Universität St.Gallen

Combining Matching and Nonparametric IV Estimation: Theory and an Application to the Evaluation of Active Labour Market Policies

Markus Frölich and Michael Lechner

June 2010 Discussion Paper no. 2010-21

Department of Economics University of St. Gallen

Editor:	Martina Flockerzi University of St. Gallen Department of Economics Varnbüelstrasse 19 CH-9000 St. Gallen	
	Phone	+41 71 224 23 25
	Fax	+41 71 224 31 35
	Email	vwaabtass@unisg.ch
Publisher:	Department of Economics University of St. Gallen Varnbüelstrasse 19 CH-9000 St. Gallen	
	Phone	+41 71 224 23 25
	Fax	+41 71 224 31 35
Electronic Publication:	http://www.vwa.unisg.ch	

Combining Matching and Nonparametric IV Estimation: Theory and an Application to the Evaluation of Active Labour Market Policies [1]

Markus Frölich and Michael Lechner[2]

Author's address: Prof. Dr. Michael Lechner
SEW-HSG
Varnbüelstrasse 14
CH-9000 St. Gallen
Phone +41 71 224 2320
Fax +41 71 224 2302
Email Michael.Lechner@unisg.ch
Website www.sew.unisg.ch/lechner

Prof. Dr. Markus Frölich
Universität Mannheim
L7, 3-5
D-68131 Mannheim
Email froelich@uni-mannheim.de
Website froelich.vwl.uni-mannheim.de

[1] We gratefully acknowledge financial support from the Swiss National Science Foundation (projects 4043-058311 and 4045-050673), the Marie Curie Individual Fellowship MEIF-CT-2004-006873, the Grundlagen-forschungsfonds HSG (project G02110112), and COST A23. Part of the data originated from a database generated for the evaluation of the Swiss active labour market policies together with Michael Gerfin. We are grateful to the Swiss State Secretariat for Economic Affairs (seco; Arbeitsmarktstatistik), particularly Jonathan Gast, and the Swiss Federal Office for Social Security (BSV) for providing the data and to Dragana Djurdjevic, Michael Gerfin and Heidi Steiger for their substantial input in preparing them.

[2] Markus Frölich is also affiliated with IZA, Bonn and IFAU, Uppsala. Michael Lechner has further affiliations with ZEW, Mannheim, CEPR, London, CES-ifo, Munich, IZA, Bonn, and PSI, London.

Abstract

In this paper, we show how instrumental variable and matching estimators can be combined in order to identify a broader array of treatment effects. Instrumental variable estimators are known to estimate effects only for the compliers, which often represent only a small subset of the entire population. By combining IV with matching, we can estimate also the treatment effects for the always- and never-takers. In our application to the active labour market programmes in Switzerland, we find large positive employment effects for at least 8 years after treatment for the compliers. On the other hand, the effects for the always- and never-participants are small. In addition, when examining the potential outcomes separately, we find that the compliers have the worst employment outcomes without treatment. Hence, the assignment policy of the caseworkers was inefficient in that the always-participants were neither those with the highest treatment effect nor those with the largest need for assistance.

Keywords

Local average treatment effect, conditional local IV, matching estimation, heterogeneous treatment effects, active labour market policy, state borders, geographic variation.

JEL Classification

C14, C2, J68

1 Introduction

Although many empirical studies find it challenging to uncover one set of credible assumptions that allows point-identification identification of causal effects of interventions (treatments), sometimes several such assumptions may be available. If treatment effects are heterogeneous, such different identifying assumptions are likely to identify treatment effects for different populations. We show how the combination of these assumptions might lead to an increased policy relevance of the findings. One case where two sets of identifying assumptions plausibly hold simultaneously is the evaluation of the Swiss active labour market policy based on rich administrative data: Gerfin and Lechner (2002) and Gerfin, Lechner, and Steiger (2005) argue that the data is informative enough such that a conditional independence (or no confounding, or matching) assumption holds. In addition, Frölich and Lechner (2010) constructed an instrument that allowed them to estimate the effect of marginally changing the size of the policy.

One important difference between instrumental variable (IV) estimators and matching estimators is that IV estimates only treatment effects for the compliers (LATE, Imbens and Angrist 1994), whereas matching estimators estimate average population effects. In many situations, however, one would also like to know the treatment effects for the always- and never-takers, which are not identified by either of those approaches. In this paper, we show how IV and matching estimators can be combined in order to identify treatment effects for the compliers, and the always- and never-participants. These methods are then used to obtain a broader understanding of the effects of active labour market policies in Switzerland. Frölich and Lechner (2010) found rather large positive treatment effects for the group of compliers. Since in their work the compliers represented only a rather small fraction of the total population, effect estimates are also needed for the always- and never-participants in order to draw broader policy conclusions. In a similar setting Gerfin and

Lechner (2002) found considerably smaller average population effects, which were however uninformative about the effects for marginally increasing programme sizes, which is exactly what Frölich and Lechner (2010) identified.

In this paper, we find that the treatment effects for the always- and for the never-participants are much smaller than for the compliers. In addition to the treatment effects, we also examine the potential outcomes separately, in particular the potential outcome in the absence of treatment. Here, we find that the compliers have the worst employment chances, compared to the other groups. This means, that the assignment policy of the caseworkers was inefficient in that the always-participants were neither those with the highest treatment effect nor those with the largest need for assistance. This is important information for the design of a cost-effective active labour market policy.

Also we further analyze the long-term effects of labour market programmes and find that the short-term effects do not fade away. This is important since the previous evaluation studies of labour market programmes for Switzerland (and many other countries) have only been able to look at short-term effects (e.g. one to three years after participation).

The main contribution of this paper to the methodological literature is the combination of IV and matching methods in a fully nonparametric framework.[1] Heckman (1997) and others have questioned the use of the local average treatment effect (LATE, i.e. the effect on compliers), because it provides an incomplete description of the overall impact of the programme, refers to an unknown population and is unsuitable for a cost-benefit calculation.[2] If *continuous* instruments happen to be available, an alternative is to estimate marginal treatment effects (Heckman and Vyt-

[1] Instead of matching, one could also use weighting techniques as e.g. in Hirano, Imbens, and Ridder (2003).

lacil, 1999, 2005) and, if the instruments are sufficiently strong, by integrating the marginal treatment effects one could even obtain the average treatment effect or the average treatment effect on the treated. In many situations, however, such strong continuous instruments are not available. In such situations, one could impose more (parametric) structure on the model in order to identify more than just the LATE (implicitly restricting effect heterogeneity). An alternative approach we suggest in this paper is to combine IV and matching estimators. Identification via instrumental variables requires a selection-on-observables or unconfoundedness assumption for the compliers (Frölich, 2007). If we extend this assumption to the always- and never-takers, i.e. believe that we have sufficiently informative data to permit this, we can identify not only the LATE but also the treatment effects for the always- and for the never-takers. We will suggest fully nonparametric estimation approaches to estimate the latter effects.

Our main contribution to the empirical literature on the evaluation of active labour market policies is threefold:[3] First, we find that the positive programme effects of active labour market programmes are long lasting, for at least 8 years, confirming, for example, the results for Germany by Lechner, Miquel, and Wunsch (2005). Second, the allocation of unemployed to the programmes does not appear to have been fully effective in the sense that the group of always-participants has a lower programme effect *and* higher re-employment chances in the absence of the programmes than the group of compliers. These estimates are not consistent with an assignment policy where unemployed most in need or with largest programme effects were given priority. Otherwise, the always-participants should have had a higher programme effect than the compliers

---

[2] For example, Manski (2003) notes that results for unobserved populations cannot be used by a planner to choose a treatment. In addition, for cost-benefit calculations we must measure the impacts on those who actually have been treated.

[3] For recent surveys of this literature see e.g. Kluve (2006), Kluve and Schmidt (2002), Lalive, van Ours, and Zweimüller (2008), and Martin and Grubb (2001).

or should have had worse employment outcomes in absence of the programmes than the compliers. Neither is observed. Third, the quantity extension stimulated by the central government, however, appears to have had a positive effect on targeting. The extension of the programmes was targeted at those unemployed (i.e. these are the compliers) with large positive programme effects and poor employment chances without assistance.

2 Combination of nonparametric IV and matching

2.1 Nonparametric instrumental variable estimation

Let D be a binary treatment variable and Y^1 and Y^0 be the potential outcomes in case of participating ($D=1$) or not participating ($D=0$) in the treatment. In our application, D will refer to participation in an active labour market programme and the outcome variable Y will be employment status or earnings, respectively, several years later. Let $Z \in [\underline{z}, \overline{z}]$ be an instrumental variable which takes values in the interval $[\underline{z}, \overline{z}]$ with masspoints at the boundaries of its support. Next, we will illustrate and motivate the results for the case where $Z \in \{0,1\}$ is *binary*, as this is the case in our application, but all results work for the more general case of non-binary Z. (As we will point out later, if Z is non-binary, we will use the two values corresponding to the endpoints of its support in order to obtain the effects for the largest complier subpopulation.) Let D_z denote the potential treatment status of an individual i if the level of the instrument were externally set to z. With the instrument taking only two different values, the potential treatment variable D_z defines four different types of individuals denoted by $T \in \{a, n, c, d\}$. Following the literature, we call these different groups always-treated (a), never-treated (n), compliers (c) and defiers (d). The treatment status of the first two groups is not affected by the instrument. The group of compliers would not be treated if $Z = \underline{z}$, but would be treated if $Z = \overline{z}$. For the defiers, this pattern is reversed. Under conditions described more precisely below, Imbens and Angrist (1994) have shown that the treatment effect for the subpopulation of *compliers* is identified as:

$$E[Y^1 - Y^0 \mid T = c] = \frac{E[Y \mid Z = \overline{z}] - E[Y \mid Z = \underline{z}]}{E[D \mid Z = \overline{z}] - E[D \mid Z = \underline{z}]}. \tag{1}$$

