

COMBINING MATCHING AND NONPARAMETRIC IV ESTIMATION: THEORY AND AN APPLICATION TO THE EVALUATION OF ACTIVE LABOUR MARKET POLICIES

Markus Frölich

Economics Department, University of Mannheim

Michael Lechner

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

This version: 11 June 2014

Abstract: 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, representing a subset of the entire population. By combining IV with matching, we can estimate the treatment effects for the always- and never-takers as well. Since in many cases, these groups are the (endogenous) outcome of some assignment process, such estimates also help in judging the implications of such a selection process. In our application to the effects of participation in active labour market programmes in Switzerland, we find large and lasting positive employment effects for the compliers, whereas the effects for the always- and never-participants are small. In addition, the compliers have worse employment outcomes without treatment than those who participate in the programme with or without the intervention under investigation. This suggests that the earlier assignment policy of the caseworkers was inefficient in that the always-participants were neither those unemployed who would experience the highest expected treatment effects nor those unemployed who had the largest need for assistance.

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

JEL classification: C14, C2, J68

Addresses for correspondence

Markus Frölich, Economics Department, Universität Mannheim, froelich@uni-mannheim.de

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

1 Introduction¹

Analysing heterogeneity of treatment effects is a major aim of many empirical studies. The extent to which average treatment effects as well as their heterogeneity are identified critically depends on the strength of the identifying assumptions. This paper adds to the literature on nonparametric identification of treatment effect heterogeneity when a binary instrumental variable as well as informative data on the selection process are available. With the instrumental variable, the local average treatment effect can be identified. This, however, only provides an estimate for the complier subpopulation, which roughly speaking represents the population at the margin of participation in a treatment. If, in addition, a conditional independence assumption for treatment assignment and potential outcomes can be exploited, the treatment effects for those who are not at the margin of participation, i.e. the always- and never-takers, are identified as well.

Such additional information is valuable. On the one hand, it allows to understand whether the particular policy change (that is the basis for the instrument used) shifted a group of individuals into the treatment because they benefited more (or less) than those individuals not affected by the policy change. In many situations, it will also be interesting to learn whether individuals, who partici-

¹ This paper emerged from the IZA Discussion paper 2144 (2006). Parts of the discussion paper are published in Frölich and Lechner (2010). However, neither the combination of IV with matching nor the *long-term effects* of ALMP, which are the topics of this paper, are published so far.

Markus Frölich is also affiliated with IZA, Bonn and IFAU, Uppsala. Michael Lechner has affiliations with CEPR, London, CES-ifo, Munich, IZA, Bonn, and PSI, London. The first author acknowledges financial support from the Research Center (SFB) 884 “Political Economy of Reforms” Project B5, funded by the German Research Foundation (DFG). The second author acknowledges financial support from the Swiss National Science Foundation (projects 4043-058311 and 4045-050673) 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

pate irrespective of the policy captured by the instrument, gain more by the treatment than those who never participate irrespective of the value of this instrument. On the other hand, the absence of such heterogeneity implies that the local average treatment effect may be generalizable to a larger population than just the one captured by compliance to this particular binary instrument.

We apply this new concept to the evaluation of the long-term effects of active labour market programmes in Switzerland. The Swiss administrative data system contains very detailed register data including subjective caseworker assessments. Therefore, a conditional independence assumption appears reasonable, see Gerfin and Lechner (2002) and Gerfin, Lechner, and Steiger (2005). In addition, Frölich and Lechner (2010) showed that a particular institutional rule assigning regional quotas could be exploited as an IV in this setting. In this paper, we combine both identification strategies to provide additional policy relevant information about labour market policies. In particular, we provide evidence on long-term effects and on heterogeneity in employment outcomes.

Learning about heterogeneity is indeed important for many policy considerations. If effects are heterogeneous, instrumental variable estimators and matching estimators estimate effects for different populations: IV estimates only (local) average treatment effects for the compliers (LATE). Matching estimators however estimate average population effects, separately for participants and non-participants. While the latter parameters also permit some judgement about whether individuals in or out of the programme benefit more from participation, the implicit policy change examined is usually too extreme to be a relevant policy option. In other words, the matching parameters are helpful for a policy decision about whether one should completely abolish the programmes (average treatment effect on the treated, ATET), or force everybody into it (average treatment effect on the non-treated). The average treatment effect (ATE) compares a situation where every-

Steiger for their substantial input in preparing them. We are also indebted to four anonymous referees who provided very valuable comments and suggestions on a previous version of this paper.

body or nobody participates. However, usually policy options are incremental, i.e. extending or reducing programmes only to some extent. For example, to our knowledge, no European country has completely abolished its active labour market programmes. However, many countries changed the size of their programmes over time to some extent. To investigate whether it is worthwhile to change the size of the programme, a LATE parameter that shifts exactly the population in or out of the programme corresponding to the policy change is the object of interest. Heckman and Vytlacil (1999, 2005) provided the analytical framework for such analysis with their concept of the *marginal treatment effect* (MTE). They show that the MTE is identified as the limit of an IV estimate. They also show that by integrating the MTE appropriately, one obtains the LATE, the ATE, the ATENT, the ATET, as well as other treatment parameters. The marginal treatment effect framework is thus a versatile tool, which recovers average effects *and* their entire heterogeneity. Its applicability, however, rests on two conditions that are very demanding in practice: First, a continuous IV is required. Second, in order to estimate all effects, the IV has to be strong enough to be able to move every individual into treatment as well as out of treatment. In most applications, these conditions are not satisfied, and we examine the situation where no continuous, powerful IV is available. Instead, we focus on the case of a *binary* instrumental variable, which is common in IV based observational studies. It is also relevant for randomized trials with imperfect compliance. From this perspective, our approach could be considered as an adaptation of the marginal treatment effects framework to the case when identification has to rely on a binary instrument. In the continuous instrument case, infinitely many complier groups can be ordered according to their inclination to participate in the treatment. If the instrument is sufficiently strong, every individual could be made to comply for some values of the instrument. In the binary IV setup, the marginal treatment effect framework reduces to only *three* identifiable groups: The never-treated are the individuals least inclined to participate and for whom the change in the instrument is not strong

enough to make them participate. The always-treated are the individuals most inclined to participate and for whom the change in the instrument is not strong enough to make them not participate. Finally, the compliers are the intermediate group with respect to their inclination to participate in the treatment. This intermediate group can be induced to change the participation status through the instrument. Thus, compared to the MTE, our approach aims at a more modest set of treatment effects that are to be identified for the case when only a less informative instrument is at hand.

The estimates of the treatment effects for the compliers, always- and never-takers, obtained via a combination of nonparametric IV and matching estimators, can be interpreted as *approximations* to large policy changes. While the complier treatment effect usually corresponds to a small change in the policy, such as an extension of active labour market programmes in our application, the effects for the always- and never-takers approximate the implications of larger policy changes. In our example, being able to identify the effects for always-takers allows us to learn something about the likely results of a large contraction of the programmes. Similarly, the effects for the never-takers give some indication about the effects of a large expansion of the programmes.

