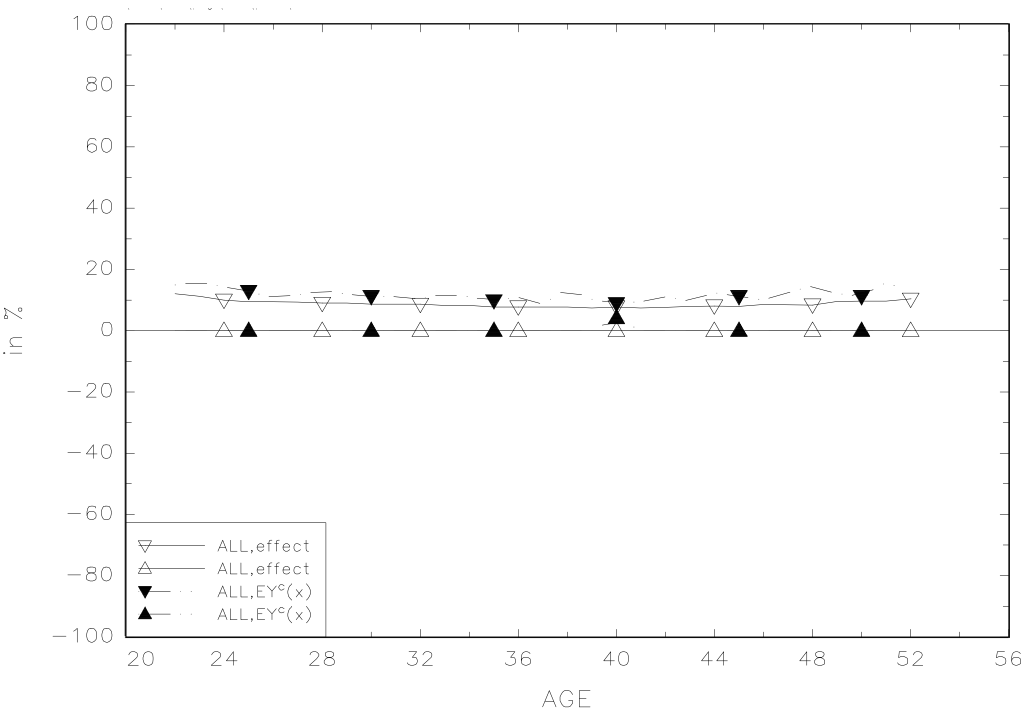
[31] The only evidence against this assumption would be a resulting empty set for the treatment effect. Given the large size of the intervals, this test has extremely low power in the current setting.

Figure 8: Bounds of 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 $E(Y^c | X = x)$



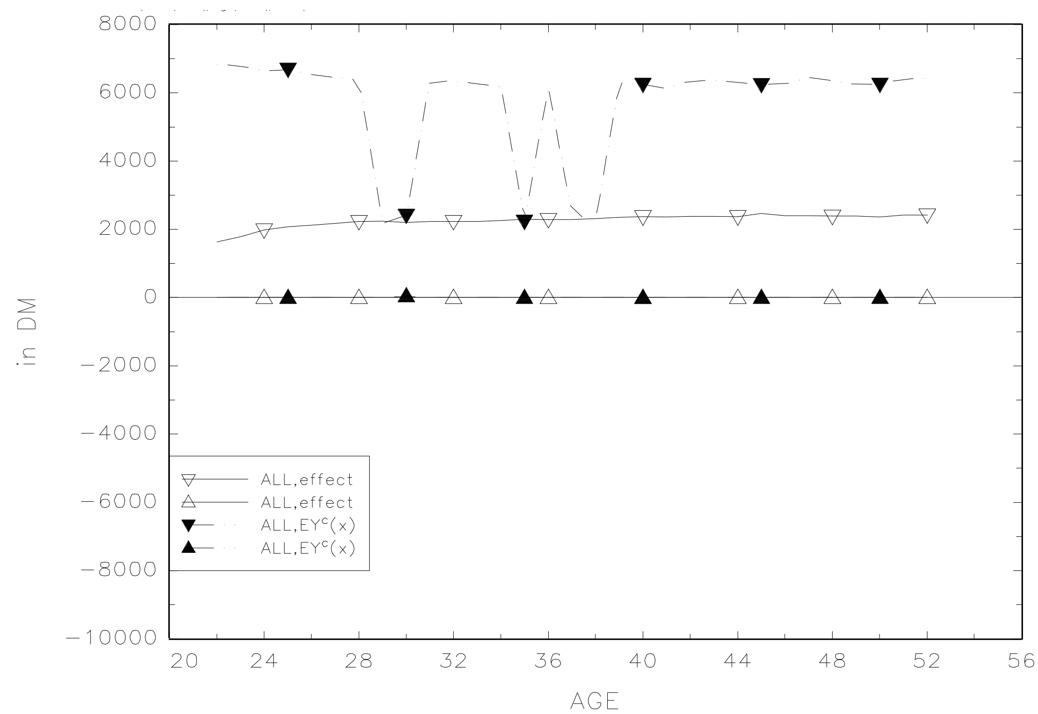
Note: Men only. Level-set restriction is for ± 2 ***years.***