For identification, they require that the instrument Z is unconfounded. Such an assumption is reasonable when the instrument Z has been completely randomly assigned. In many situations, however, Z may be a choice variable or it may be affected by various other characteristics, such that the assumption of unconfounded Z is often questionable. We extend their setup in that we require Z to be unconfounded only *conditional* on some characteristics X.[4] In our application, for example, Z is determined by a rule that depends on three characteristics of the local sites. These characteristics, as we discuss later, are likely to be related to the potential outcomes, thus violating the conventional instrumental variables assumption. However, *conditional* on the characteristics that determine Z, the instrumental variable assumption appears reasonable. We assume in the following:

Assumption 1 (Conditional Instrumental Variable Assumption):

(CIV.1) **No defiers:** $\Pr(T = d) = 0$

(CIV.2) **Compliers:** $\Pr(T = c) > 0$

(CIV.3) **Unconfounded type:** For almost every X

[4] Note that the nonparametric approach permits, to some extent, endogenous control variables X, i.e. the variables X may be correlated with the unobservables affecting the outcome Y, see e.g. Frölich (2008). This matters in our application, because some of the X variables, e.g. past employment history, may well be correlated with the unobservables affecting our outcome variable future employment status.

$$\Pr(T=t \mid X, Z=\underline{z}) = \Pr(T=t \mid X, Z=\overline{z}) \qquad \textit{for } t \in \{\text{a,n,c}\}$$

(CIV.4) **Exclusion restriction:** For almost every X

$$E[Y^0 \mid X, Z=\underline{z}, T=t] = E[Y^0 \mid X, Z=\overline{z}, T=t] \qquad \textit{for } t \in \{\text{n,c}\}$$
$$E[Y^1 \mid X, Z=\underline{z}, T=t] = E[Y^1 \mid X, Z=\overline{z}, T=t] \qquad \textit{for } t \in \{\text{a,c}\}$$

(CIV.5) **Common support:** $Supp(X \mid Z=\underline{z}) = Supp(X \mid Z=\overline{z})$,

Frölich (2007) showed that under Assumption 1[5] the treatment effect for all compliers $E[Y^1 - Y^0 \mid T=c]$ is nonparametrically identified as:

$$E[Y^1 - Y^0 \mid T=c] = \frac{\int \left(E[Y \mid X, Z=\overline{z}] - E[Y \mid X, Z=\underline{z}]\right) dF_X}{\int \left(E[D \mid X, Z=\overline{z}] - E[D \mid X, Z=\underline{z}]\right) dF_X}. \tag{2}$$

This formula is obtained by integrating out the distribution of X in the unknown complier population, and the effect can be estimated as

$$\widehat{E[Y^1 - Y^0 \mid T=c]} = \frac{\sum_{i:z_i=\overline{z}} \left(y_i - \hat{m}_{\underline{z}}(x_i)\right) - \sum_{i:z_i=\underline{z}} \left(y_i - \hat{m}_{\overline{z}}(x_i)\right)}{\sum_{i:z_i=\overline{z}} \left(d_i - \hat{\mu}_{\underline{z}}(x_i)\right) - \sum_{i:z_i=\underline{z}} \left(d_i - \hat{\mu}_{\overline{z}}(x_i)\right)}, \tag{3}$$

[5] Note that if the set of control variables X is *empty*, Assumption 1 is basically the same as in Imbens and Angrist (1994). Their assumptions were (Exclusion restriction): $Y^d_{i,z'} = Y^d_{i,z''}$ for all d, z', z''. (Existence of instrument): $Y_i^0, Y_i^1, D_{i,z} \coprod Z_i$ for all $z \in Supp(Z)$ where $\coprod$ denotes independence. (Relevance of instrument): $E[D \mid Z=z]$ is a nontrivial function of z. (Monotonicity): for all pairs (z, z') either $D_i(z) \geq D_i(z')$ for all *i* or vice versa $D_i(z) \leq D_i(z')$ for all *i*.

where $\hat{m}_z$ and $\hat{\mu}_z$ are nonparametric estimators of $m_z(x) = E[Y \mid X = x, Z = z]$ and $\mu_z(x) = E[D \mid X = x, Z = z]$. Notice that the denominator in the above formula is also an estimate of the fraction of compliers $\Pr(T = c)$.

Next, we derive additional results to Frölich (2007), which will be helpful in obtaining a more detailed picture of the evaluation results. First, the population of compliers consists of two subpopulations: those who actually receive treatment and those who do not. When Assumption 1 is valid *without* any X, the effects for these two subpopulations are identical; otherwise they are not. We can show that the treatment effect on the treated compliers is identified as:

$$E[Y^1 - Y^0 \mid D = 1, T = c] = \frac{\int \left(E[Y \mid X, Z = \overline{z}] - E[Y \mid X, Z = \underline{z}] \right) \cdot \pi(X) \cdot dF_X}{\int \left(E[D \mid X, Z = \overline{z}] - E[D \mid X, Z = \underline{z}] \right) \cdot \pi(X) \cdot dF_X}, \tag{4}$$

where $\pi(x) = \mathrm{P}(Z = \overline{z} \mid X = x, Z \in \{\underline{z}, \overline{z}\})$, see this proof in the supplementary appendix.

Second, we can identify the potential outcomes of the complier population separately as

$$E[Y^1 \mid T = c] = \frac{\int \left(E[YD \mid X, Z = \overline{z}] - E[YD \mid X, Z = \underline{z}] \right) dF_X}{\int \left(E[D \mid X, Z = \overline{z}] - E[D \mid X, Z = \underline{z}] \right) dF_X},$$

$$E[Y^0 \mid T = c] = -\frac{\int \left(E[Y(1-D) \mid X, Z = \overline{z}] - E[Y(1-D) \mid X, Z = \underline{z}] \right) dF_X}{\int \left(E[D \mid X, Z = \overline{z}] - E[D \mid X, Z = \underline{z}] \right) dF_X}.$$

(The proofs are similar to the proof of the preceding expression and are omitted.)

We obtain similar expressions for the treated compliers:

$$E[Y^1 \mid D = 1, T = c] = \frac{\int \left(E[YD \mid X, Z = \overline{z}] - E[YD \mid X, Z = \underline{z}] \right) \cdot \pi(X) \cdot dF_X}{\int \left(E[D \mid X, Z = \overline{z}] - E[D \mid X, Z = \underline{z}] \right) \cdot \pi(X) \cdot dF_X},$$

$$E[Y^0 \mid D=1, T=c] = -\frac{\int \left(E[Y(1-D) \mid X, Z=\overline{z}] - E[Y(1-D) \mid X, Z=\underline{z}] \right) \cdot \pi(X) \cdot dF_X}{\int \left(E[D \mid X, Z=\overline{z}] - E[D \mid X, Z=\underline{z}] \right) \cdot \pi(X) \cdot dF_X}.$$

The corresponding estimators are analogous to equation (3).

Hence, via Assumption 1 we can identify $E[Y^1 \mid T=c]$ and $E[Y^0 \mid T=c]$ separately. Knowing the non-treatment outcome for the compliers will help us later to better understand who the compliers are. (In addition, estimating $E[Y^1 \mid T=c]$ and $E[Y^0 \mid T=c]$ separately allows imposing restrictions on the range of the outcome variables in a straightforward way by capping them at the logical boundaries of their supports. That is, if the outcome variable is binary, both mean values should lie between 0 and 1.)

Similar to the literature on matching estimators, a dimension reduction via a "propensity score" is possible. Let $\hat{\pi}_i$ be a consistent estimator of $\pi_i = \mathrm{P}(Z=\overline{z} \mid X=x_i, Z \in \{\underline{z}, \overline{z}\})$, then the propensity score based estimator

$$\widehat{E[Y^1 - Y^0 \mid T=c]} = \frac{\sum_{i:z_i=\overline{z}} \left(y_i - \hat{\mathrm{m}}_{\underline{z}}(\hat{\pi}_i) \right) - \sum_{i:z_i=\underline{z}} \left(y_i - \hat{\mathrm{m}}_{\overline{z}}(\hat{\pi}_i) \right)}{\sum_{i:z_i=\overline{z}} \left(d_i - \hat{\mu}_{\underline{z}}(\hat{\pi}_i) \right) - \sum_{i:z_i=\underline{z}} \left(d_i - \hat{\mu}_{\overline{z}}(\hat{\pi}_i) \right)}, \tag{5}$$

where $\mathrm{m}_z(\rho) = E[Y \mid \pi(X)=\rho, Z=z]$ and $\mu_z(\rho) = E[D \mid \pi(X)=\rho, Z=z]$, is a consistent estimator of the LATE, as shown in Frölich (2007). Compared to (3) it has the advantage that it requires only one-dimensional nonparametric regression, given estimates of π_i. Analogously, a propensity score based estimator of the potential outcomes $E[Y^1 \mid T=c]$ and $E[Y^0 \mid T=c]$ and of $E[Y^1 \mid D=1, T=c]$ and $E[Y^0 \mid D=1, T=c]$ can be obtained.