As mentioned above we apply the previously discussed identification results to the evaluation of Swiss active labour market policies and make three contributions to the literature (for recent surveys see e.g. Kluge, 2010, and Lalivé, van Ours, and Zweimüller, 2008). First and independent of the methodological issues, we find that the positive programme effects are longer lasting, for at least 8 years. The positive effects are thus long-term and confirm, e.g., the results for Germany of Lechner, Miquel, and Wunsch (2011), who also found persistent long-term effects. Second, the results reveal that the assignment policy before the policy-change was not effective in the sense of selecting unemployed either with the highest effects or with the most pressing needs for such programmes. Instead, we find that always-participants have lower, albeit positive, programme effects than compliers. Considering their re-employment chances in the absence of the programmes, al-

ways-takers appear to be the ‘good risks’ among the unemployed, which is consistent with cream-skimming behaviour by the caseworkers. Third, the extension of the programmes stimulated by the central government through the participation quota, however, had a positive effect on targeting. The extension reached those unemployed (i.e. these are the ‘compliers’) with large positive effects and poor employment chances without assistance.²

2 Combination of nonparametric IV and matching

2.1 Identification of potential outcomes for compliers, always- and never-takers

Consider a non-separable model of the type

$$Y = \varphi(D, U_Y, U_{YD}, U_{YZ}, U_{YDZ}),$$

$$D = \xi(Z, U_D, U_{YD}, U_{DZ}, U_{YDZ}), \quad (1)$$

$$Z = \zeta(U_Z, U_{YZ}, U_{DZ}, U_{YDZ}),$$

were φ, ξ, ζ are unknown functions and $U_Y, U_D, U_Z, U_{YD}, U_{YZ}, U_{DZ}, U_{YDZ}$ are *mutually independent* random variables. The construction as mutually independent variables clarifies that there are some factors U_Y that only affect the variable Y , some factors U_D that *directly* only affect D , some factors U_{YD} that directly affect Y and D , some factors U_{YDZ} that directly affect Y and D and Z , etc. In addition, we suppose that the function ξ is weakly *monotone* (increasing) in its first argument, which we will refer to as the monotonicity assumption later.

² Our paper bears similarities to Frumento, Mealli, Pacini, and Rubin (2012), who also examine different complier and non-complier groups in an IV context and characterize unobservable subpopulations, although for randomized assignment and in a likelihood context. They also found heterogeneous treatment effects and concluded that the Job Corps programme should have been better targeted. They found positive treatment effects of Job Corps for the always-employed compliers, but no positive effects for the more disadvantaged groups.

Y is the outcome variable of interest, D is the treatment variable, and Z is exploited as an instrumental variable. The model already embodies an exclusion restriction, i.e. Z does not enter the first equation.³ The treatment variable D is *binary*. We denote by Y^1 and Y^0 the potential outcomes in case of participating ($D=1$) or not participating ($D=0$) in the treatment. The potential outcomes for individual i are given as $Y_i^d \equiv \varphi(d, U_{i,Y}, U_{i,YD}, U_{i,YZ}, U_{i,YDZ})$. In our application, D will refer to participation in an active labour market programme. The outcome variable Y will denote the labour market status several years later. Z is the instrumental variable. We focus on the case where $Z \in \{\underline{z}, \bar{z}\}$ is *binary*, which is a situation frequently encountered in many applications, including randomized trials in which treatment *offer* is randomly assigned.

We define $D_i^z \equiv \xi(z, U_{i,D}, U_{i,YD}, U_{i,DZ}, U_{i,YDZ})$ as the potential treatment states of individual i if the level of the instrument were exogenously set to z . With the instrument taking only two different values, the potential treatment states $D_i^{\underline{z}}$ and $D_i^{\bar{z}}$ define four different types of individuals: Individuals with $D_i^{\underline{z}} = D_i^{\bar{z}} = 1$ are called always-treated (a), those with $D_i^{\underline{z}} = D_i^{\bar{z}} = 0$ are called never-treated (n), those with $D_i^{\underline{z}} = 0, D_i^{\bar{z}} = 1$ are called compliers (c), and those with $D_i^{\underline{z}} = 1, D_i^{\bar{z}} = 0$ are called defiers (d). As a shortcut notation for the type, we use the symbol $T_i \in \{a, n, c, d\}$. Note that the type depends on the instrument and the values $\{\underline{z}, \bar{z}\}$ used. The monotonicity assumption for ξ already implies the non-existence of defiers. Model (1) implies that the type is a function of $U_{i,D}, U_{i,YD}, U_{i,DZ}, U_{i,YDZ}$. To simplify the following expressions, we define, $\underline{z} = 0$ and $\bar{z} = 1$.

³ This exclusion restriction is required for identification via instrumental variables. It would not be necessary in a selection-on-observables approach. The model actually embodies the more general case if one were to permit that U_{YD} contains the variable Z . However, there would be no exclusion restriction then. In order to permit instrumental variable identification, we make the exclusion restriction explicit in the notation throughout.

If the variables U_{YZ}, U_{DZ}, U_{YDZ} had no impact on the observed variables, we would have that

$(Y^d, T) \perp\!\!\!\perp Z$, i.e. the pair Y^d and T is independent of Z . Together with the assumption that the function ξ is (weakly) monotonous in its first argument, which implies no defiers, we obtain the result of Imbens and Angrist (1994), who have shown that the effect for *compliers* is identified as:⁴

$$E[Y^1 - Y^0 | T = c] = \frac{E[Y | Z = 1] - E[Y | Z = 0]}{E[D | Z = 1] - E[D | Z = 0]}. \quad (2)$$

If, on the other hand, the variables U_{YD}, U_{YZ}, U_{DZ} did not exist (or had no impact on observed variables), we would have $Y^d \perp\!\!\!\perp D$ and obtain the ATE $E[Y^1 - Y^0]$ by a simple regression of Y on D .

In most observational studies, however, there are factors such as $U_{YD}, U_{YZ}, U_{DZ}, U_{YDZ}$, which lead to confounding of the instrument and/or the treatment variable. If we were to observe the variables U_{YZ}, U_{DZ}, U_{YDZ} , which for convenience of notation we define as

$$X_1 = (U_{YDZ}, U_{YZ}, U_{DZ}),$$

we obtain that $(Y^d, T) \perp\!\!\!\perp Z | X_1$. As we will show below, this will permit us to nonparametrically identify $E[Y^1 - Y^0 | T = c]$ and, in fact, $E[Y^1 | T = c]$, and $E[Y^0 | T = c]$ as well.⁵

⁴ In other words, Angrist and Imbens (1994) required the instrument Z to be unconfounded. Such an assumption is reasonable when the instrument Z has been 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_1 . (See also Abadie (2003) and Frölich (2007) for a similar extension.) 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 these characteristics the IV assumption appears reasonable.

⁵ Note that choosing X_I as $X_1 = (U_{YDZ}, U_{YZ}, U_{DZ})$ is not the only choice that guarantees $(Y^d, T) \perp\!\!\!\perp Z | X_1$. Other choices of X_I may also imply independence of instrument Z . (X_I may also be related to unobservables as in Frölich, 2008).

If we also were able to observe U_{YD}, U_{YZ}, U_{YDZ} , which for convenience of notation we define as

$$X_2 = (U_{YDZ}, U_{YZ}, U_{YD}),$$

we obtain that $Y^d \coprod D | X_2$. This will allow to also identify the potential outcomes for the non-compliers $E[Y^0 | T = a]$, $E[Y^1 | T = a]$, $E[Y^0 | T = n]$, and $E[Y^1 | T = n]$. Thus, it enables us to learn something about the always- and the never-participants.

It is this combination of selection-on-observables and IV assumptions that permits identification of $E[Y^0 | T = a]$ and $E[Y^1 | T = n]$. Instrumental variable estimation alone would only deliver us the potential outcomes for the compliers. Identifying the potential outcomes for all groups will not only permit us to compare their treatment effects $Y^1 - Y^0$, but the 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 and always-treated have on average a larger non-treatment outcome, Y^0 , than the compliers. Hence, in our population of unemployed, this indicates that the never-treated and always-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.

As already mentioned in the introduction, there is an intuitive relationship to the marginal treatment effects framework of Heckman and Vytlacil (1999, 2005), because the separate effects for the always-treated, never-treated, and compliers can be seen as an analogue to the marginal treatment effects for the case when identification has to rely on discrete instruments. This is so, because the effects are obtained separately for those most inclined to treatment, those least inclined to treatment, and for the middle group which is becoming treated because of the discrete change in the instrument. If the instrument is discrete but takes on more than two different values, there

would be several complier groups for each pair of these values, in addition to the groups not affected by the treatment. The potential outcomes can be identified for each of these groups. If the instrument takes many different values, there will be many different complier groups that can be ordered according to their treatment inclination. For the limit case where the instrument becomes continuous, the marginal treatment effect framework is obtained. While the information on effect heterogeneity in the marginal treatment effect framework is certainly much richer than in our case, in many applications no continuous instrument is available so that treatment effect heterogeneity can only be assessed for a finite number of groups.