Figure 9: Bounds of the treatment effects conditional on age and training participation in %-points (not being unemployed) for on-the-job training versus no training: combining several restrictions



Note: Men only.

Figure 10: Bounds of 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 9 and 10 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 9), they collapse for some age groups to almost a single point.[32] For example, the effect for 40 year old men is almost exactly a 10%-point increase in the probability of not being unemployed. It is quite remarkable that such a precise estimate could be obtained with this approach. For the income variable, the width of about DM 2000 is still substantial (Figure 10) and a zero effect could not be excluded. This is at least partly due to income being observed only in categories. This additional uncertainty widens the interval too much to allow any policy relevant conclusions.

When considering the other types of training (results available from my website), similar patterns appear. For the case of off-the-job training vs. no training there are positive income effects for males of age 36 and 50 and for women of age 52 with RLS on $E(Y^c | X = x)$. The width of the interval of the effect on income is also smaller than in the case of on-the-job

[32] 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.

training, but still too large to be really informative. The direct comparison of on-the-job training and off-the-job training is based on the expected treatment effects for an individual drawn randomly from the population of on-the-job and off-the-job participants ($\gamma(x)$ instead of $\theta(x)$) because here the most interesting question is whether ONJ or OFFJ is more effective for the same group of people. The results show that only the combination of all assumptions leads to positive effects for the probability of not being unemployed (for men aged 24 and 26, and women aged 40 and 50).

7 Conclusion and outlook

This paper is one of the 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. Nevertheless, in some cases suitable exclusion restrictions are indeed capable of bounding the treatment effects away from zero. However, in most, but not all, cases they still allow a wide range of possible values for the treatment effects. Therefore, the use of bounds as tests for predictions of parametric selection models will result in a test of 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 and location of the intervals.

Future research might be directed to the issue of exploring additional restrictions that could be used to shrink the intervals further. Ideally, a smooth process of imposing restrictions would start with the no-information case and end with consistent point estimates of different selection models.

References

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.

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 (1997): "Assessing the Impact of Training on Employment: The Case of East Germany", *IFO Studien*, 43, 71-116.

Granger, C. W. J. (1969): "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods", *Econometrica*, 37, 424-438.

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 (with discussion).

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

Holland, P.W. (1986): "Statistics and Causal Inference", *Journal of the American Statistical Association*, 81, 945-970 (with discussion).

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", *Journal of Economic Perspectives*, 9, 149-168.

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; forthcoming in *Annales d'Economie et de Statistique*.

Lechner, M. (1999): "Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification", *Journal of Business & Economic Statistics*, 17, 74-90.

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 Resources*, 24, 343-360.

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

Manski, C.F. (1993): "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., 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 (eds), *Handbook of Econometrics, Vol. 4*, 2444-2521, Amsterdam: North-Holland.

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

In the main body of the text, the lower and upper bounds of the support of Y^t and Y^c as L^t, L^c U^t, and U^c. In this appendix the support is allowed to be different for different populations defined by *X* and *S*. For a given value of $X = x$ the bounds are defined as $L^t(x)$, $L^c(x)$, $U^t(x)$, and $U^c(x)$. If the treatment status is also taken into account, the notation $\ell^t(x,s)$, $\ell^c(x,s)$ is used for the lower bounds, and the upper bounds are denoted $u^t(x,s)$, $u^c(x,s)$. 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 $U^t(x) = \sup\{u^t(x,0), u^t(x,1)\}$, and $U^c(x) = \sup\{u^c(x,0), u^c(x,1)\}$. This notation is used to derive the bounds and interval widths for restrictions discussed in the main body of the text. Results for $\gamma(x)$ are presented in Tables A.1 and A.2. Tables A.3 and A.4 contain the results for $\theta(x)$. The respective quantities are derived exactly as explained in sections 5 and 6.