Besides the potential outcomes for the compliers, we identify the fractions of compliers, always-participants and never-participants, as well as the expected treatment outcome for the always-participants and the expected non-treatment outcome for the never-participants:

$$\Pr(T=c)=\int\left(E[D\mid X,Z=\overline{z}]-E[D\mid X,Z=\underline{z}]\right)dF_X ,$$

$$\Pr(T=a)=\int E[D\mid X,Z=\underline{z}]dF_X ,$$

$$\Pr(T=n)=\int E[1-D\mid X,Z=\overline{z}]dF_X ,$$

$$E[Y^1\mid T=a]=\frac{\int E[YD\mid X,Z=\underline{z}]dF_X}{\int E[D\mid X,Z=\underline{z}]dF_X},$$

$$E[Y^0\mid T=n]=\frac{\int E[Y(1-D)\mid X,Z=\overline{z}]dF_X}{\int E[1-D\mid X,Z=\overline{z}]dF_X}.$$

The proofs are analogous to those for the previous results and are omitted.

Hence, as already pointed out in Imbens and Angrist (1994), via instrumental variables we can estimate the treatment effect for the compliers but *not* for the always- nor for the never-treated.[6]

---

[6] So far, we discussed the case where the instrument $Z\in\{\underline{z},\overline{z}\}$ is *binary*. We mentioned in the beginning of this section that we can also permit for Z to be non-binary or continuously distributed. In this case, we can define LATE for any pair of values of z or even marginal treatment effects if Z is continuous (Heckman and Vytlacil, 1999, 2005). Any pair of values of z would define a different subpopulation of compliers. Nevertheless, the effect on the *largest* population of compliers is identified by using the boundaries of the support of Z. As long as $Z\in[\underline{z},\overline{z}]$ has masspoints at the boundaries of its support, the previously defined nonparametric estimators using $\underline{z}$ and $\overline{z}$ are root-n consistent under smoothness assumptions discussed in more detail in Frölich (2007). (Of course, we could

2.2 Combination of Matching and IV estimation

In many situations, however, it would be helpful to know also the potential outcomes for the always- and never-treated. The missing pieces not identified in the previous section are $E[Y^0 \mid T=a]$ and $E[Y^1 \mid T=n]$. Next, we show that under an additional assumption, we can identify $E[Y^0 \mid T=c]$, $E[Y^0 \mid T=a]$ and $E[Y^0 \mid T=n]$ and $E[Y^1 \mid T=c]$, $E[Y^1 \mid T=a]$ and $E[Y^1 \mid T=n]$. This will permit us not only to compare the treatment effects for these groups, but information about the non-treatment outcome Y^0 for each group will also enhance our understanding of who the compliers are. In our application, where Y refers to employment, we will find that the never-treated have on average a larger non-treatment outcome Y^0 than the always-treated, who themselves have a larger Y^0 than the compliers. Hence, in our population of unemployed, this indicates that the never-treated are the "good-risks", who most likely find a job even without assistance, whereas the compliers are the "bad-risks", who have the least chances to find a job without assistance. The always-treated represent an intermediate group.

Following, we lay out that the IV estimator of Section 2.1 is identical to a "Selection-on-observables" or "Matching" estimator *for the compliers*. Hence, the IV estimator relies on a selection-on-observables assumption for the compliers. If, furthermore, we believe that we can extend the selection-on-observables assumption also to the *non-compliers*, i.e. have sufficiently informative data to justify this, we will find the missing pieces.

Matching estimators have been popularized by Rosenbaum and Rubin (1983), Heckman, Ichimura and Todd (1998), Lechner (1999), Imbens (2004) and others. Under an assumption known as "selection on observables" or "ignorable treatment assignment":

---

obtain even more information by examining the various complier groups separately, but since Z is essentially binary

(CIA) $$Y^1, Y^0 \perp D \mid X\ , \tag{6}$$

where A⊥B|C denotes *mean* independence of A and B conditional on C, the average treatment effect (ATE) is identified as

$$E[Y^1 - Y^0] = \int \left(E[Y \mid X, D=1] - E[Y \mid X, D=0] \right) dF_X\ , \tag{7}$$

provided there is common support.

To better understand the relationship between the IV and the matching estimator, let us assume selection-on-observables in the complier population:

$$Y^1, Y^0 \perp D \mid X, T=c\ . \tag{CIV.4'}$$

Developing a matching estimator for the complier treatment effect analogous to (7) would result in

$$E[Y^1 - Y^0 \mid T=c] = \int \left(E[Y \mid X, D=1, T=c] - E[Y \mid X, D=0, T=c] \right) dF_{X \mid T=c}\ . \tag{8}$$

In contrast to (7), however, this is not directly identified, because the type (a, n, c) is unobserved. Nevertheless, in the supplementary appendix it is shown that using Bayes' theorem this expression is exactly equivalent to (2). Therefore, the IV estimator is essentially a matching estimator for the compliers. It thus requires conditional independence for the compliers, whereas identification of the ATE requires conditional independence to hold also for the *non-compliers*.

in our application, we do not discuss this further here.)

In fact, assumptions CIV.4' and CIV.4 for the compliers are equivalent, given the direct correspondence between D and Z for the compliers. Hence, CIV-4' can be thought of as the core "selection-on-observables" assumption. The IV estimator requires additionally the assumptions CIV.1, CIV.2, CIV.3, CIV.5 and CIV.4 for the never- and always-participants, whereas, loosely speaking, matching estimators of ATE require "selection-on-observables" to hold for always- and never-participants as well (and for defiers, if they exist).

The "selection-on-observables" assumption may often be more plausible for the compliers than for the always- and never-participants, as they are at the margin of changing participation status, i.e. their participation status may be more or less random. Nevertheless, in various applications one may be willing to extend this assumption also to the always- and never-participants. If, in addition to Assumption 1, we assume

$$Y^0 \perp D \mid X \,, \tag{9}$$

the expected potential outcome $E[Y^0 \mid T=a]$ is identified: By noting that

$$E[Y^0] = E[Y^0 \mid T=a]\Pr(T=a) + E[Y^0 \mid T=c]\Pr(T=c) + E[Y^0 \mid T=n]\Pr(T=n)\,,$$

it follows that

$$E[Y^0 \mid T=a] = \frac{E[Y^0] - E[Y^0 \mid T=c]\Pr(T=c) - E[Y^0 \mid T=n]\Pr(T=n)}{\Pr(T=a)}\,. \tag{10}$$

Alternatively, a similar decomposition of $E[Y^0 \mid D=1]$ leads to:

$$E[Y^0 \mid T=a] = \frac{E[Y^0 \mid D=1] - E[Y^0 \mid T=c, D=1]\Pr(T=c \mid D=1)}{\Pr(T=a \mid D=1)}\,. \tag{11}$$

Since all terms on the right-hand side are identified, also noting that $\Pr(a \mid D=1) = \Pr(a)/\Pr(D=1)$ using Bayes' theorem, the outcome $E[Y^0 \mid T=a]$ can be estimated. With the corresponding outcome $E[Y^1 \mid T=a]$ already having been identified in Section 2.1, the treatment effect for the always-participants can be estimated.

To identify the effect for never-participants, we assume that

$$Y^1 \perp D \mid X \,, \tag{12}$$

to obtain the corresponding quantity for the never-participants (analogous derivations as before):

$$E[Y^1 \mid T=n] = \frac{E[Y^1] - E[Y^1 \mid T=c]\Pr(T=c) - E[Y^1 \mid T=a]\Pr(T=a)}{\Pr(T=n)}, \tag{13}$$

$$E[Y^1 \mid T=n] = \frac{E[Y^1 \mid D=0] - E[Y^1 \mid T=c, D=0]\Pr(T=c \mid D=0)}{\Pr(T=n \mid D=0)}. \tag{14}$$

Hence, $E[Y^0 \mid T=c]$, $E[Y^0 \mid T=a]$ and $E[Y^0 \mid T=n]$ and $E[Y^1 \mid T=c]$, $E[Y^1 \mid T=a]$ and $E[Y^1 \mid T=n]$ are all identified if we combine CIV with CIA.

3 Evaluation of active labour market policies in Switzerland

3.1 Active labour market programmes, regional quota, local labour markets, and the resulting instrument

As in many European countries, active labour market programmes were widely introduced in Switzerland during the 1990s. Until the recession of the early 1990s, unemployment was very low in Switzerland, a small country with 26 different administrative regions, called *cantons*. With the recession, the unemployment rate rose rapidly to 5% and triggered a comprehensive revision of the federal unemployment insurance act. This revision, which became effective partly in January 1996 and partly in January 1997, introduced active labour market programmes (ALMP) on a much wider scale than before. Although different in some details, the main components of the Swiss ALMP can be found in various programmes in the USA, Germany and the UK as well, and include training and subsidized employment and on-the-job training in private as well as public sector jobs. A key element of the reform, which will permit an instrumental variables approach, was the introduction of a *minimum quota* in order to provide a sufficiently large number of programme places.

The 26 Swiss cantons enjoy a high degree of autonomy with respect to taxation, expenditure and many other policies. Therefore, there was a suspicion that the cantons might have been slow or even reluctant to implement the reform. To accelerate the implementation of the reform and the provision of active labour market programmes, the federal government mandated by law a minimum number of places in labour market programmes to be filled per year. For the year 1998, the minimum number was 25000 year-places (each representing 220 programme days) and was distributed across the cantons according to the formula

$$12'500 \cdot \left(\text{population share}_{1996} + \text{unemployment share}_{1996}\right),$$

where population share is the fraction of the population living in the respective canton as of 1996 and unemployment share is the average number of unemployment benefit recipients in the period April 1996 to March 1997 in the respective canton relative to the total for Switzerland.[7]

[7] The costs of active labour market programmes and of their administration generally are borne by the federal unemployment insurance fund. The cantons pay a very small lump sum contribution of 3000 Swiss Francs (CHF) per year-place for their assigned minimum quota. They can reduce this lump sum payment by up to 25% if the average

This formula for the computation of the minimum quota induced regional variation in programme participation, because, *relative* to the number of unemployed persons, the quota was rather high in cantons with a low unemployment rate in 1996 since 50% of the quota was distributed according to the population share.

