published in Journal of the American Statistical Association, Sep 2010 Volume 105, Nr. 491

EXPLOITING REGIONAL TREATMENT INTENSITY FOR THE EVALUATION OF LABOUR MARKET POLICIES

Markus Frölich[+] and Michael Lechner[++]
[+]*University of Mannheim*
[++]*SEW, University of St. Gallen*

This version: 08 January 2010

Abstract

We estimate the effects of active labour market policies (ALMP) on subsequent employment and earnings by nonparametric instrumental variable estimators. Very informative administrative Swiss data with detailed regional information are combined with exogenous regional variation in programme participation probabilities, which generates an instrument within well-defined local labour markets. We find that ALMP increases individual employment probabilities by about 15% for unemployed that may be called 'marginal' participants.

Keywords: Local average treatment effect, conditional local IV, active labour market policy, state borders, geographic variation, Switzerland, Fuller estimator

JEL classification: J68, C14, C21

Addresses for correspondence

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

1 Introduction*

In the 1990's many European countries used active labour market policies (ALMP) with the aim to reduce Europe's notoriously high levels of unemployment. Switzerland followed the same strategy: When unemployment increased rapidly in the first half of the 1990s, reaching its peak in the winter of 1997, active labour market policies were reformed and considerably extended. There is a rapidly developing literature, surveyed e.g. by Kluve (2006), and Kluve and Schmidt (2002), which attempts to assess whether such policies actually benefited those unemployed who participated in these programmes. (For Switzerland, see e.g. Gerfin and Lechner, 2002.) The almost complete absence of random experiments in Europe brings the issue of identification from observational data to the forefront of the discussion. It is probably fair to say that the vast majority of European studies use either matching estimation or the timing of events approach (Abbring and van den Berg, 2003) to identify the causal effects of such programmes at the individual level. Although the credibility of both assumptions depends on the institutional setting and the data available, they were frequently attacked as not being plausible in general.

* Markus Frölich is also affiliated with IZA, Bonn and IFAU, Uppsala. Michael Lechner has further affiliations with ZEW, Mannheim, CEPR, London, IZA, Bonn, and PSI, London. 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 Grundlagenforschungsfonds 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. We thank the editor and three anonymous referees for their very helpful and detailed comments. We also thank Joshua Angrist, Manuel Arellano, Martin Browning, Kevin Denny, Juan Dolado, Horst Entorf, Bernd Fitzenberger, Lance Lochner, Blaise Melly, Ruth Miquel, Tommaso Nannicini, Jean-Marc Robin, Jeff Smith, Heidi Steiger, Steven Stillman, Coen Teulings and Ernesto Villanueva, and seminar participants in Ann Arbor (UM), Basel, Bern, Canberra, East Lansing (MSU), Essen, Konstanz, London (cemmap), Madrid (CEMFI and Carlos III), Mannheim, San Antonio and Vienna

In this paper, we want to contribute to the knowledge about the effects of ALMP by considering a different route to identification. We use exogenous differences in participation probabilities within local labour markets in Switzerland to construct a plausible instrument that allows us to obtain reliable estimates, at least for the subpopulation of those unemployed responsive to differences in this probability. We base our estimates on a large and very informative administrative database.

The idea of exploiting geographic borders as an instrumental variable to uncover the effects of policy interventions has been used in other studies, although none of them used the fully nonparametric approaches we introduce here. For example, Card and Krueger (1994), Holmes (1998), Black (1999), and Pence (2006) use the U.S. state or district borders. In all these cases, the argument is that policies change abruptly when crossing borders, but that the economic environment changes only little within areas close to it. In other words, crossing the border changes the impact of the policy or the likelihood of being subjected to it, but has no direct effect on individual outcomes that would occur in the absence of the policy differences. The border acts like an instrumental variable. If there are many state borders, each border identifies a separate effect, because – without further homogeneity assumptions – the instrument is valid only locally. The standard approach consists of specifying linear instrumental variable (IV) models and thus implicitly aggregating these different heterogeneous local effects into a single parameter. Since the homogeneity and functional form assumptions implicit in that approach are undesirable and most likely lead to inconsistent estimates, we introduce a *nonparametric* instrumental variable approach that allows for differences in observable characteristics across the local borders, and we propose specific aggregating schemes for the local effects. We thereby go beyond the usual regression discontinuity design in that we exploit multiple border contrasts and estimate the treatment effects for the subpopulation of all compliers.

for helpful comments. Furthermore, we are grateful to the editor, David Banks, and the referees for helping us to improve the paper considerably. The usual disclaimer applies.

In the application, we exploit the fact that Switzerland is a small country with autonomous administrative regions, and runs an extensive active labour market policy. With short commuting times and a good private and public transport infrastructure within regions, local labour markets are integrated across administrative borders and individuals residing on opposite sides close to the border essentially live in the same economic environment. Opportunities for wage arbitrage through relocating instead of commuting hardly exist within such a small region. However, when an employed person becomes unemployed, a specific regional difference concerning active labour market programmes becomes relevant: Although the labour market programmes are largely similar throughout the country, treatment incidence is not. This variation is generated by a regional *minimum quota* requirement, which the Swiss federal government enacted to accelerate the local implementation of a federal labour market reform. As a result, the probability of participating in labour market programmes varies between regions. This exposes unemployed persons within the same labour market to different treatment probabilities. Thus, an instrumental variable strategy becomes available to identify local average effects of participating in the programme, after identifying neighbourhoods on both sides of a regional border that belong to the same local labour market. The border splits the local labour market, and the instrument indicates the part in which the individual lives.

We aim to contribute to the literature in various ways. First, we show how the specific exogenous variation in treatment intensity can be used to build a credible instrument. Second, having constructed the instrument, we propose to correct for remaining differences in observables by using a propensity-score conditional IV estimator recently proposed by Frölich (2007). We show how we can not only identify and estimate local average treatment effects in the specific labour markets, but also additional quantities of interest. Third, as is typical for such studies, each local labour market is small and the nonparametric estimates within each labour market are very noisy. Therefore, we propose various aggregation schemes.