One restriction is contained in the tables and is not discussed in the main body of text, because it only relevant when the support is not the same for all populations. These are level set restriction on the expectations of the counterfactuals $E(Y^c | X = x, S = 1)$ and $E(Y^t | X = x, S = 0)$. It is assumed that $E(Y^c | X = x, S = 1)$ is constant in $\chi^{0,c}$ and that $E(Y^t | X = x, S = 0)$ is constant in $\chi^{0,t}$. For example for $\theta(x)$ where only $E(Y^c | X = x, S = 1)$ is relevant, using the same reasoning as before this assumption implies that the upper bound for $\theta(x)$ is $\inf_{x \in \chi^{0,c}} u^c(x,1)$ and the lower bound is $\sup_{x \in \chi^{0,c}} \ell^c(x,1)$. This set of assumptions does not generally include the ones that impose level set restriction on $E(Y^c | X = x)$ or $E(Y^t | X = x)$, hence there is no guarantee that the bounds 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.

Table A.1: 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 $\gamma^0(x)$, $\forall x \in \chi^0$	$\sup_{x\in\chi^0}[\{g^t(x) - u^c(x,1)\}p(x) +$ $\{\ell^t(x,0) - g^c(x)\}\{1 - p(x)\}]$	$\inf_{x\in\chi^0}[\{g^t(x) - \ell^c(x,1)\}p(x) +$ $\{u^t(x,0) - g^c(x)\}\{1 - p(x)\}]$
Local exclusion for $E(Y^t \| X = x)$, $\forall x \in \chi^{0,t}$, and $E(Y^c \| X = x)$, $\forall x \in \chi^{0,c}$	$\sup_{x\in\chi^{0,t}}[g^t(x)p(x) + \ell^t(x,0)\{1 - p(x)\}] -$ $\inf_{x\in\chi^{0,c}}[g^c(x)\{1 - p(x)\} + u^c(x,1)\}p(x)]$	$\inf_{x\in\chi^{0,t}}[g^t(x)p(x) + u^t(x,0)\{1 - p(x)\}] -$ $\sup_{x\in\chi^{0,c}}[g^c(x)\{1 - p(x)\} - \ell^c(x,1)p(x)]$
Local exclusion for $E(Y^c \| X = x, S = 1)$, $\forall x \in \chi^{0,c}$, and $E(Y^t \| X = x, S = 0)$, $\forall x \in \chi^{0,t}$	$\{g^t(x) - \inf_{x\in\chi^{0,c}} u^c(x,1)\}p(x) +$ $\{\sup_{x\in\chi^{0,t}} \ell^t(x,0) - g^c(x)\}\{1 - p(x)\}$	$\{g^t(x) - \sup_{x\in\chi^{c,0}} \ell^c(x,1)\}p(x) +$ $\{\inf_{x\in\chi^{0,t}} u^t(x,0) - g^c(x)\}\{1 - p(x)\}$
Selection on nonnegative expected effect $E(Y^t \| X = x, S = 1) -$ $E(Y^c \| X = x, S = 1) \geq 0$	$\{\ell^t(x,0) - g^c(x)\}\{1 - p(x)\}$	$\{g^t(x) - \ell^c(x,1)\}p(x) +$ $\{u^t(x,0) - g^c(x)\}\{1 - p(x)\}$
Same effect for treated and controls: $\gamma^0(x) = \theta^0(x) = \xi^0(x)$, $\forall x \in \chi^0$	$\max\{B_\theta^L(x), B_\xi^L(x)\}$	$\min\{B_\theta^U(x), B_\xi^U(x)\}$