Our main identification result can be expressed in different ways. In Theorem 1, we show a type of propensity score weighting representation, which represents the most concise form of presenting the identification results. In Corollaries 1 and 2 (in the appendix) we express the identification in a form of a matching or propensity-score matching representation. In Theorem 1, we use two ‘propensity scores’: Define the functions $\pi(x_1) = E[Z | X_1 = x_1]$ and $p(x_2) = E[D | X_2 = x_2]$. Furthermore, define the random variables Π and P as $\Pi \equiv \pi(X_1)$ and $P \equiv p(X_2)$. While P is the usual propensity score, i.e. the conditional probability to participate in treatment, the ‘instrument propensity score’ Π refers to the probability that the instrument takes a particular value. Roughly speaking, while the propensity score Π exploits the exclusion restriction and identifies potential outcomes for compliers, the propensity score P is based on the selection-on-observables argument. For identifying potential outcomes for always- and never-treated both scores are needed.

In addition to the identification results for the always- and never-treated we also obtain a finer identification result for the compliers. The population of compliers consists of two subpopulations: those who actually receive treatment and those who do not. If the variables U_{YZ}, U_{DZ}, U_{YDZ} did not exist, the treatment effects for treated and non-treated compliers would be identical. However,

with the need to control for U_{YZ}, U_{DZ}, U_{YDZ} the treatment effects on the treated compliers (LATET) and on the non-treated compliers (LATEN) are different.

For identifying all potential outcomes in Theorem 1, including $E[Y^0 | T = a]$ and $E[Y^1 | T = n]$, a common support assumption with respect to both propensity scores is required, i.e. we need $0 < \Pi < 1$ and $0 < P < 1$ almost surely, which means we need that $Supp(X_1 | Z = 1) = Supp(X_1 | Z = 0)$ and $Supp(X_2 | D = 1) = Supp(X_2 | D = 0)$. This condition can be verified, for example, by plotting densities of the estimated $\hat{\Pi}$ and \hat{P} . If the support condition is not satisfied, we can restrict our population parameter to a subset χ with

$$\chi = \{(x_1, x_2) : f_{X_1|Z=1}(x_1) \cdot f_{X_1|Z=0}(x_1) \cdot f_{X_2|D=1}(x_2) \cdot f_{X_2|D=0}(x_2) > 0\},$$

i.e. where all the previous conditional densities are larger than zero. In this case, all expected values in Theorem 1 are defined with respect to this set χ .

Theorem 1: For Model 1 with binary D and binary Z and ξ weakly monotonously increasing in its first argument and the support conditions $E[D | Z = 1] \neq E[D | Z = 0]$ and $0 < \Pi < 1$ a.s. and $0 < P < 1$ a.s., we obtain the following identification results for compliers, treated compliers, non-treated compliers, always-participants and never-participants:

$$E[Y^1 | T = c] = \frac{E\left[YD \frac{Z - \Pi}{\Pi(1 - \Pi)}\right]}{E\left[D \frac{Z - \Pi}{\Pi(1 - \Pi)}\right]}, \quad E[Y^0 | T = c] = \frac{E\left[Y(1 - D) \frac{Z - \Pi}{\Pi(1 - \Pi)}\right]}{E\left[(1 - D) \frac{Z - \Pi}{\Pi(1 - \Pi)}\right]}, \quad (3)$$

$$E[Y^1 | D = 1, T = c] = \frac{E\left[YD \frac{Z - \Pi}{1 - \Pi}\right]}{E\left[D \frac{Z - \Pi}{1 - \Pi}\right]}, \quad E[Y^0 | D = 1, T = c] = \frac{E\left[Y(1 - D) \frac{Z - \Pi}{1 - \Pi}\right]}{E\left[(1 - D) \frac{Z - \Pi}{1 - \Pi}\right]}, \quad (4)$$

$$E[Y^1 | D = 0, T = c] = \frac{E\left[YD \frac{Z - \Pi}{\Pi}\right]}{E\left[D \frac{Z - \Pi}{\Pi}\right]}, \quad E[Y^0 | D = 0, T = c] = \frac{E\left[Y(1 - D) \frac{Z - \Pi}{\Pi}\right]}{E\left[(1 - D) \frac{Z - \Pi}{\Pi}\right]}, \quad (5)$$

$$E[Y^1 | T = n] = \frac{E\left[YD \frac{\Pi - PZ}{P\Pi}\right]}{E\left[(1 - D) \frac{Z}{\Pi}\right]}, \quad E[Y^0 | T = n] = \frac{E\left[Y(1 - D) \frac{Z}{\Pi}\right]}{E\left[(1 - D) \frac{Z}{\Pi}\right]}, \quad (6)$$

$$E[Y^1 | T = a] = \frac{E\left[YD \frac{1 - Z}{1 - \Pi}\right]}{E\left[D \frac{1 - Z}{1 - \Pi}\right]}, \quad E[Y^0 | T = a] = \frac{E\left[Y(1 - D) \frac{(1 - \Pi) - (1 - P)(1 - Z)}{(1 - \Pi)(1 - P)}\right]}{E\left[D \frac{1 - Z}{1 - \Pi}\right]}, \quad (7)$$

$$\text{and } \Pr(T = c) = E\left[D \frac{Z - \Pi}{\Pi(1 - \Pi)}\right], \quad \Pr(T = a) = E\left[D \frac{1 - Z}{1 - \Pi}\right], \text{ and } \quad \Pr(T = n) = E\left[(1 - D) \frac{Z}{\Pi}\right].$$

In Theorem 1 we obtained identification of *mean* outcomes. We could easily extend these results to the identification of quantiles and distributional effects, e.g. by using methods similar to Frölich and Melly (2013) or Frandsen, Frölich and Melly, 2012.

While the identification results for the potential outcomes for the compliers are closely related to the results of Abadie (2003) and Frölich (2007), the results for $E[Y^1 | T = n]$ and $E[Y^0 | T = a]$ are new and based on the combination of selection-on-observables and IV identification strategies.

Estimation of the previous objects is straightforward by replacing the expectation operator by a sample average and plugging in consistent estimates of $\hat{\pi}_i$ and \hat{p}_i for Π and P . These can be obtained by parametric or nonparametric estimation of the functions $\pi(x_1) = E[Z | X_1 = x_1]$ and $p(x_2) = E[D | X_2 = x_2]$. Alternatively, estimation can be based on the (propensity) score matching representations given in Corollaries 1 and 2 in the appendix.

2.2 Relationship to the literature

2.2.1 Marginal treatment effects

Our paper is related to the literature on local and marginal treatment effects. The seminal contribution of Imbens and Angrist (1994) analysed the implications of treatment effect heterogeneity on nonparametric identification of instrumental variable (IV) estimators. Since then the concept that population members differ in their treatment effect and may actively sort on their gains or losses has become widely accepted. An important further extension was the development of the marginal treatment effect (MTE) and its identification by local IV in Heckman and Vytlacil (1999, 2005). Applying the MTE concept, Carneiro and Lee (2009) and Carneiro, Heckman, and Vytlacil (2011) for example found substantial heterogeneity in the returns to college. Carneiro and Lee (2009) analyzed which individuals select into college and how their MTE affects inequality. By examining the estimated MTE patterns, they concluded “individuals sort into the sector where they have both comparative and absolute advantage”. Similarly, Carneiro, Heckman, and Vytlacil (2011) found substantial evidence for large differences in returns and for active sorting, where individuals choose the schooling sector where they have comparative advantage. Widespread heterogeneity in treatment effects was also found in many other studies.

Learning about effect heterogeneity is essential for judging whether alternative allocation patterns would improve overall performance. If the marginal treatment effects can be identified for all values of the covariate space and for all values of the unobservables, the effectiveness of alternative allocation patterns can be compared. Heckman and Vytlacil (2005) suggested the policy relevant treatment effect (PRTE) as a weighted average of marginal treatment effects in order to do exactly that. They also showed that by integrating the MTE over appropriate regions, one could also obtain the ATE, the ATENT, and the ATET. Carneiro, Heckman, and Vytlacil (2010), however,

pointed out that identification and estimation of the PRTE is usually difficult because it requires large support conditions, and because root-n consistent estimation is generally not possible. As an alternative, they proposed the marginal policy relevant treatment effect (MPRTE) as a summary measure to assess the performance of marginal changes in allocation or treatment group compositions. While the MTE, the PRTE, and the MPRTE are important parameters to assess allocation efficiency, they all require at least one *continuously distributed* IV for identification. This requirement of a continuous IV is a limitation of the MTE framework. In many applications, there is only a single discrete instrumental variable available, which is often binary. This is, e.g., usually the case in randomized trials and in (fuzzy) regression discontinuity designs (RDD).