The paper is organised as follows: Section 2 presents the Swiss active labour market policies and the origins of the regional variation in treatment intention intensity. Section 3 describes the econometric approach, and Section 4 discusses the particular implementation for Switzerland. Section 5 gives the estimation results, and Section 6 concludes. Additional background material is available in the discussion paper version (Frölich and Lechner, 2006), which is accompanied by an internet appendix available on *www.sew.unisg.ch/lechner* ***<please insert url of JASA supplemental archives here>***.

2 The Swiss labour market and active labour market policies

2.1 Regional employment offices, unemployment insurance and active labour policies

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 larger scale than before. Although different in some details, the main components of the Swiss ALMP can be found in various programmes in Germany, USA and UK as well. (More details on Swiss active labour market policies can be found in Gerfin and Lechner, 2002, Gerfin, Lechner and Steiger, 2005, and Lalive, van Ours and Zweimüller, 2000.)

With the reform, benefit entitlement was prolonged to two years, but benefit payments were made conditional on willingness to participate in labour market programmes. This *activation principle* empowered the caseworker to assign an unemployed person at any time to any programme if participation is expected to be beneficial to her employability. Non-cooperation by the unemployed person can be (and often is) sanctioned through the suspension of benefits.

Another element of the reform was the consolidation of the 3000 municipal unemployment offices into about 150 regional employment offices (REOs) supervised by 26 cantonal centres. These centres contract private and public organisations to provide programmes and seek to ensure that a sufficient number of programme places can be offered. The REOs 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 REOs.

2.2 The minimum quota to solve an agency problem within the federalist state

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

$$12500 \times (\text{population share}_{1996} + \text{unemployment share}_{1996}) \,,$$

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 unemployment of Switzerland.

The costs of active labour market programmes and of their administration are borne by the federal unemployment insurance funds. The cantons pay a very small lump sum contribution of 3000 Swiss Francs (CHF) per year-place for their assigned minimum quota. 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 funds 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.

The formula for the calculation of the quota for 1998 was codified in November 1996, and in October 1997 the minimum quotas for 1998 were proclaimed. (See first columns of Table 2.1). The formula for the computation of the minimum quota induced regional variation in treatment intention. Relative to the number of unemployed, the quota was rather high in cantons with a low unemployment rate in 1996 because 50% of the quota was distributed according to the population share.

Table 2.1: Minimum quotas and number of unemployed

Canton	Minimum quota[a] 1997	Minimum quota[a] 1998	Number of unemployed Jan 1998	Quota per unemployed[b] Jan 1998	Quota per unemployed[c] Dec 1998	Realised places[d] total 1998	Number of unemployed average 1998	Places per unemployed[e] average 1998
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ZH	4,258	4,325	33,802	**12.8**	18.0	3,976	27,985	14.2
BE	2,947	2,966	19,591	**15.1**	25.4	3,665	14,151	25.9
LU	1,000	1,040	6,885	**15.1**	23.4	1,187	4,967	23.9
UR	64	89	394	**22.6**	35.0	83	244	33.9
SZ	342	370	1,739	**21.3**	29.3	533	1,228	43.4
OW	60	75	273	**27.5**	43.1	56	200	27.8
NW	76	90	381	**23.6**	46.4	81	263	30.9
GL	111	119	636	**18.7**	29.2	110	405	27.3
ZG	283	288	1,737	**16.6**	21.3	305	1,480	20.6
FR	841	805	5,256	**15.3**	22.1	1,319	4,023	32.8
SO	743	773	6,908	**11.2**	21.2	820	4,536	18.1
BS	712	685	4,926	**13.9**	21.2	812	3,855	21.1
BL	774	758	4,740	**16.0**	27.4	805	3,521	22.9
SH	249	242	2,091	**11.6**	18.2	285	1,527	18.6
AR	117	142	633	**22.4**	45.1	118	363	32.5
AI	15	28	112	**25.0**	73.7	8	56	15.0
SG	1,311	1,370	7,899	**17.3**	25.1	1,146	6,079	18.8
GR	369	478	3,172	**15.1**	24.4	433	2,230	19.4
AG	1,629	1,697	10,411	**16.3**	23.8	1,859	8,276	22.5
TG	656	694	4,742	**14.6**	23.6	751	3,455	21.7
TI	1,514	1,445	12,383	**11.7**	16.6	1,828	8,844	20.7
VD	2,833	2,669	21,758	**12.3**	16.5	2,914	17,885	16.3
VS	1,246	1,194	9,197	**13.0**	18.8	1,258	5,710	22.0
NE	715	652	5,449	**12.0**	15.6	1,036	4,513	23.0
GE	1,875	1,750	15,277	**11.5**	15.1	1,219	12,607	9.7
JU	260	256	2,100	**12.2**	23.6	327	1,255	26.1
Total	25,000	25,000	182,492	**13.7**	20.1	26,934	139,658	19.3

Note: Abbreviations of cantons: ZH: Zürich, BE: Bern, LU: Luzern, UR: Uri, SZ: Schwyz, OW: Obwalden, NW: Nidwalden, GL: Glarus, ZG: Zug, FR: Fribourg, SO: Solothurn, BS: Basel-Town, BL: Basel-Country, SH: Schaffhausen, AR: Appenzell-Ausserrhoden, AI : Appenzell-Innerrhoden, SG: St. Gallen, GR: Graubünden, AG: Argau, TG: Thurgau, TI: Ticino, VD: Vaud, VS: Valais (Wallis), NE: Neuchâtel, GE: Geneva, JU: Jura.

a) The minimum quota is the minimum number of 'year-places' to be provided by the canton. A year-place corresponds to 220 programme days.

b) Minimum quota divided by the number of registered unemployed in *January* 1998 (and multiplied by 100).

c) Minimum quota divided by the number of registered unemployed in *December* 1998 (and multiplied by 100).

d) Realised places contain only courses, employment programmes and internships (Berufspraktika). Other smaller programmes are not included.

e) Realised programme places in 1998 divided by the average number of unemployed in 1998 (and multiplied by 100).

Source: Jonathan Gast, seco, Arbeitsmarktstatistik; own calculations.