This minimum quota was used by Frölich and Lechner (2010) to estimate treatment effects for the compliers, and we roughly present it here. In this paper, we complement their empirical analysis by also estimating the treatment effects for the always- and the never-treated. Such treatment effects are interesting in that the compliers actually represent only a rather small fraction of the total population. In addition, we also provide estimates for the long-term effects, whereas Frölich and Lechner (2010) only contained short-term effects.

Frölich and Lechner (2010) analyzed and discussed in detail that the proclaimed minimum quotas, which were codified in law in November 1996, indeed induced a regional variation in the probability of being treated: The correlation between the quota per unemployed and treatment incidence for the population of unemployed was 0.53 across the cantons. However, using the minimum quota in a conventional instrumental variable analysis might not be a valid approach as one would be comparing Western and urban regions of Switzerland (where the quota was relatively lower) to regions of Eastern and Central Switzerland (where the quota was relatively higher). These regions, however, differ not only in their quota but also in many other respects, including past unemployment rates and industry structure, such that the needed exclusion restriction might

unit costs of the purchased programme slots are below the national average within defined programme categories. No financial contribution has to be paid for places filled beyond the required minimum. On the other hand, cantons which fill less than the required minimum number of year-places, have to compensate the federal unemployment insurance fund with 20% of the unemployment benefits paid to those persons to whom no ALMP could be offered. Hence, there are financial and political incentives for the cantons to meet their quota. In fact, they were encouraged to provide even more ALMP places.

not be plausible. (Generally speaking, the formula used for calculating the quota implies that the quota per unemployed is higher in regions where unemployment was low in 1996.)

As an alternative, therefore they used the minimum quota as an instrument only within *local labour markets* that are partitioned by a cantonal border, i.e. within a neighbourhood of a cantonal border. Individuals living left and right of a cantonal border essentially live in the same labour market and have access to the same job opportunities. When actually becoming unemployed, however, they have to register with the local employment office according to their place of residence. Their chances of being sent to active labour market programmes then depend on the management strategy of the local employment office, which is governed by the canton and seeking to fulfil the quota for the entire canton. Although living in the same local neighbourhood, the treatment probability in case of becoming unemployed is thus differentially affected by the rest of the canton.

Despite living in the same local area, we might potentially still be concerned about differences in the characteristics of the populations living left and right of the border. We therefore will also control for a large number of characteristics X_i, which includes among other things the unemployment history, which is a key determinant of the quota as seen from the formula above.[8]

3.2 Construction of the local labour markets

In this section, we briefly discuss the construction of the local labour markets. (More details can be found in the supplementary appendix and in Frölich and Lechner, 2010.) With the reform in 1996 the municipal unemployment offices were consolidated into about 150 regional em-

[8] Frölich and Lechner (2010) found that, in fact, there were no important differences in X for the compliers, since the estimates with and without controlling for X were very similar. Hence, endogeneity concerns *within* the local labour

ployment offices (REO), supervised by the cantonal centres. The REO are geographically organised, each REO serving several municipalities. For each unemployed there is one unique REO defined by place of residence. They cannot change their assigned REO other than by moving to another municipality. (Exceptions are the city centres of Zurich and Geneva, which are served by several REO, and which we exclude from our analysis.)

We think of a local labour market if the value of different job opportunities does not depend on the location of residence. In other words, all relevant employment opportunities can be reached within convenient commuting distance (e.g. half an hour) from both sides of the border, such that the choice of workplace location and the choice of residence are not immediately tied. There should then be no opportunities for wage arbitrage by moving residence. Switzerland, with its numerous winding administrative borders and a very good commuting infrastructure, is a candidate country for finding such local labour markets stretching across internal borders.

As in Frölich and Lechner (2010), we use the following criteria to define integrated *local labour markets* with internal administrative borders. A local labour market is defined in terms of the area corresponding to one or more regional employment offices (REO), which satisfy: (1) The REO is spread over 2 cantons, (2) commuting times by car between these REOs are 30 minutes or shorter, (3) the same language (French, German or Italian) is spoken in the areas belonging to the REO, (4) the ALMP composition is similar in the REO. With the first criterion, we identify local labour markets pair-wise between cantons. For the econometric analysis, this will imply that the instrumental variable *quota per unemployed* will take only *two* different values within each local labour market. The second criterion ensures that all potential employers can be reached within convenient commuting distance from both sides of the cantonal border. This criterion is imple-

markets do not seem to be important. In this paper, however, we only show the results with controlling for X, in ormented by examining the distances between any pair of regional employment offices in terms of commuting times by car.[9] The third criterion takes into account the different language regions, as Switzerland consists of German, French and Italian speaking parts.[10] The fourth criterion requires that the allocation of the treated to the different ALMP categories is similar on both sides of the border. As discussed in Section 2, the identification assumption CIV.4 or CIV.4' requires that the potential outcomes Y^0 and Y^1 are independent of the instrument. For Y^1 to be independent of the instrument, the quality and type of treatment should be identical on both sides of the border. It appears reasonable to assume that the quality of the services does not vary systematically between neighbouring regions, since these are offered by private profit or non-profit organisations which usually operate nationwide (or at least within the language region). On the other hand, there was some variation in the types of programmes used. We therefore restrict our analysis to those 18 local labour markets, which have the same ALMP-structure on each side. The following table provides some summary measures for these labour markets. (More details can be found in Frölich and Lechner (2010) and the discussion paper.) Column (1) indicates the cantonal border that partitions the labour market, and columns (2) and (3) report the REO belonging to this labour market, left and right of the border. Columns (4) and (5) report the number of observations in our dataset, left and right of the border. (The dataset is explained in the next section.) Columns (6) and (7) present which percentage of these observations was treated, left and right of the border. An unemployed person is defined as treated, if he/she entered a labour market programme (with duration of at least

---

der to save space (and also for comparability with the estimator used for the always- and never-participants).

9 Switzerland is one of the countries with the highest per capita car ownership worldwide. In addition, public transportation is very good and reaches every village.

10 Local labour markets where French is spoken on the one side of the border and German on the other side are excluded. French-German bilingual regions bordering to German speaking regions are not excluded, though. In such local labour markets, all observations with French mother tongue are deleted, as they may not consider the neighbouring German-speaking region as part of their labour market when searching for jobs.

one week) during January to March 1998. Finally, column (8) gives the difference between columns (6) and (7), i.e. the cross-border difference in the treatment probability. This difference represents an estimate of the fraction of *compliers* (when no covariates X are controlled for). This percentage of compliers lies in the range of ±18 percentage points, with many small values. It is highly correlated with the cantonal quota. As shown in Frölich and Lechner (2010), the correlation between the compliers and the differences in the quota is larger than 0.5. Hence, the instrument *quota per unemployed* does indeed have an effect on the probability of receiving treatment.

From Table 1, we also see that the compliers represent only a small portion of the total population. The largest number in column (8) is 18%, hence more than 80% of the unemployed in each labour market are always- and never-participants. Therefore, it is interesting to learn not only the treatment effects on the compliers but also for the always- and the never-participants. Before we show the latter estimates in Section 4, we first provide some more information on the individual-level dataset.

Frölich and Lechner (2010) discuss potential threats to the validity of an instrument constructed in such a way at substantial length and come to a positive conclusion. Here, we do not want to repeat these arguments but refer the reader to their paper.

Table 1: The 18 local labour markets divided by administrative border

Cantons	Regional employment offices (REO), left and right of the border		Number of observations N_1	N_2	% Treated		% Complier[b]
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SO-BE	Solothurn, Oensingen, Biberist, Zuchwil	Wangen, Langenthal, Burgdorf	877	818	51.7	48.2	3.5
BE-AG	Langenthal	Zofingen	313	472	45.7	49.2	-3.5
BE-FR	Gümligen, Zollikofen, Köniz, Bern (2x)	Murten, Tafers, Fribourg	2660[a]	763[a]	45.5	46.9	-1.4
FR-VD	ChatelSt.Denis	Oron la Ville	107	107	51.4	44.9	6.5
FR-VD	Romont, Estavayer	Payerne, Moudon	371	355	47.4	40.6	6.9
VD-GE	Nyon	Genf (6x)	576	5700	35.8	33.3	2.5
VD-VS	Vevey, Aigle, Montreux	Monthey (2x)	1580	609	40.3	50.1	-9.8
BL-BS	Pratteln, München-stein, Binningen	Basel (3x)	934	2081	52.0	34.3	17.8
LU-NWOW	Luzern, Emmen, Em-menbrücke, Kriens	Hergiswil (2x)	1607	265	49.0	52.5	-3.5
LU-ZG	Luzern, Emmen, Em-menbrücke, Kriens	Zug	1607	571	49.0	49.6	-0.6
SZ-UR	Goldau	Altdorf	337	150	57.9	39.3	18.5
AG-ZH	Baden, Wettingen, Wohlen	Opfikon, Effretikon, Uster, Wetzikon, Bülach, Dietikon, Regensdorf	1529	4165	45.5	39.7	5.8
ZH-TG	Winterthur	Frauenfeld	1221	537	39.0	53.6	-14.7
ZH-SG	Meilen, Thalwil	Rapperswil	1421	360	40.2	43.9	-3.7
ZH-SZ	Meilen, Thalwil	Lachen	1421	529	40.2	53.3	-13.1
TG-SH	Frauenfeld	Schaffhausen	537	605	53.6	45.6	8.0
TG-SG	Amriswil	Rohrschach, Oberuzwil	474	853	42.4	42.9	-0.5
SG-SZ	Rapperswil	Lachen	360	529	43.9	53.3	-9.4

Note: [a] Number of observations after deleting individuals with French mother tongue, because a French-German bilingual region is bordering a German-speaking region.
[b] The estimate of the fraction of compliers is the difference between the previous two columns. Here we do not control for differences in *X*.