In addition, if one aims to trace the entire MTE trajectory and identify the ATE, the instrument must not only be continuous but also powerful enough to move the treatment probability from zero to one (conditional on other covariates), see Carneiro and Lee (2009) and Carneiro et al (2011).

2.2.2 Extensions of LATE framework

Given these shortcomings, several recent contributions examined extensions of the LATE framework based on one or several binary instruments. In this case, different instrumental variables identify effects for different complier populations. For example, Oreopoulos (2006) exploits various changes in compulsory schooling laws to estimate returns to compulsory education. Within a parametric example, he shows the links between the LATE and the ATE. He concludes that any difference between ATE and LATE becomes smaller the larger the complier fraction affected by the instrument, and that from any two LATE estimates for different instruments, one can calculate the ATE (although the latter finding hinges on the assumed bivariate normal model).

Angrist and Fernandez-Val (2013) extend the idea of combining different LATE estimates within a formal nonparametric framework. Since different instruments imply different complier popula-

tions, they impose some homogeneity assumption to make these effects comparable. Their rather strong conditional constant-treatment effect assumption permits them to reconcile different IV estimates because in their model differences between different IV estimators can only be due to differences in observables. Then, the ATE is identified and overidentification tests are possible.

Hirano, Imbens, Rubin, and Zhou (2000) present an early analysis of treatment effects for always-and never-takers, albeit in a parametric approach. They analyzed a design where encouragement to treatment is randomized, instead of the treatment itself. Using this randomized encouragement design, they estimate distributions of potential outcomes for always-takers, never-takers, and compliers. Their definition of potential outcomes refers to the randomization status Z , i.e. they estimate intention to treat effects (ITT) of the random encouragement design, and identification comes through the functional form assumption. If the usual exclusion restriction on the instrument was valid, the ITT on the always-takers and never-takers should be zero (as their treatment status does not change by randomization). Employing parametric logit specifications within a Bayesian framework, they model the distributions of compliance types and of the potential outcomes for all types. Our paper complements theirs in that we do not impose any parametric restrictions and in that we identify effects of the treatment (and not only of the ITT) for the compliance subgroups.

The previously discussed literature (as well as our paper) imposed assumptions that are sufficiently strong to obtain point identification. An alternative approach examines set identification, which can be obtained under weaker assumptions. An early discussion is given in Angrist, Imbens, and Rubin (1996, in their response to Robins and Greenland) who propose bounds on the treatment effects for always-takers and never-takers (and thus also for the average treatment effect). Huber and Mellace (2010) complement their approach in deriving sharp IV bounds for the average treatment effects on several (sub)populations based on treatment monotonicity combined with dominance assumptions on the mean potential outcomes across subpopulations.

Two recent papers are related to ours in that they examine the combination of instrumental variables and matching assumptions in order to learn about heterogeneity. Donald, Hsu, and Lieli (2014) examine a special case of our setup, where a binary instrumental variable is available but only one-sided non-compliance is permitted. In such a design, only compliers and always-takers exist. They use a selection-on-observables and an IV assumption to estimate the separate treatment effects. The case with only one-sided non-compliance actually implies overidentification, which they exploit to test the selection-on-observables assumption. Since we permit two-sided non-compliance, such a test is not available in our setup.

Angrist and Rokkanen (2013) study a setup related to ours in a RDD framework. The conventional RDD identification only provides the treatment effect at the threshold of the running variable that is valid for the marginal participants. Being left or right of the threshold of the running variable in a fuzzy RDD essentially represents a *binary* instrument, as in our paper, although being valid only in a neighbourhood of the RDD threshold. However, as Angrist and Rokkanen (2013) note, finding a zero treatment effect for these marginal participants does not permit to conclude that the effects for the intramarginal students are zero as well. Finally, to identify effects for other subpopulations as well, Angrist and Rokkanen (2013) augment the instrumental variable identification approach with a conditional independence assumption. Their focus is on effect heterogeneity due to the running variable, though, and not across unobserved compliance types, which is our focus.

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, in Switzerland active labour market programmes (ALMP) were widely introduced during the 1990s. Before the recession of the early 1990s, the unemployment

rate was very low in Switzerland, but increased up to 5% during the recession. This stimulated a comprehensive reform of the unemployment insurance (UI) system. This reform, which became effective partly in January 1996 and partly in January 1997, stipulated the use of ALMP on a wide scale, including training, subsidized employment and on-the-job training in private as well as in public sector jobs. A key element of the reform was the introduction of a *minimum quota* in order to provide a sufficiently large number of programme places. Switzerland consists of 26 administrative regions, called *cantons*, which enjoy a high degree of autonomy. In order to ensure a rapid implementation of the reform of the federal unemployment insurance system as well as the provision of sufficiently many ALMP, the federal government mandated a minimum quota for each canton. Each canton was obliged to fill a minimum number of places in ALMP per year. For the year 1998, the nationwide minimum was set to 25,000 year-places (where a year-place represents 220 programme days). These places were distributed across cantons according to the formula:

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

In this formula, the population share is defined as the fraction of the population living in the canton as of 1996. The unemployment share is defined as the average number of unemployment benefit recipients in the period April 1996 to March 1997 in the respective canton. It is measured relative to the Swiss total. This formula for the distribution of the minimum number of places introduced a regional variation in programme participation, because half of the places were distributed according to the population share, which implied that *relative* to the number of unemployed persons, the quota was rather high in cantons with a low unemployment rate in 1996.

Frölich and Lechner (2010) used this minimum quota to estimate treatment effects for the compliers. In this paper, we complement their empirical analysis by estimating effects also for always-and never-treated. In addition, we estimate long-term effects, whereas Frölich and Lechner (2010)

only contained short-term effects. Here, we find that the short-term effects do *not* fade away and persist in the long run. Frölich and Lechner (2010) discussed that the proclaimed minimum quotas, which were codified in law in November 1996, indeed induced a regional variation in the probability of being treated. 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 lower) to regions of Eastern and Central Switzerland (where the quota was higher). These regions, however, differ also in many other respects, including past unemployment rates and industry structure. Thus, the exclusion restriction might not be plausible.

As an alternative, they used the minimum quota as an instrument only within *local labour markets*, i.e. within confined regions that are divided by a cantonal border. Individuals living in these local labour markets have access to the same job opportunities, irrespective of whether they live left or right of the cantonal border. Yet, if they become unemployed and want to receive unemployment insurance benefits and job offers, they must register at the local employment office where they live. Thus, it will matter then whether they live left or right of the border. The likelihood of participating in an active labour market programme then depends on the strategy of this local employment office. As the local office is governed by the canton, which is obliged to fulfil the quota for the entire canton, the treatment probability in the local employment office depends on the unemployment rate in the entire canton. We thus only compare individuals left and right of the border within local neighbourhoods, where the side of the border should only matter for the probability of being treated. To alleviate remaining concerns about possible differences in the characteristics of the populations living left and right of the border, we will also control for many covariates including the unemployment history, which is a main determinant of the minimum quota as is visible from the formula discussed above. (In fact, Frölich and Lechner (2010) found that control-

ling for these covariates did not make a difference for the IV estimates, such that endogeneity concerns with respect to the quota *within* the local labour markets do not seem to be important.)