Consider the situation of the management of the cantonal and the regional employment offices in the beginning of 1998 when planning their strategy to provide active labour market programmes. In the fourth and fifth columns of Table 2.1, the number of registered unemployed in January 1998 and the ratio of the quota to the number of unemployed in January 1998 are given. Suppose the management of the cantonal employment offices forecasted that the number of unemployed would remain constant during the whole year. Then, in cantons such as Uri (UR), Schwyz (SZ), Obwalden (OW), Nidwalden (NW), Appenzell (AR&AI), Glarus (GL), and St. Gallen (SG) with high ratios of the quota to the number of unemployed, the management was under pressure to make sure that many of the relatively few unemployed persons will be assigned to programmes. In the cantons Zurich (ZH), Solothurn (SO), Schaffhausen (SH), Ticino (TI), Vaud (VD), Neuchâtel (NE), Geneva (GE), and Jura (JU), on the other hand, the relative quota was much lower and the administration was under less pressure to fill this quota.

These differences in the quota per unemployed became even more pronounced in the middle or at the end of 1998. Suppose the management did not assume that the number of unemployed would remain constant throughout 1998. Their forecasts may have varied between the cantons. Indeed, the number of unemployed decreased by about 30% during the year and this decrease differed between the cantons: from -22% in Zug (ZG) to -66% in Appenzell (AI). These differential developments even exacerbated the differences in the quota per unemployed, as cantons with relatively few unemployed persons in January (relative to the quota) experienced *larger* decreases in the number of unemployed than cantons where there were relatively many unemployed. Due to these developments, the differences in the quota per unemployed were even more pronounced in the middle or at the end of 1998 (column 6 in Table 2.1). If the cantonal authorities forecasted these trends even roughly, the pressure on those cantons with a high quota per unemployed (with respect to the January figures) was even larger, while it was even less in cantons with a low quota in January. Hence, the *quota per unemployed* in January 1998 indicates the intensity of the pressure to which the local administrations were subjected to assign a sufficient number of unemployed to programmes.

Columns seven to nine of Table 2.1 show that this measure of treatment intention is indeed correlated with subsequent treatment incidence during the year 1998. Column 7 gives the number of year-places that actually were filled in the year 1998, while column eight displays the average number of unemployed in 1998. Column 9 shows the actual extent of treatment per unemployed as the ratio of the two previous columns. The correlation between treatment intention (column 5) and actual treatment incidence (column 9) is 0.53, thus indicating that the quota indeed induced a higher treatment incidence in cantons with a high quota. Hence, the first stage regression on a cantonal level shows a clear relationship.

While the formula for the calculation of the minimum quota generated a regional variation in treatment intensity, the quota per unemployed is unlikely to be a valid instrument per se as it is related to the unemployment rate in 1996. The quota is higher in cantons where unemployment was low in 1996 and vice versa. Nevertheless, it might be a valid instrument *locally* if we compare only individuals living in *local labour markets* that are partitioned by a cantonal border. This identification strategy is described in the next section. Before that, we give some details on the data.

2.3 Administrative data from the Swiss unemployment and pension system

The basis of this study is a large random sample of Swiss unemployed, with individual information from very detailed administrative records. 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 2006. They were obtained from databases of the unemployment insurance system and the social security records. 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. More details on the data and the definition of the local labour markets are given in Section 4.

3 Identification of the individual effects of the active labour market policy

In this section, we discuss the nonparametric identification strategy for identifying and estimating the effects of participation in ALMP on employment. The approach permits selection on unobservables and is based on the concept of *local labour markets* as a key input into a fully *nonparametric* identification strategy, which is derived from the local average treatment effect in Imbens and Angrist (1994) and its extension in Frölich (2007).

Regional variation in treatment intention intensity (i.e. the quota per unemployed) is a candidate instrumental variable for identifying the effects of actual treatment receipt. Since the cantonal minimum quotas are determined by federal law based on the past labour market situation, they are not endogenously chosen according to the preferences of local authorities. The extent to which programmes are provided is subject to different regional perceptions in the cantonal administrations about the desirability of ALMP and the number of unemployed persons. Nevertheless, there is a strong relation between the minimum quota and the share of unemployed persons assigned to programmes (as we saw in Table 2.1). Let the binary variable D indicate whether an individual participates in ALMP and let Y denote the outcome variable, e.g. the employment status or earnings. (Note that we will consider D as a binary variable throughout. Therefore, we will estimate only the effect of the active labour market policy as a mix of programmes, because we have only one instrument.)

Let Z denote the instrument *quota per unemployed*. From a nonparametric perspective, we can consider the regression as separate cross-border comparisons for each labour market. Within each local labour market, we compare those living on the one side of the border with those residing on the other side. The quota takes only two different values and the framework of the local average treatment effect (LATE) of Imbens and Angrist (1994) becomes convenient. We first discuss instrumental variable estimation without any control variables X in the next section, before we turn to including control variables in Section 3.2.

3.1 Identification of LATE with regional variation in treatment intention intensity

Consider every cross-border comparison within a local labour market as a separate estimate of a local average treatment effect (LATE). In a local labour market the instrument takes only two different values $Z \in \{\overline{z}, \underline{z}\}$, where $\overline{z}$ is observed on the one side of the border and $\underline{z}$ on the other side. Suppose $\overline{z}$ is larger than $\underline{z}$. Consider two persons living close to the border but on opposite sides of it. Both persons live in the same economic environment and have the same employment opportunities, but when becoming unemployed they have to register with different regional employment offices. This affects their probability of being assigned to ALMP. The REO pursue different reintegration strategies, which are partly influenced by the minimum quota each canton has to fulfil. REOs in cantons with an ambitious quota will assign more persons to programmes, and will assign earlier, than cantons with a lower quota.

Let D_z denote the potential participation status of an individual i if the level of the instrument were externally set to z. As the instrument takes only two different values, the potential participation variable D_z defines four different types of individuals denoted by $T \in \{a, n, c, d\}$. Following the literature, we call these groups always-participants (a), never-participants (n), compliers (c) and defiers (d). The always-participants would be assigned to ALMP in both cantons. The never-participants would be assigned in neither canton. The compliers are those who are assigned to ALMP in the canton with the higher quota per unemployed $\overline{z}$, but not in the canton with the lower quota $\underline{z}$. For the defiers, this pattern is reversed.