3.3 Administrative data from the Swiss unemployment and pension system and the matching assumption

The basis of this study is a large random sample of Swiss unemployed, with individual information from very detailed administrative records from the unemployment insurance system and the social security / pension system. Those records contain ten years of employment histories (including self-employment), monthly earnings, monthly unemployment benefits, participation in ALMP and personal characteristics for the years 1988 to December 2006. (From the year 2000 onwards, only the data from the unemployment insurance system are available.) The personal information includes age, gender, marital status, household size, place of residence, nationality, type of work permit, mother tongue, foreign language skills, education, qualification, caseworker's rating of employability, position in last job, occupation and industry of last job, size of town where worked before, looking for part-time or full-time job, occupation and industry of desired job, information on earnings in last job, duration of contribution to unemployment insurance, disability, etc. The population for our study are all individuals who were unemployed on January 1, 1998, for at most one year.[11] We use the same dataset as in Frölich and Lechner (2010), which contains 32634 individuals who live in one of the 18 local labour markets of Table 1.[12]

In our empirical analysis, we estimate the effects of participation in ALMP during 1998 on employment and earnings in 1999 and the following years. Participation (=treatment) is defined as entering in a programme of at least one week duration during January to March 1998. We also

[11] Persons who were unemployed for more than one year are excluded because they entered in unemployment before the reform was enforced in January 1997 and were thus subject to different rules and regulations at the entry in unemployment.

[12] The original dataset, which is a random sample from the population of all individuals who were unemployed on January 1, 1998, for at most one year, contained 81399 individuals. Several sample selection criteria are applied to restrict the population to individuals who are *eligible* to take part in ALMP, and for whom no restrictions to their mobility are known or probable. In particular, disabled persons are excluded, as well as foreigners with a working permit of less than a year (i.e. without a 'B' or 'C' permit) since there are legal restrictions to their mobility. In addition, persons with very low earnings (monthly earnings in last job below 1000 CHF, ≈ 650 EURO) are excluded, because monetary costs of commuting might be an obstacle to taking advantage of job opportunities that are not nearby. We restrict the sample to the prime age group (25-55), and excluded students, apprentices and home workers, and persons registered as part-time employees. The remaining sample contains 66713 individuals, of which 32634 individuals live in one of the 18 local labour markets of Table 1.

examine robustness of the results when using a treatment window of four months (January to April 1998), respectively.[13]

For making the conditional independence assumption (CIA) of Section 2.2 credible, we must control for individual characteristics *X* that jointly affect treatment status as well as potential outcomes. Our selection of these control variables is based on earlier studies by Gerfin and Lechner (2002), Gerfin, Lechner and Steiger (2005) and Frölich and Lechner (2010). We control for *socio-demographic characteristics* (age, gender, marital status, household size, nationality, year of immigration), *language skills* (mother tongue, number of languages, first foreign lan-

[13] In Frölich and Lechner (2010), a treatment window of 12 months was used. (Therefore, their estimates for the compliers differ slightly from our estimates in Section 4 of this paper.) In contrast to Frölich and Lechner (2010), here we require conditional independence between the potential outcomes and treatment status not only for the compliers, but also for the entire population, i.e. including the always- and never-participants. In our matching estimator, we thus compare individuals with different values of *D*, rather than the quota *Z*. A problem discussed in detail in Fredriksson and Johansson (2008) is that the definition of treatment automatically implies some 'conditioning-on-the-future'. Suppose that treatment assignment *D* and finding a job are two competing processes. This will imply that the group of non-treated contains disproportionately many observations who found a job before treatment started, whereas the group of treated contains more individuals who would have found a job on their own rather late and treatment happened to start before it. Hence, individuals with good labour market chances (which may be partly unobservable) are over-represented in the D=0 group and under-represented in the D=1 group. Our matching estimates would thus tend to be downward biased, with the size of the effect increasing with the length of the treatment window. For a very short window, e.g. one day, the bias would be zero since in our population everybody is unemployed on January 1, 1998. On the other hand, a short treatment window leads to a very small number of treated and thus to imprecise estimates. In addition, a short treatment window also implies that many of those defined as untreated during this window might actually receive treatment shortly afterwards, thus making the interpretation of such estimates difficult. A treatment window of three or four months appears as a reasonable trade-off between sample sizes and bias in this application. Of our total sample, 60% of all unemployed entered active labour market programmes during the year 1998. Of these, 70% entered during the first three months of 1998, 87% entered during the first six months, 95% entered during the first nine months, and only 1% entered ALMP in December (for the first time in 1998). Note that various other approaches exist as well, e.g. the random start date setup (as in Lechner, 1999) or a dynamic treatment model (as in Lechner and Miquel, 2005) or the integrated hazard approach of Fredriksson and Johansson (2008). However, combining such approaches with our instrumental variables estimator would be beyond the scope of this paper.

guage), *skills and qualifications* (professional qualification, education), *employability rating* (assessment made by the caseworker about ease of finding a job and how much assistance required), *characteristics of last job* (earnings, previous job position, industry and type of last job, industry unemployment rate), *characteristics of job search* (searching for part-time or full-time job, job preferences), *unemployment and employment history* (length of current unemployment spell, unemployment and employment history in last 10 years), and *history of participation in ALMP*. The following table presents the descriptive statistics of these 67 variables (including several square and interaction terms). Those descriptive statistics reveal that the differences between participants and nonparticipants are not dramatic, although visible e.g. in the short- and long-term labour market histories. In addition, we also observed that the descriptive statistics of the 32634 individuals who live in one of the 18 local labour markets are on average not very different from the total population. Therefore, we think that our estimation results, which only make use of those individuals living in these 18 local labour markets, might roughly carry over to the remaining parts of the country.

The first rows of Table 2 show the average employment outcomes during 1999 to 2006. The average employment outcome is about 0.55, i.e between 6 to 7 months employed per year. The average monthly earnings are between 2100 and 2400 CHF.[14]

[14] The earnings data are only available for 1999. For the employment outcomes, note that these are based on the unemployment insurance data. When we construct employment indicators on the basis of the social security/pension data, which are only available for 1999, the average employment outcome is about 0.60 to 0.63, and thus higher than the numbers shown in Table 2. The impact estimates, i.e. the average difference between Y^1 and Y^0, however, are similar irrespective of the data source used.

Table 2: Descriptive statistics (means or shares multiplied by 100)

Variable name	66713		32634 individuals	
	ALMP	Non-ALMP	ALMP	Non-ALMP
Observations	28122	38591	13746	18888
Outcome variables 1999 to 2006				
Employment 1999: Number of months employed in 1999, divided by 12	0.49	0.46	0.50	0.47
Employment 2000: Number of months employed in 2000, divided by 12	0.59	0.54	0.59	0.54
Employment 2002: Number of months employed in 2002, divided by 12	0.60	0.55	0.60	0.54
Employment 2003: Number of months employed in 2003, divided by 12	0.58	0.52	0.58	0.52
Employment 2004: Number of months employed in 2004, divided by 12	0.57	0.51	0.56	0.51
Employment 2006: Number of months employed in 2006, divided by 12	0.56	0.51	0.55	0.50
Employment 1990-2006: Number of months employed in 1999-2006, divided by 96	0.57	0.52	0.57	0.51
Earnings 1999: Total earnings from employment & self-employment, divided by 12	2318	2141	2408	2222
Control variables *X*				
Age in years	38	38	38	38
older than 50 years (%)	11	10	11	11
30 years and younger (%)	23	25	22	25
Female (%)	45	42	45	43
Marital status: married (%)	59	60	58	59
single (%)	28	27	27	28
Number of (dependent) persons in household	2.4	2.4	2.4	2.4
interacted with foreigner status	1.2	1.3	1.3	1.3
interacted with marital status	1.9	1.9	1.9	1.9
Foreigner with yearly permit (%)	15	17	16	16
Swiss national (%)	58	53	57	55
Mother tongue not German, French or Italian (%)	33	37	35	37
Immigrant who migrated to Switzerland in 1988-1992 (and ≥ 25 years old then) (%)	5	6	5	5
in 1993-1997 (and ≥ 25 years old then) (%)	6	5	6	5
Number of languages known, other than mother tongue (0-3)	1.4	1.4	1.4	1.4
First foreign language is German, French or Italian (%)	64	64	62	62
English, Spanish or Portuguese (%)	14	14	16	18
Qualification: skilled (%)	58	54	58	56
semi-skilled (%)	14	16	15	17
Job position: unqualified labourer (%)	37	38	36	36
management (%)	6	5	7	7
Industry unemployment rate (January 1998, unemployment rate in percent)	6.4	6.6	6.3	6.3
Job type: office (%)	16	14	16	15
hotels, restaurant, catering (%)	15	16	15	14
construction (%)	7	8	7	8
chemistry, metal (%)	8	8	8	8
painting, technical drawing (%)	7	7	7	7
scientists, teaching, education (%)	5	4	5	4
agriculture, food processing (%)	2	3	2	3
health care (%)	3	3	3	3
management, entrepreneurs, senior officials, justice (%)	3	3	3	4
transportation, traffic (%)	3	4	3	3
Preferred job equals last job (%)	72	74	72	73
Looking for a part time job (%)	12	14	13	15