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.2: 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 $E(Y^t\|X=x)$, $\forall x\in\chi^{0,t}$, and $E(Y^c\|X=x)$, $\forall x\in\chi^{0,c}$	$\inf_{x\in\chi^{0,t}}[g^t(x)p(x)+u^t(x,0)\{1-p(x)\}]-\sup_{x\in\chi^{0,c}}[g^c(x)\{1-p(x)\}+\ell^c(x,1)p(x)]-$ $\sup_{x\in\chi^{0,t}}[g^t(x)p(x)+\ell^t(x,0)\{1-p(x)\}]+\inf_{x\in\chi^{0,c}}[g^c(x)\{1-p(x)\}+u^c(x,1)p(x)]$
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 on nonnegative expected effect: $E(Y^t\|X=x,S=1)-E(Y^c\|X=x,S=1)\geq 0$	$\{g^t(x)-\ell^c(x,1)\}p(x)+\{u^t(x,0)-\ell^t(x,0)\}\{1-p(x)\}$
Same effect for treated and controls: $\gamma^0(x)=\theta^0(x)=\xi^0(x)$, $\forall x\in\chi^0$	$\min\{B^U_\theta(x),B^U_\xi(x)\}-\max\{B^L_\theta(x),B^L_\xi(x)\}$

Table A.3: 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 $E(Y^c\|X=x)$, $\forall x\in\chi^0$ *)	$g^t(x)-g^c(x)+\frac{g^c(x)-\inf_{x\in\chi^0}v^u(x)}{p(x)}$	$g^t(x)-g^c(x)+\frac{g^c(x)-\sup_{x\in\chi^0}v^\ell(x)}{p(x)}$
Local exclusion for $E(Y^c\|X=x,S=1)$, $\forall x\in\chi^{0,c}$	$g^t(x)-\inf_{x\in\chi^{0,c}}u^c(x,1)$	$g^t(x)-\sup_{x\in\chi^{0,c}}\ell^c(x,1)$
Selection on nonnegative expected effect: $E(Y^t\|X=x,S=1)-E(Y^c\|X=x,S=1)\geq 0$	0	$g^t(x)-\ell^c(x,1)$
Same effect for treated and controls: $\gamma^0(x)=\theta^0(x)=\xi^0(x)$, $\forall x\in\chi^0$	$\max\{B_\theta^L(x),B_\xi^L(x)\}$	$\min\{B_\theta^U(x),B_\xi^U(x)\}$

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

Table A.4: 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 $\theta^0(x)$, $\forall x\in\chi^0$	$\inf_{x\in\chi^0}\{g^t(x)-\ell^c(x,1)\}-\sup_{x\in\chi^0}\{g^t(x)-u^c(x,1)\}$
Local exclusion for $E(Y^c\|X=x)$, $\forall x\in\chi^0$	$\frac{\inf_{x\in\chi^0}v^u(x)-\sup_{x\in\chi^0}v^\ell(x)}{p(x)}$
Local exclusion for $E(Y^c\|X=x,S=1)$, $\forall x\in\chi^{0,c}$	$\inf_{x\in\chi^{0,c}}u^c(x,1)-\sup_{x\in\chi^{0,c}}\ell^c(x,1)$
Selection on nonnegative expected effect: $E(Y^t\|X=x,S=1)-E(Y^c\|X=x,S=1)\geq 0$	$g^t(x)-\ell^c(x,1)$
Same effect for treated and controls: $\gamma^0(x)=\theta^0(x)=\xi^0(x)$, $\forall x\in\chi^0$	$\min\{B_\theta^U(x),B_\xi^U(x)\}-\max\{B_\theta^L(x),B_\xi^L(x)\}$

Note: See note on Table A.3.

Combining the cases given in Tables A.1 and A.3 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 and sharpens the bounds: The lower bound is then given by $\sup_{x \in \chi^0} \{ \theta^0(x),$ $\gamma^0(x) \}$ and the upper bound is equal to $\inf_{x \in \chi^0} \{ \theta^0(x), \; \gamma^0(x) \}$.

Finally, there are several other restriction discussed by Manski in various papers that are however not used in the empirical part, because they appeared to be too restrictive for the particular application. Manski (1990) introduced one particular restriction stating that only individuals with a nonnegative effect (all of them, and not the average as used above!) are selected. However, in a social context this is hardly plausible, because it may very well require too many resources and too much information (about the future!) for those who select participants. Several other restrictions appear in the different papers by Manski. Those most closely related to our problem are discussed in Manski (1993, p. 163, 164). However, the assumption of *ordered outcomes* a priori means 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.