Let Y denote an outcome variable of interest (e.g. earnings, employment status) and denote the potential outcomes by Y_z^d for $d \in \{0,1\}$ and $Z \in \{\overline{z}, \underline{z}\}$. Define $Y_z = Y_z^{D_z}$ as the outcome that would be observed if z were fixed externally. The potential outcomes of interest are $Y^d = Y_Z^d$ where d denotes a value of D fixed externally without a change in Z. Since these potential outcomes might be confounded with the assignment to ALMP, the causal effect of labour market programmes cannot be inferred directly by simple means comparisons.

Under conditions discussed below, the effect for the subpopulation of *compliers* is identified as:

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

This effect is the impact of treatment on those individuals who would switch their treatment status if the value of their instrument would be changed exogenously. They would not participate if living in the canton with the lower quota, but would participate if they were subject to the labour market policy in the neighbouring canton. In other words, this is the marginal group being induced to enter in treatment due to the differing quotas.

Imbens and Angrist (1994) give the conditions for identification of the treatment effect on compliers. Frölich (2007) extends those results to incorporate additional control variables X. The following assumptions, which condition on X, are taken from Frölich (2007). If the set of control variables X is *empty*, we will refer to them as the unconditional IV assumptions (IV.1) to (IV.4).

(CIV.1) **No defiers:** $\mathrm{P}(T = \mathit{defier}) = 0$

(CIV.2) **Compliers:** $\mathrm{P}(T = c) > 0$

(CIV.3) **Unconfounded type:** For all $x \in Supp(X)$

$$\mathrm{P}(T = t \mid X = x, Z = \underline{z}) = \mathrm{P}(T = t \mid X = x, Z = \overline{z}) \qquad \mathit{for}\ t \in \{\text{a,n,c}\}$$

(CIV.4) **Exclusion restriction:** For all $x \in Supp(X)$

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

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

The first assumption requires that there are no defiers, i.e. that an increase in the quota does not induce any unemployed person to switch from participation to non-participation. This holds if an increase in the quota would lead to more unemployed persons being treated, without any substantial organizational changes. Although it cannot be ruled out entirely that an increase in the quota would also have changed the patterns of people being assigned to programmes (for which we do not find any empirical or anecdotal evidence, though), the number of defiers would likely be small. Furthermore, if the treatment effects were identical for compliers and defiers, there would be no bias if assumption (CIV.1) was invalid.

The second assumption requires that the quota has an effect on the treatment probability. This assumption is testable and Table 2.1 already presented some first evidence in this regard. Note that, when including covariates *X*, it is not required that compliers exist for every value of *X*, i.e. $\mathrm{P}(T = c \mid X)$ can be zero for some *X*, because values of *X* where $\mathrm{P}(T = c \mid X) = 0$ receive a weight of zero in the averaging of the effects in the complier subpopulation.

The third assumption requires the fractions of compliers, always- and never-participants to be the same on both sides of the border. In particular, this rules out a selective choice of residence by the compliers. Unemployed persons might have realized that the probability of being assigned to ALMP is different in the neighbouring canton. As some of them had preferences to take part in programmes, or conversely, to avoid participation, they might have preferred attending a REO in the other canton. This, however, would have required moving to the neighbouring canton (before being assigned to a programme). While the costs of changing residence are quite substantial, its benefits are small and highly uncertain: the differences in the probability of being assigned to treatment are not very large between neighbouring cantons. Hence, such a selective choice of residence appears rather unlikely. In any case, assumption CIV.3 becomes more plausible if we compare only individuals who are identical on a large number of characteristics X. We need to control for all variables X that affected the instrument as well as the type of the individual. We are going to control for a large number of characteristics in the CIV estimator, which, nevertheless, turns out to give similar estimates of the fractions of compliers, thereby indicating that confounding with respect to type might indeed not have been a serious problem.

The fourth assumption represents an exclusion restriction on the population level. It requires that those living to the one side of the border have the same expected potential outcomes as those living on the other side of the border. One could think of this as a combination of an *exclusion restriction* on the individual level (i.e. no direct effect of Z on the outcome) and an *unconfoundedness-of-the-instrument* condition on the population level (i.e. the populations on either side of the border do not differ systematically).

For the potential outcome Y^0, this assumption requires that the quota does not directly affect the employment prospects of an unemployed person. In other words, the employment chances should be the same on both sides of the border. To take account of this requirement, we will consider only local labour markets with very good commuting infrastructure and short commuting times (at most 30 minutes by car). We will also restrict our analysis to unemployed persons without (known or probable) restrictions to their mobility, and control for individual characteristics X.

For the potential outcome Y^1, assumption (CIV.4) not only requires the labour markets to be integrated, but also that the type and quality of ALMP is the same. In other words, that the quota did not affect the composition of ALMP. It appears reasonable that the courses and programmes are of comparable quality because the cantons frequently purchase them from private providers that operate in the entire country. Neighbouring cantons may even buy places in the same courses. We also tested whether the quota is systematically related to the composition of the programmes, in terms of training, employment programmes and temporary wage subsidies, and did not find any systematic relationship. (See the appendix (Table IC.3) to Frölich and Lechner, 2006) Nevertheless, to be on the safe side, we will include in our analyses only those labour markets with a similar ALMP structure on both sides of the border.

Besides the absence of a direct link between the quota and the outcomes, we also need the instrument to be *unconfounded*. In other words, the populations residing on the two sides of the border should be identical in terms of their employability. The clear definition of the instrument as defined according to a pre-determined formula in the unemployment insurance law certainly adds credence to the plausibility of this assumption, because the instrument is *not* an outcome of a concurrent unobserved bargaining process or the result of (biased) forecasts of labour market developments. Instead, it was publicized long before it affected the actual implementation of the programmes.

Nevertheless, other *common factors* could introduce a confounding between the instrument and the outcome variables. The average employability of the population in 1996 might be such a common factor, which affected the quota per capita and, most likely, also the outcomes in 1999. Nevertheless, once we examine only *local* labour markets along a border, this correlation is unlikely to be large. It could still be, though, that other differences in public policies, such as local tax rates, social assistance, policies supporting families, child allowances, which would attract different types of individuals, correlate by chance with the instrument. In fact, we do find that the populations differ somewhat across the border. To be on the safe side, we thus control for a large number of individual characteristics X in our estimations, including entire earnings and employment histories, family composition, skills and job information, industry unemployment rates etc. By controlling for these characteristics, we ensure that the instrument is not picking up differences in other characteristics.