3.2 Construction of the local labour markets

The unemployment insurance system is administered through about 150 regional employment offices (REO), which are directly supervised by the cantonal centres. Each REO serves several municipalities within a local area. The unemployed persons cannot change their REO which is assigned according to their residence, unless they move to another municipality. (The only exceptions are the city centres of Zurich and Geneva, which are served by several REO. We will therefore exclude them.) As in Frölich and Lechner (2010), we define an integrated *local labour market* in terms of the area corresponding to a set of regional employment offices (REO). Here, we only sketch the construction of the local labour markets, and refer to the supplementary appendix for more details. We will use only those local labour markets that are partitioned by a cantonal border, as otherwise there would be no variation in the minimum quota. As discussed in Section 2.1, the identification assumption requires that the potential outcomes Y^0 and Y^1 are independent of the instrument, conditional on individual covariates. For Y^1 to be independent of the instrument, the quality and type of treatment should be identical on both sides of the border.

Therefore, the set of REO that define a local labour market have to satisfy the following conditions: (1) They are partitioned by a cantonal border, (2) the commuting time by car between the REOs is at most 30 minutes, (3) in each of these REOs and their areas served the same main language is spoken, and (4) the composition of the offered active labour market programmes is similar in the REOs. By the first criterion, we define local labour markets pair-wise between cantons. This implies for the econometric analysis that the instrumental variable *quota per unemployed* takes only two different values within each labour market. The second and third criterion ensures

that the same jobs are reachable within reasonable commuting time from both sides of the border, and that there is no language divide in that area, which might prohibit some job opportunities if different language skills are expected on the other side of the border.⁶ Finally, the last criterion requires that the labour market programmes offered are similar on both sides of the border.

We identified 18 local labour markets that satisfied the criteria outlined above. Table 1 provides some summary measures for these labour markets. In Column (1) the acronym of the cantonal border that divides the local labour market is given. Columns (2) and (3) report the names of the regional employment offices belonging to this labour market. In columns (4) and (5), we show the cantonal quota per unemployed in January 1998 on either side of the border. (This quota is also used for ordering the 18 labour markets.) Columns (6) and (7) report the number of observations in our dataset, to be described in the next section, left and right of the border. Columns (8) and (9) present the percentages of these unemployed who were treated, where an unemployed person is defined as treated, if he entered a labour market programme (with at least one week duration) during January to March 1998. Column (10) shows the difference between columns (8) and (9), i.e. the cross-border difference in the treatment incidence. We also indicate whether this difference is significantly different from zero. This difference corresponds to an estimate of the fraction of *compliers* (when no covariates are controlled for). This percentage of compliers lies in the range of ±18 percentage points, with many small values. In the subsequent Column (11), we show the product of the fraction of compliers multiplied with the number of observations in that labour market. (I.e. the sum of (6) and (7) multiplied with the difference between (8) and (9), divided by 100.)

⁶ Note that local labour markets where German is spoken on the one side and French on the other side of the border are excluded. On the other hand, French-German bilingual regions bordering to German speaking regions are not excluded. In these latter local labour markets, all individuals 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. On the other hand, all native German-speaking persons are retained, as they can access jobs on both sides of the border.

This represents the estimated *number of compliers* in our dataset in that labour market. In the last columns we report a few descriptive statistics of the unemployed on the two sides of the border, and indicate whether cross-border differences are statistically significant.

--- Table 1 here ---

3.3 Administrative data from the Swiss unemployment and pension system

Our empirical analysis is based on administrative records from the Swiss unemployment insurance and the pension system. The data base contains monthly employment and earnings histories for up to ten years, histories of participation in labour market programmes, as well as many socio-demographic characteristics. Of particular importance to our identification strategy is the case-worker's *rating of employability*, which is a subjective assessment of an unemployed person's employability and captures many 'soft' features of the unemployed. In our empirical analysis, we use the dataset of Frölich and Lechner (2010), which contains 66,713 individuals who were unemployed on the first of January 1998 and eligible to participate in active labour market programmes. Of these, 32,634 individuals lived in the 18 labour markets of Table 1, which is our main sample.

We define participation in ALMP as entering a programme of at least one week duration during January to March 1998.⁷ We will estimate the effects on employment in subsequent years.

As control variables, we define two sets of characteristics \tilde{X}_1 and \tilde{X}_2 that are used in the estimations. We assume that these control variables satisfy the assumptions laid out for X_1 and X_2 in Section 2. To ease notation, in the following we will only be using the symbols X_1 and X_2 , in-

⁷ We also examined alternative treatment windows of two and four months, respectively. Note that these treatment windows are shorter than in Frölich and Lechner (2010), where a treatment window of 12 months was used, because here we also exploit a selection-on-observables assumption, which is less credible for long treatment windows, see e.g. Fredriksson and Johansson, 2008.

stead of \tilde{X}_1 and \tilde{X}_2 , with the implicit assumption that these regressors satisfy the conditions of Section 2. For comparability, we will use the same X_1 variables as in Frölich and Lechner (2010).

For the estimations based on the selection-on-observables (CIA) assumption, we include *additional* control variables that presumably could jointly affect treatment status as well as potential outcomes. Our selection of these variables builds on Gerfin and Lechner (2002) and Gerfin, Lechner, and Steiger (2005), who discussed relevant variables that are jointly related to treatment and potential outcomes. In particular, the detailed long employment, earnings, and programmes histories, and the subjective employability rating by the caseworker cover many factors that are usually not observable. The variables X_2 thus contain X_1 and additionally several variables describing the *employability rating* and the *history of participation in ALMP*. These variables are important individual predictors U_{YD} of the likelihood of participating in ALMP, but not of the quota Z .

Referring back to our discussion of Section 2, we essentially assumed that there are no variables U_{DZ} because all the variables on which the calculation of the minimum quota Z is based are related to employment outcomes such that they must be included in U_{YDZ} or in U_{YZ} . In Section 2 we defined $X_1 = (U_{YDZ}, U_{YZ}, U_{DZ})$ and $X_2 = (U_{YDZ}, U_{YZ}, U_{YD})$. The fact that in our application we could not think of any variables that cause Z and perhaps D but for sure neither Y^0 nor Y^1 , led us to include all variables in U_{YDZ} to be on the safe side. Hence, U_{DZ} is empty. This then implies that X_1 is a strict *subset* of X_2 , where the latter additionally includes the variables U_{YD} .

--- Table 2 here ---

Table 2 lists the outcome variables as well as X_1 and X_2 . The first rows show employment outcomes during 1999 to 2006. Average employment is about 0.55, i.e. between 6 to 7 months employed per year. The subsequent rows show the X_1 variables (all included in X_2), and finally the

eight variables that are included in X_2 but not in X_1 . Statistics are given for the 32,634 individuals who live in the 18 labour markets and for the total sample of 66,713 observations.

3.4 Descriptive evidence relating to the plausibility of the identifying assumptions

Having defined the various local labour markets, we further examine whether there are differences in the individual characteristics X_l within a labour market. If the distribution of X_l seems to be balanced on the two sides of the border, there might not be any need to control for covariates. It may also be an indication of similarity with respect to other variables not observed. On the other hand, if differences in X_l are related to the quota per unemployed, we might want to control for these characteristics in the IV estimator. Furthermore, if there are many such differences, the labour markets may not be as homogenous as expected.

Table 3 provides evidence for the four largest relevant labour markets. This table gives the estimated coefficients of a probit regression of the quota Z (which takes only two different values within each labour market) on the characteristics X_l . I.e. the coefficients correspond to the probit estimation of the instrument propensity score $\pi(x_l) = \Pr(Z = \bar{z} | X_l = x_l)$, which will be used in the next section. To ease the reading of the table, only those coefficients significant at the 5% level are shown. From the table it is obvious that the distribution of X_l is not perfectly balanced within the labour markets. However, most differences do not appear to be systematic. There is only one variable that is significant in all four comparisons ('preferred future job is the same as the last job').

--- Table 3 here ---

In the appendix we show plots of the estimated instrument propensity scores for both sides of the border for all 18 labour markets. These graphs provide a two-dimensional summary measure of cross-border differences in the 59 variables. In many labour markets the distributions of the instrument propensity scores are similar across the border.