Table 2 to be continued

Table 2: … continued

Variable name	66'713		32'634 individuals	
	ALMP	Non-ALMP	ALMP	Non-ALMP
Unemployment duration in days (as of 1.1.1998)	178	160	180	165
squared (divided by 10000)	4.3	3.8	4.3	3.9
Part time unemployed (i.e. not available for a full time job) (%)	10	13	10	14
Insured earnings (CHF)	4030	3840	4130	3960
Earnings < 2000 CHF	7	10	7	10
> 6000 CHF	11	9	12	11
Never been unemployed in last 10 years (1988-1997) (%)	49	44	50	46
5 years (1993-1997) (%)	53	48	54	50
Number of unemployment spells in the period 1988-1992	0.29	0.33	0.26	0.29
last 5 years (1993-1997)	0.92	1.08	0.87	0.99
Fraction of time spent in unemployment (since first registration in pension data)	0.12	0.12	0.12	0.12
interacted with immigrant status	0.03	0.02	0.03	0.02
Duration of last employment spell (months)	44	41	45	43
Wage increase during last employment spell (last wage compared to first wage)	0.004	0.003	0.003	0.003
Number of employment spells in last 10 years (1988-1997)	2.50	2.68	2.41	2.55
Fraction of time spent in employment (since first registration in pension data)	0.79	0.77	0.79	0.78
interacted with immigrant status	0.07	0.07	0.07	0.07
Number of contribution months to unemployment insurance	18	18	18	18
Continuously increasing annual earnings (since first registration in pension data) (%)	10	10	9	9
decreasing annual earnings (since first registration in pension data) (%)	8	7	8	8
Yearly earnings 1997 (CHF)	27090	25280	27240	25440
1996 (CHF)	40520	37570	41880	38960
1995 (CHF)	39610	37510	41310	39260
Ever been self-employed in the period 1988-1992 (%)	7	8	7	7
last 5 years (1993-1997) (%)	5	5	5	5
Employability rating: unknown (%)	4	5	3	3
does not need assistance (%)	5	6	2	2
good (%)	17	16	18	16
intermediate (%)	57	55	57	56
Participated in employment programme in 1997 (%)	11	3	9	3
in a temporary wage subsidy in 1997 (%)	37	19	36	19
in training in 1997 (%)	4	3	5	5
Treatment started on 1.1.1998 (%)	13	0	12	0

Note: 1 Swiss Franc (CHF) ≈ 2/3 Euro. For non-binary variables the means are given. For binary variables (=dummies) the means multiplied by 100 are given.

Gerfin and Lechner (2002) and Gerfin, Lechner, and Steiger (2005) argue at length why in the Swiss institutional setting, it is plausible that these data sources contain all variables jointly related to treatment and potential outcomes. Thus, controlling for them is sufficient to remove confounding. Again, we do to repeat these arguments in detail, but refer the reader to their paper.

4 Empirical results

In this section, we present the estimated effects of participation in ALMP on subsequent labour market outcomes. To see the dynamics of the effects, we follow the individual labour market situation over the years 1999 to 2006 and create the following outcome variables: Employment is defined as the number of months with positive earnings in a non-subsidized job in a particular year, divided by 12. Employment in a subsidized job, e.g. temporary wage subsidies, is not counted as regular employment. Earnings are defined as total earnings from employment or self-employment during a year, divided by 12. In the following tables, we show the estimated effects on employment during the years 1999, 2003 and 2006,[15] and in addition, the average effect over the 8 years 1999 to 2006, i.e. the number of months employed between January 1999 and December 2006, divided by 96 months. The earnings estimates are only available for the year 1999.[16]

4.1 Implementation of the nonparametric estimators

All the objects of Section 2 we need to estimate are functionals of various potential means, which can be obtained via nonparametric regressions on X and subsequently taking averages. As an alternative, nonparametric regression on the propensity score is used here. For the treatment

[15] We also examined the effects for each of the years from 1999 to 2005, which did not provide additional insight beyond what is shown in the following tables.

[16] The earnings data are obtained from the social security/pension system, to which we have access only until the end of 1999. The employment data are taken from the unemployment insurance system, to which we had access until the end of 2006. We define an individual as employed if he is de-registered because of having entered employment. This definition only relies on the exit code and is thus somewhat imprecise since we cannot observe subsequent movements between employment and out-of-labour-force if no intermittent unemployment spell is registered. For the year 1999, we had employment data from *both* the unemployment insurance and the social security/pension system (i.e. earnings and earnings sources each month) and could thus cross-validate whether exits into employment and positive earnings in the pension data are jointly observed. There was a rather high correspondence, and perhaps more importantly, the treatment effects on employment were very similar irrespective of whether the employment outcomes were constructed from the pension or the unemployment insurance data.

effects for the *compliers*, and in fact all other objects defined in Section 2.1, the relevant propensity score is $\pi(x)$, within a particular labour market.

For the treatment effects for the always- and the never-participants, we additionally need estimates of $E[Y^0]$ and of $E[Y^1]$ (see Section 2), which are identified via a selection-on-observables assumption. These potential outcomes are obtained via propensity score matching where the relevant propensity score now is $p(x) = \Pr(D = 1 \mid X = x)$, within a particular labour market, i.e. the probability of participating in ALMP among all individuals within the same labour market. With an estimate $\hat{p}_i$ of $p(x_i)$, the expected potential outcomes are estimated as

$$\widehat{E[Y^0]} = \frac{1}{N}\left(\sum_{i:d_i=1} \hat{m}_0(\hat{p}_i) + \sum_{i:d_i=0} y_i\right), \tag{15}$$

$$\widehat{E[Y^0 \mid D=1]} = \frac{1}{N_1}\sum_{i:d_i=1} \hat{m}_0(\hat{p}_i), \tag{16}$$

where $m_d(\rho) = E[Y \mid p(X) = \rho, D = d]$, and analogously for $E[Y^0]$ and $E[Y^1]$.

The estimation of $E[Y^0]$, $E[Y^1]$ and $E[Y^0 \mid T = c]$, and $E[Y^1 \mid T = c]$ is done in three steps, separately for each local labour market and each outcome. First, the propensity scores $\pi(x)$ and $p(x)$ are estimated by a binary probit to obtain predicted probabilities $\hat{\pi}_i$ and $\hat{p}_i$ for all observations. Second, the conditional expectation functions $\mathrm{m}_z(\rho) = E[Y \mid \pi(X) = \rho, Z = z]$ and $\mu_z(\rho) = E[D \mid \pi(X) = \rho, Z = z]$ or $m_d(\rho) = E[Y \mid p(X) = \rho, D = d]$, respectively, are estimated via (one-dimensional) nonparametric regression on the respective propensity score. Separately for each conditional expectation function, the bandwidth value is selected by leave-one-out least squares cross-validation. Bandwidths are chosen from the expanding grid with 10 values: {1/100, 1.9/100, $1.9^2/100$, ..., $1.9^8/100$, ∞}.[17] With the selected bandwidth values, the conditional expectation functions are estimated nonparametrically and sample averages are computed to obtain estimates of $E[Y^0]$, $E[Y^1]$, $E[Y^0 \mid T = c]$, $E[Y^1 \mid T = c]$ etc. These estimates are then restricted to be within the support of the respective outcome variables, i.e. to be non-negative for earnings and to be within [0,1] for the employment variable.

For the nonparametric regression, we use nonparametric *ridge* regression, which performed best in Frölich (2004). Ridge regression is a variant of local linear regression with a ridge term added to the denominator to reduce its variance.[18] Given a sample of observations $(y_i, w_i) \in \Re \times \Re$, where y_i is the outcome variable and w_i the (one-dimensional) regressor, i.e. one of the two estimated propensity scores defined above, and a bandwidth value *h*, the ridge regression estimate at location *w* is defined as

$$\widehat{E[Y \mid W = w]} = \frac{T_{1,0}}{T_{0,0}} + \frac{T_{1,1} \cdot (w - \bar{w})}{T_{0,2} + rh \mid w - \bar{w} \mid},$$

where $T_{a,b} = \sum_i y_i^a \cdot (w_i - \bar{w})^b K\left(\frac{w_i - \bar{w}}{h}\right)$ and $\bar{w} = \sum_i w_i K\left(\frac{w_i - w}{h}\right) / \sum_i K\left(\frac{w_i - w}{h}\right)$. The ridge parameter *r* is set to 0.35 for the Gaussian kernel (see Seifert and Gasser, 1996, 2000, and Frölich, 2004).

---

17 For ease of comparison, all estimators use the same *X* variables, and the same nonparametric estimator and bandwidth search grid. The implementation of the estimator follows Frölich (2004, 2007). Bandwidth choice via cross-validation is not optimal but performed very well in the simulations of Frölich (2004). All estimates are based on the 32634 individuals who live in the 18 labour markets defined in the previous section. All estimations, including bandwidth choice, are done exclusively for each local labour market (i.e. exact match on labour market).

18 Conventional local linear regression estimators often perform poorly. For curve estimation, this is observed by Seifert and Gasser (1996) and for matching estimation by Frölich (2004).

4.2 Aggregated treatment effects

For each of the five different outcome variables defined before, we estimate the potential outcomes separately for each of the 18 local labour markets. This leads to a large number of estimates, which are displayed in Tables 1 to 10 in the supplementary appendix. (One table for each outcome variable for the treatment definition window January to March, and again for the alternative treatment definition window January to April.) This large number of estimates makes it difficult to find any discernible patterns, and many of the estimates are very noisy due to the small number of observations in most labour markets. To reduce the dimensionality of the estimates and to increase statistical precision, we will compute weighted averages of the estimated outcomes across these 18 labour markets. These aggregated effects are self-weighted averages for the populations in these labour markets. More precisely, the average potential outcomes for the compliers are obtained by weighting the 18 estimates with the *number of compliers* (i.e. the fraction of compliers multiplied with the sample size) in each labour market. The average potential outcomes for the always-participants are obtained by weighting with the *number of always-participants*, and analogously for the never-participants.[19] Table 3 presents the average effects of ALMP on employment and earnings, in addition to bootstrap standard errors.[20] We further test whether the

---

[19] Hence, for the compliers the average outcome is a weighted average of all 18 labour markets, where the weights are $w_c = \Pr(T = c) \cdot N$ where N is the number of observations in the respective labour market. The weights for the always-participants are $w_a = \Pr(T = a) \cdot N$, and analogously for the never-participants. We also examined alternative weighting schemes, where we used the sample size only as weight. The results were similar.