This discussion of the assumptions CIV.1 to CIV.4 already indicates which variables *X* we want to control for, which are all variables that are jointly related with the instrument and the potential outcomes. These control variables are permitted to be endogenous in some sense (see e.g. Frölich, 2008). Yet, we do not want to control for variables that are already causally affected by the instrument, though. In Section 5, we control for a large number of covariates in the CIV estimator (defined below). Yet, we will see that the results with and without *X* do not differ very much. Hence, in retrospect, we may say that confoundedness of the instrument was not a real concern.

3.2 Conditional instrumental variables (CIV)

The previous assumptions required the instrument to be unconfounded. As discussed in Section 2, this assumption is unlikely to hold if we use the entire Swiss unemployed population. Thus, we restricted the analysis to geographically limited local labour markets. In addition, as discussed above, we might additionally also like to control for various covariates *X*. Without covariates *X*, the complier effect was identified by (1). When including *X*, Frölich (2007) shows that the treatment effect for the 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 by a ratio of two generalized ‘matching’ estimators:

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

where $\hat{m}_z$ and $\hat{\mu}_z$ are 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 an estimate of the fraction of compliers $\mathrm{P}(T = c)$.

(Please note that we use the word matching estimator in the more general sense introduced by Heckman et al. (1998) and Imbens (2004), where each observation of the $Z = \bar{z}$ group is matched to multiple observations of the $Z = \underline{z}$ group, and vice versa, using weights that may not be integers. In the classical sense in the statistics literature each observation is matched to one or several of the other group, corresponding to taking weighted averages where each weight is an integer. Below we use a type of kernel weights where each observation with $Z = \bar{z}$ is matched to all observations from the $Z = \underline{z}$ group but with (non-integer) weights that decline by distance.)

The population of compliers consists of two subpopulations: those who actually receive treatment and those who do not. When the instrumental variables assumptions are valid without conditioning on *X*, the effects for these two subpopulations are identical, otherwise they are not. With a similar reasoning as before, the treatment effect on the treated compliers is identified as well:

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

where $\pi(x) = \mathrm{P}(Z = \bar{z} \mid X = x)$, see proof in the discussion paper (Frölich and Lechner, 2006).

Similarly, we can identify the potential outcomes of the complier population separately. (These relationships can be derived by obvious changes in the proofs in Frölich (2007). We omit the proofs.)

$$E[Y^1 \mid T = c] = \frac{\int E[YD \mid X, Z = \bar{z}] - E[YD \mid X, Z = \underline{z}] \;\; dF_X}{\int E[D \mid X, Z = \bar{z}] - E[D \mid X, Z = \underline{z}] \;\; dF_X}, \qquad (5)$$

$$E[Y^0 \mid T=c] = -\frac{\int \left(E[Y(1-D) \mid X, Z=\bar{z}] - E[Y(1-D) \mid X, Z=\underline{z}] \right) dF_X}{\int \left(E[D \mid X, Z=\bar{z}] - E[D \mid X, Z=\underline{z}] \right) dF_X}. \tag{6}$$

We obtain similar expressions for the treated compliers:

$$E[Y^1 \mid D=1, T=c] = \frac{\int \left(E[YD \mid X, Z=\bar{z}] - E[YD \mid X, Z=\underline{z}] \right) \cdot \pi(X) \cdot dF_X}{\int \left(E[D \mid X, Z=\bar{z}] - E[D \mid X, Z=\underline{z}] \right) \cdot \pi(X) \cdot dF_X}, \tag{7}$$

$$E[Y^0 \mid D=1, T=c] = -\frac{\int \left(E[Y(1-D) \mid X, Z=\bar{z}] - E[Y(1-D) \mid X, Z=\underline{z}] \right) \cdot \pi(X) \cdot dF_X}{\int \left(E[D \mid X, Z=\bar{z}] - E[D \mid X, Z=\underline{z}] \right) \cdot \pi(X) \cdot dF_X}. \tag{8}$$

The corresponding estimators are analogous to equation (3). Estimating these expected potential outcomes separately permits us to impose restrictions on the range of the outcome variables in a straightforward way by capping them at the logical boundaries of their supports.

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

$$\hat{E}[Y^1 - Y^0 \mid T=c] = \frac{\sum_{i:z_i=\bar{z}} \left(y_i - \hat{\mathrm{m}}_{\underline{z}}(\hat{\pi}_i) \right) - \sum_{i:z_i=\underline{z}} \left(y_i - \hat{\mathrm{m}}_{\bar{z}}(\hat{\pi}_i) \right)}{\sum_{i:z_i=\bar{z}} \left(d_i - \hat{\mu}_{\underline{z}}(\hat{\pi}_i) \right) - \sum_{i:z_i=\underline{z}} \left(d_i - \hat{\mu}_{\bar{z}}(\hat{\pi}_i) \right)}, \tag{9}$$

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. (See Frölich and Lechner, 2006.)

Besides the potential outcomes for the compliers, we also identify the fractions of compliers, always-participants and never-participants in the respective local labour markets, as well as the expected treatment outcome for the always-participants and the expected non-treatment outcome for the never-participants. (The proofs are analogous to those for the previous results and are omitted.)

$$\mathrm{P}(T=c)=\int E[D \mid X, Z=\bar{z}]-E[D \mid X, Z=\underline{z}]\; dF_X\ , \tag{10}$$

$$\mathrm{P}(T=a)=\int E[D \mid X, Z=\underline{z}] dF_X\ , \tag{11}$$

$$\mathrm{P}(T=n)=\int E[1-D \mid X, Z=\bar{z}] dF_X\ , \tag{12}$$

$$E[Y^1 \mid T=a]=\frac{\int E[YD \mid X, Z=\underline{z}] dF_X}{\int E[D \mid X, Z=\underline{z}] dF_X}\ , \tag{13}$$

$$E[Y^0 \mid T=n]=\frac{\int E[Y(1-D) \mid X, Z=\bar{z}] dF_X}{\int E[1-D \mid X, Z=\bar{z}] dF_X}\ . \tag{14}$$

Hence, the CIV (and IV) assumptions permit identification of $E[Y^1 \mid T=a]$ and $E[Y^0 \mid T=n]$, but do not identify treatment *effects* for these two groups.