4 Empirical results

In this section, we present the estimated effects of participation in ALMP on subsequent labour market outcomes. We follow the individual labour market situation over the years 1999 to 2006. For every year, we define the variable *employment* as the number of months employed in a *non-subsidized job*, divided by 12. (Thus employment in a subsidized job, e.g. temporary wage subsidies, is not counted in this outcome variable.) In the following tables, we show the estimated effects on employment during the years 1999, 2003 and 2006, and, the average effect over the 8 years 1999 to 2006, i.e. the number of months employed in this period, divided by 96 months.

4.1 Implementation of the nonparametric estimators

In our implementation of the estimators, we estimate all the objects of Section 2 based on the propensity score matching representation of Corollary 2. All estimations were done separately for each labour market and each outcome variable. In a first step, we estimate the two propensity scores $\pi(x_1) = \Pr(Z = \bar{z} | X_1 = x_1)$ and $p(x_2) = \Pr(D = 1 | X_2 = x_2)$ by probit. In the second step, we obtain nonparametric estimates of conditional expectations, such as $E[YD | \Pi = \pi, Z = z]$, $E[D | \Pi = \pi, Z = z]$, and $E[Y | P = \rho, D = d]$, by local linear regressions. Separately for each conditional expectation function, the bandwidth value is chosen by leave-one-out least-squares cross-validation. (Note that cross-validation does not deliver the optimal bandwidth choice for our final parameters of interest, which are the mean potential outcomes for compliers, always- and never-takers. For efficient estimation of these latter objects we would need some asymptotic undersmoothing, relative to cross-validation. Hence, our estimators are not optimal. On the other hand, we also examined robustness to bandwidth choice, e.g. doubling and halving the bandwidth values, and obtained similar results.) With these bandwidth values, the conditional expectations are estimated, and sample averages are computed to obtain the estimates of mean potential outcomes,

see Corollary 2 in the appendix. In general, the implementation of the estimator follows Frölich (2004, 2007). All estimates are based on the 32,634 individuals who live in the 18 labour markets.

For the local linear regressions, we use nonparametric *ridge* regressions, see e.g. Frölich (2004), which is a variant of local linear regression where a ridge term is added to the denominator to reduce its variance. Given a sample of observations $(y_i, w_i) \in \mathfrak{R} \times \mathfrak{R}$, where y_i is an outcome variable and w_i a (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

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

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 $\bar{w} = \sum_i w_i K\left(\frac{w_i - \bar{w}}{h}\right) / \sum_i K\left(\frac{w_i - \bar{w}}{h}\right)$. The parameter

r is set to 0.35 for the Gaussian kernel; see Seifert and Gasser (1996, 2000) and Frölich (2004).

The recent simulation studies, Busso, DiNardo, and McCrary (2009) and Huber, Lechner, and Wunsch (2013) investigated a wide range of propensity score estimators within a matching context. In particular Huber, Lechner, and Wunsch (2013), and to some extent also Busso, DiNardo, and McCrary (2009) showed that although estimators similar to those used here may not be the optimal ones in the matching setting, they show a robust behaviour in different specifications and perform generally well. However, of course, it remains an open issue how much of these results carry over to the combined matching IV setting used here.

4.2 Common support

Having estimated the two propensity scores $\pi(x_1)$ and $p(x_2)$, we examine common support in detail in the appendix. There we plot for each of the 18 labour markets the distribution of each propensity score among the treated/non-treated or across the border, respectively. In these 36

graphs we find substantial overlap in most, although not in all, labour markets. Strictly speaking, the expected outcomes are not nonparametrically identified outside the common support, such that our local linear regression estimator will rely on the locally linear regression plane to extrapolate into regions outside the common support. An alternative would be to estimate the region of common support for each labour market and to restrict the estimator to this area. This, however, would make the interpretation of the estimates more difficult since the subpopulations across which we aggregate over the 18 labour markets would then further also differ by their common support definitions. To avoid such complications, as a robustness check we apply a suggestion of Crump, Hotz, Imbens, and Mitnik (2009) to restrict our estimates to the subsample of observations that satisfy $c < \hat{\Pi} < 1 - c$ and $c < \hat{P} < 1 - c$ for some positive value of c . We examined $c = 0.02, 0.05, 0.075, 0.10$, respectively, and found rather robust results.

4.3 Aggregated treatment effects

For the four different outcome variables defined above, we estimate the average 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 12 in the appendix. This large number of estimates makes it difficult to find any discernible patterns. While the estimated potential outcomes for the always- and never-participants vary somewhat around 0.5 and their treatment effects around ± 4 percentage points, the estimates are particularly noisy for the compliers, with some estimates of the potential outcomes even reaching the boundaries of the logical support in some labour markets, because of the small number of observations in most markets. Therefore, 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 by 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.

When interpreting the results one should note, though, that the fraction of people affected differs between the labour markets. In some cantons, many unemployed persons were sent to the programmes already before the reform, while other cantons had been more reluctant. In addition, the gap in the quota between neighbouring cantons varies (see Table 1). In particular, this means that if we hypothetically move an individual from one labour market to another, he could be a complier in the one market but an always-taker (or a never-taker) in the other market. We are thus aggregating across somewhat different types of unemployed. We interpret the aggregated effects for the compliers as self-weighted averages of the treatment effects for all people induced to enter the labour market programmes through the extension of ALMP usage because of the reform (in the subset of labour markets considered). The complier effect can be considered as an approximation to the effect of a policy change where the caseworkers, who have substantial freedom in whom they select, are induced to increase somewhat their treatment quota from the status quo.

Table 4 presents the average effects of ALMP in addition to bootstrap standard errors. We further test whether the 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, using two-sided tests. Significance levels rely on the percentiles of the bootstrap estimates. Analogously, we test whether expected Y^0 outcomes differ between types.⁸

⁸ In general, inference is complicated by the fact that we have multi-step estimators, which involve averages of nonparametric regressions on estimated covariates. For obtaining asymptotic properties of our final estimators one could build on results of Mammen et al (2012) and Hahn and Ridder (2013), who examined nonparametric regressions with generated (i.e. estimated) regressors. This is left for future research. Another concern is the observation that our instrument is weak in many labour markets. With weak instruments, finite sample variances may not exist

The one-year treatment effect on employment for the compliers is 0.17 (see Table 4), which corresponds to a little less than *two months* of additional employment during 1999. This confirms the main finding of Frölich and Lechner (2010), who had only examined short-term treatment effects.

Three new insights are obtained from Table 4: First, we find that the positive effects for the compliers are not short-lived. The effects are positive for employment in 2003 and 2006 (and in fact for all other years as well which are not shown). Furthermore, the average effect over the 8 years from 1999 to 2006 is 0.205, thus positive and similar to the short-term effect. 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 precisely estimated effects are somewhat (and sometimes statistically significantly) greater for the always- than for the never-participants. We also observe that the medium-term effects are somewhat larger than the short-term effects. Hence, after an initial lock-in period, the participation in ALMP turns out to be beneficial for always-participants as well, and perhaps even for the never-participants, but in any case much less so than for the compliers. Third, when comparing the average Y^0 outcomes, we observe that they are considerably larger for the never-participants and always-participants, compared to compliers. Hence, the compliers are a special group of bad risks in the labour market who would find it hard to return to employment without the assistance of active labour market programmes. Perhaps for this reason the programmes are more effective for them.

--- Table 4 here ---

and bootstrap standard errors could thus explode towards infinity. We try to circumvent this problem by basing inference on bootstrap percentiles, which are well defined, even if the (bootstrap) standard errors in finite samples might be infinite. One should point out, though, that this approach only partly solves the problem since the justification of bootstrapping is based on asymptotic arguments, and if the standard asymptotic approximation is poor in case of weak instruments, also the bootstrap may perform very poorly. Again this is left for future research. The nonparametric bootstrap used proceeds by drawing with replacement from the original sample with 66,713 observations and repeating the entire estimation process. 999 bootstrap replications.