[20] 499 bootstrap replications. The nonparametric bootstrap proceeded by drawing with replacement from the original sample with the 66713 observations and repeating the entire estimation process. The probits are estimated by maximum likelihood augmented with the following features to deal with collinearity problems that might occur during the bootstrapping. 1) All regressors without variation are dropped. 2) All regressors that cause local multicollinearity are dropped. For detecting (nearly) linear dependencies in the regressor matrix, the pivotal orthogonal-triangular (QR) decomposition is used, see Judd (1998, p. 58f) or Press, Flannery, Teukolsky, and Vetterling (1986, p. 357ff). This decomposition decomposes a regressor or moment matrix into an orthogonal matrix Q and an upper

treatment effects are different from zero, and whether the treatment effect for the compliers is statistically significantly different from the effect for the always-participants, or for the never-participants. (Two-sided bootstrap test of an equal effect.) Significance levels rely on the percentiles of the estimates.[21] From this table, we first confirm the main result of Frölich and Lechner (2010): The one-year-after treatment effect on the compliers is 0.155, which corresponds to a little less than *two months* of additional employment during the year 1999. This effect is rather similar to the findings in Frölich and Lechner (2010). The earnings effect is 57 CHF and thus smaller than in Frölich and Lechner (2010), but both cases are very noisy estimates, and given this large uncertainty, we cannot draw any conclusions about earnings.

In addition, we obtain three new results: First, the positive treatment effects for the compliers are not short-lived. The effects are positive for employment in 2003 and 2006 (and in fact for every other year, not shown). Furthermore, the average effect over the 8 years from 1999 to 2006 is 0.163, thus positive, and very similar to the short-term effect of 0.155. Hence, for the compliers participation in ALMP has a long-lasting effect.

Second, the treatment effects for the always- and never-participants are much smaller than for the compliers, albeit still mostly positive. The effects are somewhat (and often statistically significantly) greater for the always- than for the never-participants. We also observe that the me-

---

triangular matrix R, where diagonal elements of *R* that are close to zero indicate (nearly) linear dependencies attributable to the corresponding columns. All regressors associated with a diagonal element in *R* smaller than 10^{-5} are dropped in the local regression. (Different threshold values have been tried and did not affect the results very much. 10^{-5} is a conservative choice, in the sense that rather more than less regressors are dropped to spare local degrees of freedom for estimating the remaining coefficients.) 3) Furthermore, regressors with coefficients diverging towards infinity are dropped.

21 The bootstrap standard errors should be dealt with caution since the finite-sample standard errors might be infinite (e.g. for the earnings estimate) as it is well known that the conventional IV estimator (without over-identification)

dium-term effects are somewhat greater than the short-term effects. Hence, after an initial lock-in period, the participation in ALMP is also beneficial for the always-participants and perhaps also for the never-participants, but in any case much less than for the compliers.

Third, when comparing the potential outcomes $E[Y^0]$, we observe that these numbers are biggest for the never-participants, followed by the always-participants and are smallest for the compliers. This ordering is the same for every outcome variable, except for the earnings estimates. (This ordering is also the same in Table 4 below.) Hence, on average, the group of never-participants contains the good-risks, i.e. those who can find a job on their own, whereas the compliers are the worst group, and are least likely to find a job without ALMP. In other words, the always- and never-participants consist of unemployed people who have better chances on the labour market, or accept more job offers, than the compliers. Perhaps for this reason, the effect of ALMP is small for the always- and never-participants.[22]

In this sense, it seems that the (external) introduction of the quota was actually effective in terms of *targeting*: Increasing the quota let the worst-off people participate in ALMP, who then actually benefitted from it. On the other hand, the priority orderings of the caseworkers themselves seems not to have been as effective due to the treatment effects on the always-participants being rather small. The always-participants would have been sent to the programme even if the quota had been lower (i.e. without the external pressure). Hence, some external pressure could be helpful to overcome incorrect beliefs of the caseworkers about who benefits most from ALMP. (Had the caseworkers been effective in targeting those unemployed who benefit most from the programmes,

does not have finite moments if errors are normal. A similar problem might well exist for the nonparametric CIV estimator. Therefore, we do not use them.

[22] Interestingly, the Y^0 earnings outcomes for the compliers suggest that, although they may not be as successful in finding jobs, if they find them, they are better paid than the other two groups. Alternatively, they may have higher reservation wages that led them to reject (or not receive) job offers that the two other groups accepted.

the treatment effects should have been biggest for the always-participants and smallest for the never-participants.)

Table 3: Estimates for compliers, always- and never-participants, treatment window 3 months

	Employment 1999	Employment 2003	Employment 2006	Employment 1999-2006	Earnings 1999
Always-participants					
$E[Y^1\|T=a]$	0.502	0.579	0.550	0.572	2419
$E[Y^0\|T=a]$	0.473	0.526	0.508	0.522	2318
$E[Y^1-Y^0\|T=a]$	* 0.029 (0.019)	*** 0.053 (0.021)	** 0.042 (0.022)	*** 0.049 (0.019)	100 (123)
Never-participants					
$E[Y^1\|T=n]$	0.481	0.557	0.543	0.554	2281
$E[Y^0\|T=n]$	0.503	0.546	0.520	0.545	2245
$E[Y^1-Y^0\|T=n]$	*** -0.019 (0.013)	0.011 (0.013)	* 0.023 (0.013)	0.011 (0.012)	55 (67)
Compliers					
$E[Y^1\|T=c]$	0.548	0.609	0.542	0.584	2788
$E[Y^0\|T=c]$	0.393	0.399	0.422	0.421	2731
$E[Y^1-Y^0\|T=c]$	*** 0.155 (0.108)	*** 0.210 (0.113)	* 0.120 (0.121)	*** 0.163 (0.108)	57 (12317)
Are treatment effects statistically different ?					
$E[\Delta\|T=c] = E[\Delta\|T=a]$	**	***		**	
$E[\Delta\|T=c] = E[\Delta\|T=n]$	***	***		***	
$E[\Delta\|T=a] = E[\Delta\|T=n]$	**	***		**	

Note: An unemployed is defined as treated if entering in ALMP between January to March 1998. The employment outcomes refer to number of months employed per year divided by 12. Earnings refers to monthly earnings in CHF. The aggregated potential outcomes and treatment effects are given for the always- and never-participants and for the compliers. Bootstrap standard errors in parentheses. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively. Inference is obtained from bootstrapping the estimate (percentile method). 67 regressors. 18 local labour markets. 32634 observations.

In Table 4 we show the estimates when using a treatment definition window of four months instead of three months, i.e. an unemployed is defined as treated if he/she entered ALMP during January to April 1998. Overall, we obtain rather similar conclusions as from Table 3.

Table 4: Estimates for compliers, always- and never-participants, Treatment window 4 months

	Employment 1999	Employment 2003	Employment 2006	Employment 1999-2006	Earnings 1999
Treatment window: January to April 1998 (= 4 months)					
Always-participants					
$E[Y^1\|T=a]$	0.492	0.565	0.543	0.563	2394
$E[Y^0\|T=a]$	0.481	0.535	0.519	0.534	2333
$E[Y^1-Y^0\|T=a]$	0.011 (0.017)	0.030 (0.018)	0.024 (0.020)	0.029 (0.017)	61 (106)
Never-participants					
$E[Y^1\|T=n]$	0.491	0.553	0.533	0.559	2204
$E[Y^0\|T=n]$	0.514	0.550	0.525	0.551	2291
$E[Y^1-Y^0\|T=n]$	** - 0.024 (0.015)	- 0.001 (0.015)	0.007 (0.014)	0.004 (0.014)	- 79 (75)
Compliers					
$E[Y^1\|T=c]$	0.455	0.630	0.574	0.537	3090
$E[Y^0\|T=c]$	0.354	0.360	0.357	0.358	2210
$E[Y^1-Y^0\|T=c]$	0.102 (0.111)	*** 0.270 (0.114)	*** 0.218 (0.117)	*** 0.179 (0.110)	880 (12395)
Are treatment effects statistically different ?					
$E[\Delta\|T=c] = E[\Delta\|T=a]$		***	***	***	
$E[\Delta\|T=c] = E[\Delta\|T=n]$		***	***	***	
$E[\Delta\|T=a] = E[\Delta\|T=n]$		*			

Note: An unemployed is defined as treated if entering in ALMP between January to April 1998. See note below previous table.

Finally, in Table 5 we compare the treatment effects for the compliers and for the treated compliers. (The results for the compliers are reproduced from Tables 3 and 4.) Note that these effects differ only because of differences in the distributions of X among the treated and non-treated compliers. (In Imbens and Angrist (1994), the effects for the treated and non-treated compliers are identical since the IV was assumed to be valid without the need to condition on any X.) Overall we find that the effects on the treated compliers tend to be somewhat larger than for the non-treated compliers. In this sense, the targeting was again successful in that of all compliers, on average those who benefited more from ALMP, received it.