4 Implementation of the evaluation of Swiss active labour market policies

4.1 Data and sample selection

The population for the microeconometric evaluation consists of all individuals who were unemployed on January 1, 1998, for at most one year. For a random sample of 81399 individuals, detailed information on employment histories (including self-employment), monthly earnings, participation in ALMP, and personal characteristics for the years 1988 to 1999 were obtained from administrative databases of the unemployment insurance system and the social security records.

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, as discussed with our IV identification strategy in Section 3. 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 were 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 them to take advantage of job opportunities that are not nearby. We restrict the sample to the prime age group (25-55). Furthermore, we excluded students, apprentices and home workers, and persons registered as part-time employed. (For details see Frölich and Lechner, 2006). The final sample contains 66,713 individuals. Of these, 32,634 individuals live in one of the local labour markets that we define in Section 4.3.

4.2 Definition of outcomes, treatment and conditioning variables

We measure the individual labour market situation during the year 1999 and create the following outcome variables: *Employment* is defined as the number of months with positive earnings in a non-subsidized job in 1999, divided by 12. (Employment in a subsidized job, e.g. temporary wage subsidies, is not counted as regular employment.) *Earnings* are defined as earnings from employment or self-employment during a year, divided by 12. These outcome variables capture the policy objectives of the ALMP, namely rapid and lasting re-employment without large earnings losses. Participation is defined as entering in a programme of at least one week in duration at any time in 1998.

Table 4.1 gives descriptive statistics for the outcome and the treatment variables and for the 59 control variables X used in the CIV estimation. The means are shown for the total sample of 66,713 individuals, as well as for the 32,634 individuals in the relevant labour markets.

Overall, 60% of all unemployed entered active labour market programmes. 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 in ALMP in December (for the first time in 1998). The average employment outcome is about 0.60 to 0.63, which corresponds to about 7.5 months of employment in the subsequent 12 months. Average monthly earnings are about 2100 to 2400 CHF. (The average monthly earnings for those with positive earnings are thus about 3800 CHF.)

Table 4.1: Descriptive statistics of selected characteristics (means or shares multiplied by 100)

Variable name	66,713		32,634 individuals	
	ALMP	Non-ALMP	ALMP	Non-ALMP
Observations	40,193	26,520	19,522	13,112
Outcome variables in 1999				
Employment: Number of months employed in 1999, divided by 12	0.59	0.63	0.60	0.63
Earnings: Total earnings from employment and self-employment, divided by 12	2126	2351	2222	2419
Control variables *X*				
Age in years	38	37	38	37
older than 50 years (%)	11	10	12	10
30 years and younger (%)	23	27	22	26
Female (%)	45	41	45	43
Marital status: married (%)	59	59	59	58
single (%)	27	29	27	28
Number of (dependent) persons in household	2.5	2.4	2.5	2.4
interacted with foreigner status	1.3	1.3	1.3	1.3
interacted with marital status	1.9	1.9	1.9	1.9
Foreigner with yearly permit (%)	16	16	16	16
Swiss national (%)	56	55	56	56
Mother tongue not German, French or Italian (%)	35	36	37	35
Immigrant who migrated to Switzerland in 1988-1992 (and ≥ 25 years old then) (%)	5	5	5	5
in 1993-1997 (and ≥ 25 years old then) (%)	6	4	6	5
Number of languages known, other than mother tongue (0-3)	1.4	1.4	1.4	1.5
First foreign language is German, French or Italian (%)	64	64	63	61
English, Spanish or Portuguese (%)	14	15	16	19
Qualification: skilled (%)	56	56	56	58
semi-skilled (%)	15	15	16	17
Job position: unqualified labourer (%)	38	37	37	35
management (%)	6	6	6	7
Industry unemployment rate (January 1998, unemployment rate in percent)	6.4	6.6	6.3	6.3
Job type: office (%)	15	14	15	16
hotels, restaurant, catering (%)	16	16	15	13
construction (%)	7	9	7	8
chemistry, metal (%)	8	8	8	8
painting, technical drawing (%)	7	7	7	7
scientists, teaching, education (%)	5	4	5	5
agriculture, food processing (%)	3	3	2	3
health care (%)	3	3	3	3
management, entrepreneurs, senior officials, justice (%)	3	4	3	4
transportation, traffic (%)	3	4	3	4
Preferred job equals last job (%)	72	75	71	74
Looking for a part time job (%)	13	14	13	16

Table 4.1 to be continued

Table 4.1.: *... continued*

Variable name	66,713		32,634 individuals	
	ALMP	Non-ALMP	ALMP	Non-ALMP
Unemployment duration in days (as of 1.1.1998)	176	155	178	161
squared (divided by 10000)	4.2	3.6	4.3	3.8
Part time unemployed (i.e. not available for a full time job) (%)	11	14	11	15
Insured earnings (CHF)	3980	3832	4091	3941
Earnings < 2000 CHF	8	11	7	11
> 6000 CHF	10	9	12	11
Never been unemployed in last 10 years (1988-1997) (%)	48	42	50	45
5 years (1993-1997) (%)	53	47	54	49
Number of unemployment spells in the period 1988-1992	0.30	0.32	0.28	0.28
last 5 years (1993-1997)	0.94	1.12	0.88	1.01
Fraction of time spent in unemployment (since first registration in pension data)	0.13	0.12	0.13	0.12
interacted with immigrant status	0.03	0.02	0.03	0.02
Duration of last employment spell (months)	44	40	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)	2.52	2.73	2.44	2.58
Fraction of time spent in employment (since first registration in pension data)	0.78	0.78	0.79	0.78
interacted with immigrant status	0.071	0.066	0.072	0.064
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)	26340	25601	26657	25515
1996 (CHF)	39823	37285	41324	38503
1995 (CHF)	38853	37712	40620	39379
Ever been self-employed in the period 1988-1992 (%)	8	8	7	8
last 5 years (1993-1997) (%)	5	5	5	6

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.

The differences between participants and nonparticipants are not dramatically large, although visible in particular in the short- and long-term labour market histories. An important difference is the labour market history in the year prior to participation. Participants spent more time in unemployment and received more ALMP already in 1997. Comparing the variables for the 66713 and the 32634 samples does not reveal any important differences, so that we expect that the results we obtain for the selected regional markets carry over to the remaining parts of the country.