What do we conclude from the heterogeneous estimates of the effects and levels for compliers compared to always- and never-takers? The results indicate that the (external) introduction of the quota was indeed effective in terms of reaching *the group with higher than average effects*: Increasing the quota increased the share of worst-off people (i.e. the compliers) in ALMP, who then benefitted from it. However, had the caseworkers on their own been effective in targeting those unemployed who benefit most from the programmes, the effects should have been largest for the always-participants and smallest for the never-participants. This is not the case, though. Hence, some external pressure, here in form of the quota, was helpful to overcome incorrect beliefs of the caseworkers about who benefits most from ALMP. In addition, the caseworkers had also not been successful in selecting the “most deserving”, because, on average, Y^0 is higher for the always-participants than for the compliers. This application thus illustrates how estimates of the effects for the three different groups helps to judge the effectiveness of the implemented allocation scheme (characterised by the effects for always- and never-participants) by some benchmark coming from some external variation (compliers).

5 Conclusions

We proposed a fully nonparametric method to identify potential outcomes not only for compliers but also for always- and never-treated. 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 an instrument can be observed as well.

The suggested methods have been applied to evaluate the effects of active labour market policies in Switzerland. We found positive and long-lasting employment effects for compliers. The effects on the always- and never- participants were much smaller, but still mostly positive. Furthermore, the comparison of the estimated potential outcomes Y^0 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 effects.

References

- Abadie, A. (2003): "Semiparametric Instrumental Variable Estimation of Treatment Response Models," *Journal of Econometrics*, 113, 231-263.
- Angrist, J. and I. Fernandez-Val (2013): "ExtrapoLATE-ing: External Validity and Overidentification in the LATE Framework," in: D. Acemoglu, M. Arellano, and E. Dekel (eds.): *Advances in Economics and Econometrics 3 Volume Paperback Set, Theory and Applications*, Tenth World Congress, Series: Econometric Society Monographs, May 2013, ISBN-13: 9781107628861.
- Angrist, J., Imbens, G., and Rubin, D. (1996): "Identification of Causal Effects Using Instrumental Variables (with Discussion)," *Journal of the American Statistical Association*, 91, 444-472.
- Angrist, J. and M. Rokkanen (2013): "Wanna Get Away? RD Identification Away from the Cutoff," IZA Discussion paper 7429, May 2013.
- Busso, M., DiNardo, J., McCrary, J. (2009): "Finite sample properties of semiparametric estimators of average treatment effects", mimeo, University of California, Berkeley.
- Carneiro, P., J. Heckman and E. Vytlacil (2010): "Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin", *Econometrica*, 78 (1), 377-394.
- Carneiro, P., J. Heckman, and E. Vytlacil (2011): "Estimating Marginal Returns to Education", *American Economic Review*, 101 (October 2011): 2754-2781.
- Carneiro P., and S. Lee (2009): "Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrolment and wage inequality," *Journal of Econometrics*, 149(2), 191-208.
- Crump, R., Hotz, J., Imbens, G., and O. Mitnik (2009): "Dealing with limited overlap in estimation of average treatment effects", *Biometrika*, 96, 187-199.
- Donald, S.G., Y.-C. Hsu, and R.P. Lieli (2014): "Testing the Unconfoundedness Assumption via Inverse Probability Weighted Estimators of (L)ATT," forthcoming in *Journal of Business and Economics Statistics*.
- 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*, 105 (491), 1014-1029.

- Frumento, P., F. Mealli, B. Pacini, and D. Rubin (2012): "Evaluating the Effects of Training on Wages in the Presence of Noncompliance, Nonemployment, and Missing Outcome Data," *Journal of the American Statistical Association*, 107, 450-466.
- Gerfin, M., and M. Lechner (2002): "Microeconometric Evaluation of the Active Labour Market Policy in Switzerland," *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*, 12, 807-835.
- Hahn, J. and G. Ridder (2013): "The Asymptotic Variance of Semi-parametric Estimators with Generated Regressors", *Econometrica*, 81, 315-340.
- 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., Imbens, G., Rubin, D., and Zhou, X. (2000): "Assessing the Effect of an Influenza Vaccine in an Encouragement Design with Covariates," *Biostatistics*, 1, 69-88.
- Huber, M., M. Lechner, and C. Wunsch (2013): "The performance of estimators based on the propensity score," *Journal of Econometrics*, 175, 1-21.
- Huber, M., and Mellace, G. (2010): "Sharp IV Bounds on Average Treatment Effects under Endogeneity and Non-compliance," Discussion Paper no. 2010-31, University of St. Gallen, Dept. of Economics.
- Imbens, G.W., and J.D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-475.
- Kluve J. (2010): "The effectiveness of European active labor market programmes," *Labour Economics*, 17, 904-918.
- Lalive, R., J. C. van Ours, and J. Zweimüller (2008): "The Impact of Active Labour Market Programmes on the Duration of Unemployment," *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 (2010): "Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions," *Empirical Economics*, 39, 111-137.
- Lechner, M., R. Miquel, and C. Wunsch (2011): "Long-run effects of Public Sector Sponsored Training in West Germany," *Journal of the European Economic Association*, 9, 742-784.
- Mammen, E., C. Rothe, and M. Schienle (2012): "Nonparametric Regression with Nonparametrically Generated Covariates," *Annals of Statistics*, 40, 1132-1170.
- Oreopoulos, P. (2006): "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter", *American Economic Review*, 96(1): 152-175.
- 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.
- Vytlacil, E. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result", *Econometrica*, 70(1), 331-341.

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

Cantons	Regional employment offices (REO), different sides of the border		Quota per unemployed Jan 1998 in %		Number of observations		% Treated		% complier ^a	No. of compliers	Fraction women left / right of border		Not been unemployed in past left / right of border		Average unemployment duration left / right of border		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(8)-(9)	(11)	(12)	sig	(13)	sig	(14)	sig	
NWOW-LU	Hergiswil (2x)	Luzern, Emmen, Emmenbrücke, Kriens	25.4 ^b	15.1	265	1607	52.5	49.0	3.5	66	40	33	**	51	47	167	133 ***
UR-SZ	Altdorf	Goldau	22.2	21.3	150	337	39.3	57.9	18.5***	90	48	31	***	52	57	146	132
SZ-SG	Lachen	Rapperswil	21.3	17.3	529	360	53.3	43.9	9.4***	84	49	40	**	58	58	164	167
SZ-ZH	Lachen	Meilen, Thalwil	21.3	12.8	529	1421	53.3	40.2	13.1***	255	43	40		54	58	173	167
SG-TG	Rohrschach, Oberuzwil	Amriswil	17.3	14.6	853	474	42.9	42.4	0.5	7	43	42		57	59	162	165
SG-ZH	Rapperswil	Meilen, Thalwil	17.3	12.8	360	1421	43.9	40.2	3.7	66	43	49	*	54	58	173	164
ZG-LU	Zug	Luzern, Emmen, Emmenbrücke, Kriens	16.6	15.1	571	1607	49.6	49.0	0.6	13	40	44		51	53	167	163
AG-BE	Zofingen	Langenthal	16.3	15.1	472	313	49.2	45.7	3.5	27	39	51	***	59	57	181	174
AG-ZH	Baden, Wettingen, Wohlen	Opfikon, Effretikon, Uster, Wetzikon, Bülach, Dietikon, Regensdorf	16.3	12.8	1529	4165	45.5	39.7	5.8***	330	45	48	**	56	57	173	175
BL-BS	Pratteln, München- stein, Binningen	Basel (3x)	16.0	13.9	934	2081	52.0	34.3	17.8***	537	41	40		48	51	153	163 **
FR-BE	Murten, Tafers, Fribourg	Gümligen, Zolliko- fen, Köniz, Bern (2x)	15.3	15.1	763 ^c	2660 ^c	46.9	45.5	1.4	48	42	41		61	53	***	173 167
FR-VD	Chatel St.Denis	Oron la Ville	15.3	12.3	107	107	51.4	44.9	6.5	14	46	41		43	49	175	174
FR-VD	Romont, Estavayer	Payerne, Moudon	15.3	12.3	371	355	47.4	40.6	6.9*	50	44	38	*	51	48	186	168 **
BE-SO	Wangen, Langen- thal, Burgdorf	Solothurn, Oensin- gen, Biberist, Zuchwil	15.1	11.2	818	877	48.2	51.7	3.5	59	44	40		56	59	167	170
TG-ZH	Frauenfeld	Winterthur	14.6	12.8	537	1221	53.6	39.0	14.7***	258	40	41		54	61	***	170 160
TG-SH	Frauenfeld	Schaffhausen	14.6	11.6	537	605	53.6	45.6	8.0***	91	42	41		46	61	***	154 160
VS-VD	Monthey (2x)	Vevey, Aigle, Montreux	13.0	12.3	609	1580	50.1	40.3	9.8***	215	42	42		46	43	187	171 ***
VD-GE	Nyon	Genf (6x)	12.3	11.5	576	5700	35.8	33.3	2.5	157	48	49		43	52	***	181 186

Note: ^aThis estimate of the fraction of compliers does not control for differences in covariates.