Table 5: Estimates for compliers and treated compliers

	Employment 1999	Employment 2003	Employment 2006	Employment 1999-2006	Earnings 1999
Treatment window: January to March 1998 (= 3 months)					
$E[Y^1-Y^0 \mid T=c]$	*** 0.155 (0.108)	*** 0.210 (0.113)	* 0.120 (0.121)	*** 0.163 (0.108)	57 (12317)
$E[Y^1-Y^0 \mid D=1,T=c]$	** 0.142 (0.113)	*** 0.252 (0.113)	** 0.175 (0.124)	*** 0.205 (0.109)	109 (107905)
Treatment window: January to April 1998 (= 4 months)					
$E[Y^1-Y^0 \mid T=c]$	0.102 (0.111)	*** 0.270 (0.114)	*** 0.218 (0.117)	*** 0.179 (0.110)	880 (12395)
$E[Y^1-Y^0 \mid D=1,T=c]$	*** 0.142 (0.109)	*** 0.292 (0.117)	*** 0.295 (0.121)	*** 0.307 (0.112)	913 (14990)

Note: See note below previous table.

5 Conclusions

In this paper, we proposed a fully nonparametric method to identify potential outcomes not only for compliers but also for always- and never-treated. Learning about the potential outcomes of always- and never-treated is important since in many applications the compliers comprise only a very small subpopulation. These potential outcomes and treatment effects can be estimated by a combination of IV and matching estimators in cases when the no-confounding (conditional independence) assumptions holds and a instrument can be observed as well.

These methods have then been used to evaluate the effects of active labour market policies in Switzerland. We found positive and long-lasting employment effects of ALMP for the compliers. The effects on the always- and never- participants were much smaller, but still mostly positive.

Furthermore, the comparison of the estimated potential outcomes showed that, on average, the never-participants had the best chances to find a job, even *without* ALMP, followed by the always-participants and finally by the compliers. Hence, the compliers were the group with the worst chances on the labour market, and at the same time, those with the largest treatment effect.

Hence, the initial selection by the caseworkers was not efficient in the sense that they neither send those unemployed to treatment who benefited most from it nor picked those with the worst employment prospects. On the other hand, the expansion of ALMP was effective in reaching those unemployed who benefited strongly from it and, in addition, would have been worst off otherwise.

References

Fredriksson, P., and P. Johansson (2008): "Program Evaluation and Random Program Starts," *Journal of Business and Economics Statistics*, 26, 435-445.

Frölich, M. (2004): "Finite Sample Properties of Propensity-Score Matching and Weighting Estimators," *Review of Economics and Statistics*, 86, 77-90.

Frölich, M. (2007): "Nonparametric IV estimation of local average treatment effects with covariates," *Journal of Econometrics*, 139, 35-75.

Frölich, M. (2008): "Parametric and Nonparametric Regression in the Presence of Endogenous Control Variables," *International Statistical Review*, 76, 214-227.

Frölich, M. and M. Lechner (2010): "Exploiting regional treatment intensity for the evaluation of labour market policies," *Journal of American Statistical Association,* in press.

Gerfin, M., and M. Lechner (2002): "Microeconometric Evaluation of the Active Labour Market Policy in Switzerland," *The Economic Journal*, 112, 854-893.

Gerfin, M., M. Lechner, and H. Steiger (2005): "Does subsidised temporary employment get the unemployed back to work? An econometric analysis of two different schemes," *Labour Economics – An International Journal*, 12, 807-835.

Heckman, J.J. (1997): "Instrumental Variables", *Journal of Human Resources*, 32, 441-462.

Heckman, J., H. Ichimura, and P. Todd (1998): "Matching as an Econometric Evaluation Estimator," *Review of Economic Studies*, 65, 261-294.

Heckman, J. and E. Vytlacil (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings National Academic Sciences USA, Economic Sciences*, 96, 4730-4734.

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

Hirano, K., G. Imbens, and G. Ridder (2003): "Efficient Estimation of Average Treatment Effects using the Estimated Propensity Score," *Econometrica*, 71, 1161-1189.

Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, 86, 4-29.

Imbens, G.W., and J.D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-475.

Judd, K. (1998), *Numerical Methods in Economics*, MIT Press, Cambridge.

Kluve J. (2006): "The Effectiveness of European Active Labour Market Policy," IZA DP No. 2018.

Kluve, J., and C. Schmidt (2002): "Can training and employment subsidies combat European unemployment?," *Economic Polic*y, 35, 411-448.

Lalive, R., J. C. van Ours, and J. Zweimüller (2008): "The Impact of Active Labour Market Programs on the Duration of Unemployment," *The Economic Journal*, 118, 235-257.

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

Lechner, M. and R. Miquel (2005): "Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions," forthcoming in *Empirical Economics*, doi: 10.1007/s00181-009-0297-3.

Lechner, M., R. Miquel, and C. Wunsch (2005): "Long-run effects of Public Sector Sponsored Training in West Germany," CEPR discussion paper 4851, forthcoming in *Journal of the European Economic Association*.

Maddala, G.S., and J. Jeong (1992): "On the Exact Small Sample Distribution of the Instrumental Variable Estimator," *Econometrica*, 60, 181-183.

Manski, C. F. (2003), *Partial Identification of Probability Distributions*, New York: Springer Verlag.

Martin, J. P., and D. Grubb (2001): "What Works and for Whom: A Review of OECD Countries' experiences with active labour market policies," *Swedish Economic Policy Review*, 8(2), 9-56.

on Opportunity: State Laws and Mortgage Credit," *Review of Economics and Statistics*, 88, 177-182.

Press, W., B. Flannery, S. Teukolsky, and W. Vetterling (1986), *Numerical Recipes*, Cambridge University Press, Cambridge.

Rosenbaum, P.R., and D.B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-50.

Seifert, B., and T. Gasser (1996): "Finite-Sample Variance of Local Polynomials: Analysis and Solutions," *Journal of American Statistical Association*, 91, 267-275.

Seifert, B., and T. Gasser (2000): "Data Adaptive Ridging in Local Polynomial Regression," *Journal of Computational and Graphical Statistics*, 9, 338-360.

Appendix - Proofs of identification

Proof of expression (4): Notice first that in the subpopulation of compliers, conditioning on D =1 is equivalent to $Z=\overline{z}$ and conditioning on D=0 is equivalent to $Z=\underline{z}$. It follows that

$$\begin{aligned} E[Y^1-Y^0 \mid D=1,T=c] &= \int E[Y^1-Y^0 \mid X,Z=\overline{z},T=c]\cdot dF_{X|Z=\overline{z},T=c} \\ &= \int E[Y^1-Y^0 \mid X,T=c]\cdot dF_{X|Z=\overline{z},T=c} \end{aligned}$$

where the last equality follows from the exclusion restriction CIV.4.

By Bayes' theorem the conditional distribution of X can be written as

$$\begin{aligned} dF_{X|Z=\overline{z},T=c} &= \frac{\Pr(Z=\overline{z},T=c \mid X)\cdot dF_X}{\Pr(Z=\overline{z},T=c)} = \frac{\Pr(T=c \mid X,Z=\overline{z})\cdot \pi(X)\cdot dF_X}{\int \Pr(Z=\overline{z},T=c \mid X)dF_X} \\ &= \frac{\Pr(T=c \mid X)\pi(X)dF_X}{\int \Pr(T=c \mid X)\pi(X)dF_X} \end{aligned}$$

by the unconfounded type assumption CIV.3. By assumption CIV.1 to CIV.5 we can also show that via iterated expectations over the subpopulations complier, always- and never-treated

$$E[D \mid X, Z=\overline{z}] - E[D \mid X, Z=\underline{z}] = \Pr(T=c \mid X)\text{, and}$$

$$E[Y \mid X, Z=\overline{z}] - E[Y \mid X, Z=\underline{z}] = E[Y^1 - Y^0 \mid X, T=c] \cdot \Pr(T=c \mid X)\,.$$

Combining these results, we obtain

$$E[Y^1 - Y^0 \mid D=1, T=c] = \frac{\int E[Y \mid X, Z=\overline{z}] - E[Y \mid X, Z=\underline{z}]\pi(X)dF_X}{\int E[D \mid X, Z=\overline{z}] - E[D \mid X, Z=\underline{z}]\pi(X)dF_X}\,. \tag{17}$$

Proof of equivalence between (8) and (2): Notice that for compliers conditioning on D and conditioning on Z are equivalent. Therefore, by exploiting the assumption of conditional independence for the compliers, the expression (8) can be written as

$$E[Y^1 - Y^0 \mid T=c] = \int \left(E[Y^1 \mid X, D=1, T=c] - E[Y^0 \mid X, D=0, T=c] \right) dF_{X|T=c}$$

$$= \int \left(E[Y^1 \mid X, Z=\overline{z}, T=c] - E[Y^0 \mid X, Z=\underline{z}, T=c] \right) dF_{X|T=c}$$
$$= \int \left(E[Y^1 \mid X, T=c] - E[Y^0 \mid X, T=c] \right) dF_{X|T=c}$$

by the exclusion restriction.

By Bayes' theorem the conditional distribution of X can be written as

$$dF_{X|T=c} = \frac{\Pr(T=c \mid X) \cdot dF_X}{\Pr(T=c)} = \frac{\Pr(T=c \mid X) \cdot dF_X}{\int \Pr(T=c \mid X) dF_X}$$

by the unconfounded type assumption CIV.3. Now using the previously obtained results that

$$E[D \mid X, Z=\overline{z}] - E[D \mid X, Z=\underline{z}] = \Pr(T=c \mid X)\text{, and}$$

$$E[Y \mid X, Z=\overline{z}] - E[Y \mid X, Z=\underline{z}] = E[Y^1 - Y^0 \mid X, T=c] \cdot \Pr(T=c \mid X)\,,$$