4.3 Identifying local labour markets

To apply the evaluation strategy discussed in the previous sections, integrated *local labour markets* with internal administrative borders need to be found. We define a local labour market in terms of the area corresponding to one or more regional employment offices. In particular, we seek to identify a *cluster of REOs* that satisfies the following criteria: 1) The REOs are 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 REOs. With the first criterion, we identify local labour markets pair-wise between cantons. If a local labour market extends into three or more cantons, we consider each pair-wise comparison between the cantons separately.

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 implemented by examining the distances between any pair of regional employment offices in terms of commuting times by car. A maximum driving time of about 30 minutes seems acceptable for exploiting wage arbitrage opportunities. 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.

The third criterion takes account of the different language regions, as Switzerland consists of German, French, and Italian speaking parts. 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. Based on the criteria one to three, 30 local labour markets are identified.

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 3, one of the IV assumptions is that the quality and type of treatment is identical on both sides. It appears reasonable to assume that the quality of the services does not vary systematically between neighbouring regions, but there is variation in the types of programmes the caseworkers assign their clients to. We therefore limit our analysis to only those 18 local labour markets with a similar ALMP-structure.

Table 4.2 displays summary statistics for these 18 labour markets. Column (1) gives the number of the labour market, for future reference. Column (2) indicates the cantonal border that partitions the labour market, and columns (3) and (4) give the REOs belonging to this labour market (on the two sides of the border). Columns (5) and (6) give the number of observations in the sample.

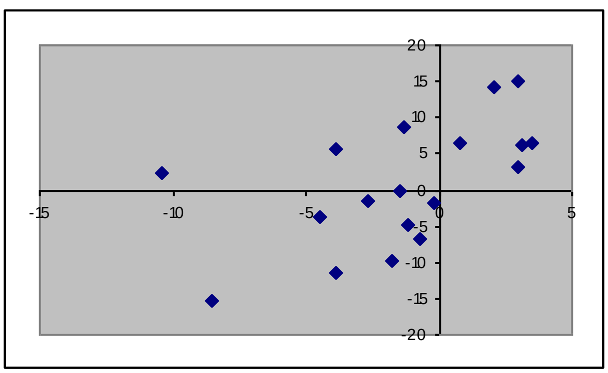
Table 4.2: Local labour markets divided by administrative border

#	Can-tons	Regional employment offices (REOs)		Number of ob-servations N_1	N_2	% Treated		Complier[b]	Diff. in instrument Table 2.1[c]
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
2	SO-BE	Solothurn, Oensingen, Biberist, Zuchwil	Wangen, Langenthal, Burgdorf	877	818	68	63	5.6	-3.9
7	BE-AG	Langenthal	Zofingen	313	472	64	68	-4.9	-1.2
8	BE-FR	Gümligen, Zollikofen, Köniz, Bern (2x)	Murten, Tafers, Fribourg	2,660[a]	763[a]	65	67	-1.7	-0.2
9	FR-VD	ChatelSt.Denis	Oron la Ville	107	107	74	59	15.0	3.0
10	FR-VD	Romont, Estavayer	Payerne, Moudon	371	355	64	60	3.3	3.0
11	VD-GE	Nyon	Genf (6x)	576	5,700	57	50	6.5	0.8
12	VD-VS	Vevey, Aigle, Montreux	Monthey (2x)	1,580	609	59	66	-6.7	-0.7
13	BL-BS	Pratteln, München-stein, Binningen	Basel (3x)	934	2,081	67	53	14.2	2.1
15	LU-NWOW	Luzern, Emmen, Em-menbrücke, Kriens	Hergiswil (2x)	1,607	265	64	62	2.4	-10.4
16	LU-ZG	Luzern, Emmen, Em-menbrücke, Kriens	Zug	1,607	571	64	64	-0.2	-1.5
17	SZ-UR	Goldau	Altdorf	337	150	69	61	8.8	-1.3
19	AG-ZH	Baden, Wettingen, Wohlen	Opfikon, Effretikon, Uster, Wetzikon, Bülach, Dietikon, Regensdorf	1,529	4,165	64	58	6.6	3.5
21	ZH-TG	Winterthur	Frauenfeld	1,221	537	59	69	-9.9	-1.8
22	ZH-SG	Meilen, Thalwil	Rapperswil	1,421	360	56	60	-3.8	-4.5
23	ZH-SZ	Meilen, Thalwil	Lachen	1,421	529	56	72	-15.2	-8.5
24	TG-SH	Frauenfeld	Schaffhausen	537	605	69	63	6.3	3.1
25	TG-SG	Amriswil	Rohrschach, Oberuzwil	474	853	64	66	-1.5	-2.7
28	SG-SZ	Rapperswil	Lachen	360	529	60	72	-11.4	-3.9

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.
[c] Difference in the instrument quota per unemployed (Table 2.1, column 5) between the two cantons.

Columns (7) and (8) display how many of these observations were treated, and column (9) gives the difference in the treatment probability. This is an estimate of the fraction of compliers (when no covariates X are controlled for) and lies in the range of ±15 percentage points, with many small values. Column (10) gives the difference in the quota per unemployed between the two cantons (calculated from column (5) of Table 2.1). We compare these last two columns of Table 4.2 to see whether the positive relationship observed in Table 2.1 between the *quota per unemployed* and treatment incidence can be confirmed for our population. The scatter plot in Figure 4.1 shows that this relationship is indeed positive, with a correlation of 0.57. Note that 15 of the 18 local labour markets are located in the upper-right or lower-left quadrant, i.e. the change in the quota and the change in treatment probability have the same direction.

Figure 4.1: Correlation between differences in the instrument and the estimated complier fraction



Note: Abscissa: Differences in the quota per unemployed (column 10 of Table 4.2).
Ordinate: differences in treatment probability (complier fraction, column 9 of Table 4.2).

5 The effects of active labour market policies in Switzerland

In this section, we exploit the cantonal quota as an instrumental variable to overcome selection on unobservables and identify local average treatment effects. All estimates are based on those 32634 individuals who live in the 18 labour markets defined in the previous section. We estimate the short-term effects for the two outcome variables, employment in 1999 (in months, divided by 12) and gross earnings in 1999 (divided by 12, in CHF).