^bThe quota for combined half-cantons NW/OW is computed as an average quota of NW and OW weighted by the number of unemployed in both half-cantons in January 1998

^cIndividuals with French mother tongue are deleted, because a French-German bilingual region is bordering a German-speaking region.

***, **, and * denotes significance at the 1%, 5%, and 10% levels respectively.

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

Variable name	Full sample 66713		Local labour markets sample 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
Control variables X_t				
Age	in years	38	38	38
	older than 50 years (%)	11	10	11
	30 years and younger (%)	23	25	22
Female (%)		45	42	45
Marital status:	married (%)	59	60	58
	single (%)	28	27	27
Number of (dependent) persons in household		2.4	2.4	2.4
	interacted with foreigner status	1.2	1.3	1.3
	interacted with marital status	1.9	1.9	1.9
Foreigner with yearly permit (%)		15	17	16
Swiss national (%)		58	53	57
Mother tongue not German, French or Italian (%)		33	37	35
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
Number of languages known, other than mother tongue (0-3)		1.4	1.4	1.4
First foreign language is	German, French or Italian (%)	64	64	62
	English, Spanish, or Portuguese (%)	14	14	16
Qualification:	skilled (%)	58	54	58
	semi-skilled (%)	14	16	15
Job position:	unqualified labourer (%)	37	38	36
	management (%)	6	5	7
Industry unemployment rate (January 1998, unemployment rate in percent)		6.4	6.6	6.3
Preferred job equals last job (%)		72	74	72
Looking for a part time job (%)		12	14	13

Table 2 to be continued ...

Table 2: ... continued

Variable name	66'713		32'634 individuals	
	ALMP	Non-ALMP	ALMP	Non-ALMP
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
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
in last 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
in 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
in last 5 years (1993-1997) (%)	5	5	5	5
Additional control variables in X_2 but not in X_1				
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) in year 2000 \approx 2/3 Euro. For non-binary variables the means are given. For binary variables (=dummies) the means multiplied by 100 are given.

Table 3: Probit estimates for four largest labour markets – significant coefficients only

Local labour market	VD-GE	BL-BS	AG-ZH	ZH-SZ
Observations	6276	3015	5694	1950
Dependent variable	GE	BS	ZH	SZ
Age in years	0.01 (1.96)			
Marital status: Single		-0.20 (2.74)		
Number of (dependent) persons in household	-0.12 (2.34)	-0.11 (2.46)		
Household size interacted with foreigner status	0.12 (3.12)	0.10 (3.25)		
Foreigner with yearly permit		0.31 (3.53)	0.41 (3.10)	
Swiss national			0.34 (3.76)	0.41 (2.22)
Mother tongue not German, French or Italian		0.17 (2.22)		
Number of foreign languages (0-3)		0.12 (3.28)	0.10 (3.55)	-0.28 (5.08)
First foreign language is German, French or Italian			-0.16 (2.20)	
Qualification: skilled		-0.44 (5.69)	0.34 (6.21)	-0.33 (3.17)
semi-skilled		-0.80 (8.67)	0.52 (7.92)	-0.56 (4.77)
Job position: unqualified labourer	-0.19 (2.17)		0.54 (10.58)	
management	-0.66 (6.22)	-0.26 (2.13)		
Job type: office	-0.21 (3.05)	-0.25 (3.01)		
hotels, restaurant, catering	-0.28 (2.90)		0.17 (2.24)	
construction	-0.42 (2.84)	0.24 (2.30)		
chemistry, metal	-0.44 (3.89)			
painting, technical drawing	-0.25 (2.14)	0.28 (2.71)	0.24 (2.95)	
scientists, teaching, education		0.35 (2.71)		
agriculture, food processing	-0.89 (5.38)			
management, entrepreneurs, senior officials, justice	-0.28 (2.53)			
transportation, traffic	-0.32 (2.51)			
Preferred job equals last job	0.25 (4.34)	0.29 (4.69)	0.23 (5.52)	-0.35 (4.95)
Looking for a part time job	-0.84 (4.14)		0.24 (1.89)	
Unemployment duration in days (as of 1.1.1998)	0.23 (2.38)			
squared (divided by 10000)	-0.07 (2.79)			
Part time unemployed (i.e. not available for a full time job)	0.57 (2.67)			
Insured earnings (CHF)			0.12 (4.43)	
Never been unemployed in last 10 years (1988-1997)			0.28 (2.81)	
in last 5 years (1993-1997)			-0.31 (2.85)	
Fraction of time spent in unemployment		-1.74 (3.79)		
Duration of last employment spell (months)		0.38 (3.71)		
Number of employment spells in last 10 years (1988-1997)	-0.08 (2.86)		0.11 (2.03)	
Fraction of time spent in employment	-0.58 (2.62)			
Number of contribution months to unemployment insurance	-0.92 (2.28)		3.30 (6.50)	
Continuously increasing annual earnings		0.28 (2.91)		

Note: 59 regressors plus a constant. t-statistics are in parentheses.

Table 4: Estimates for compliers, always- and never-participants; treatment window 3 months

	Employment 1999	Employment 2003	Employment 2006	Employment 1999-2006
Always-participants	Estimated fraction of always-participants: 40,7%			
$E[Y^1 T=a]$	0.500	0.579	0.550	0.571
$E[Y^0 T=a]$	0.483	0.524	0.515	0.527
$E[Y^1 - Y^0 T=a]$	0.017 (0.014)	*** 0.054 (0.015)	** 0.035 (0.015)	*** 0.044 (0.013)
Never-participants	Estimated fraction of never-participants: 53,0%			
$E[Y^1 T=n]$	0.482	0.561	0.546	0.555
$E[Y^0 T=n]$	0.497	0.552	0.522	0.547
$E[Y^1 - Y^0 T=n]$	-0.015 (0.014)	0.009 (0.015)	0.024 (0.015)	0.008 (0.013)
Compliers	Estimated fraction of compliers : 6,3%			
$E[Y^1 T=c]$	0.507	0.559	0.513	0.563
$E[Y^0 T=c]$	0.337	0.368	0.375	0.358
$E[Y^1 - Y^0 T=c]$	** 0.170 (0.091)	*** 0.191 (0.094)	** 0.138 (0.097)	*** 0.205 (0.089)
Are treatment effects statistically different?				
$E[\Delta T=c] = E[\Delta T=a]$	*	**		***
$E[\Delta T=c] = E[\Delta T=n]$	**	***		***
$E[\Delta T=a] = E[\Delta T=n]$	*	***		**
Are potential outcomes statistically different ?				
$E[Y^0 T=c] = E[Y^0 T=a]$	***	***	**	***
$E[Y^0 T=c] = E[Y^0 T=n]$	***	***	**	***
$E[Y^0 T=a] = E[Y^0 T=n]$		**		*
$E[Y^1]$	0.491	0.568	0.546	0.562
$E[Y^0]$	0.481	0.529	0.510	0.527
$E[Y^1 - Y^0]$	* 0.010 (0.006)	*** 0.039 (0.006)	*** 0.036 (0.006)	*** 0.035 (0.005)

Note: A participant is entering ALMP between January and March 1998. Employment is defined as number of months employed per year / 12. ***, **, * indicate significance at the 1%, 5%, 10% level, respectively, using bootstrap percentile method, 999 replications. 18 local labour markets; 32,634 observations. (Bootstrap standard errors in parentheses.)