5.1 Implementation of the nonparametric estimators

5.1.1 Unconditional IV estimates

If the IV conditions are valid without conditioning on *X*, the effect of participation in ALMP for the compliers living in the local labour market along the administrative boundary is identified by (1) and can be estimated by dividing the cross-border difference in *Y* by the cross-border difference in *D*. When the instrument has only a weak impact on *D* in that the participation probabilities do not differ much from the one side of the border to the other side, this Wald estimator (Wald, 1940) can have poor finite sample properties, because the difference in *D* appears in the denominator. This was confirmed in a Monte Carlo study where the Wald estimator often seemed to have an infinite variance but the Fuller estimator performed much better (see the previous discussion paper version of this paper).

5.2.1 Estimates conditional on covariates

In addition to the previous estimator without *X*, we also examine IV estimation with controlling for *X*. The CIV estimation proceeds in three steps, separately for each local labour market and each outcome. The implementation of the instrumental variable estimator follows Frölich (2004, 2007): (i) The probability $\pi(x) = \mathrm{P}(Z = \overline{z} \mid X = x)$ is estimated by a binary probit to obtain predicted probabilities $\hat{\pi}_i$ for all observations. (ii) Bandwidth values are selected by leave-one-out least squares cross-validation for the nonparametric regression, separately for the estimation of $E[D \mid \hat{\pi}(X) = \rho, Z = \underline{z}]$, $E[D \mid \hat{\pi}(X) = \rho, Z = \overline{z}]$, $E[YD \mid \hat{\pi}(X) = \rho, Z = \underline{z}]$, $E[YD \mid \hat{\pi}(X) = \rho, Z = \overline{z}]$, $E[Y(1-D) \mid \hat{\pi}(X) = \rho, Z = \underline{z}]$ and $E[Y(1-D) \mid \hat{\pi}(X) = \rho, Z = \overline{z}]$. Bandwidths are chosen from the expanding grid with 10 values: {1/100, 1.9/100, 1.9^2/100, ..., 1.9^8/100, ∞}. (iii) With the selected bandwidths, m_z and μ_z are estimated by 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. (We use ridge regression because local linear regression oftens perform poorly in small samples and/or with small bandwidths, see e.g. Seifert and Gasser, 1996, or Frölich, 2004, for matching estimation.) Given a sample of N observations on $(\pi_i, y_i) \in \Re \times \Re$, where $\pi_i = \hat{\pi}(X_i)$, and a bandwidth value h, the ridge regression estimate at a location π_0 is

$$\hat{E}[Y \mid \hat{\pi}(X) = \pi_0] = \frac{T_{1,0}}{T_{0,0}} + \frac{T_{1,1} \cdot (\pi_0 - \bar{\pi})}{T_{0,2} + rh \mid \pi_0 - \bar{\pi} \mid}, \tag{15}$$

where $T_{a,b} = \sum_{i=1}^{N} y_i^a \cdot (\pi_i - \bar{\pi})^b K\left(\frac{\pi_i - \bar{\pi}}{h}\right)$ and $\bar{\pi} = \sum_{i=1}^{N} \pi_i K\left(\frac{\pi_i - \pi_0}{h}\right) / \sum_{i=1}^{n} K\left(\frac{\pi_i - \pi_0}{h}\right)$. The parameter r is set to 0.35 for the Gaussian kernel (Seifert and Gasser, 1996, 2000, and Frölich, 2004).

This procedure estimates $E[Y^1 \mid T = c]$ and $E[Y^0 \mid T = c]$ for every local labour market. 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 in [0, 1] for the employment variable.

5.2 Results of the instrumental variable estimators

Table 5.1 presents the estimated effects of ALMP on employment and earnings for the 18 labour markets and their weighted averages. Columns (1) to (4) show the results for the CIV estimator with covariates X, for the compliers (columns 1 and 2) as well as for the treated compliers (3, 4). Columns (6) to (9) show the results for the compliers without controlling for X, for the Fuller estimator (6, 7) and for 2SLS (8, 9). All estimators impose range restrictions to ensure that estimated potential outcomes are non-negative and, in the case of employment, not larger than one.

The estimates for the 18 labour markets vary considerably, with several results not being plausible, such as employment effects of 1 or –1 and large negative earnings effects. Still, the estimated employment effects show similar patterns for the different estimators. The correlation for the 18 labour markets between the CIV estimates and the Fuller estimator with range restrictions is 0.68 (the correlation between Fuller and 2SLS is 0.94.) Furthermore, the estimated fractions of compliers are very similar for the CIV and the Fuller estimators (the correlation is 0.98).

Although the estimates for the labour markets are highly variable, appropriately weighted averages $\hat{\Theta}$ might reveal the mean effects with more precision. The goal of the weighting schemes we propose is to approximate in different ways the average complier effects. The different approximations implicitly attach different importance to the different components of the implicit trade-off between improving efficiency of the final estimate on the one hand and retaining the easy and clear interpretability of the aggregate estimate on the other hand. Let $\hat{\Theta} = w'\hat{\theta}$, where $\hat{\theta}$ is the vector of the 18 local estimates and w is a vector of weights. The first row of Table 5.1 gives the results for the *compliers-weighted* LATE Θ_c. These weights w_c represent the estimated number of compliers (i.e. complier fraction multiplied with the number of observations) and are given in columns (5) and (10). A comparison between the columns reveals that they are similar for the specifications with and without covariates. The subsequent rows present estimates for other weighting schemes to assess the sensitivity of the results to the particular weights used. $\Theta_{c,trim}$ is based on the number of compliers using only those labour markets where the estimates were not censored to be within logical range. Θ_{obs} and $\Theta_{obs,trim}$ weight the estimates by the number of observations, but the latter gives positive weight only to labour markets where the estimates did not need to be censored at their logical range. For Θ_{MD} the minimum distance weights are applied, i.e. the estimates of the individual labour markets are weighted by the Cholesky decomposition of the inverse of their covariance matrix. Minimum distance weights generally have advantages on efficiency grounds. However, the Θ_{MD} weights are difficult to interpret in our application, also because censored estimates may have a very small or even zero variance and would thus get a large weight. However, these labour markets should not play much of a role for the interpretation of the results. Furthermore, Θ_{MD} is not attractive when finite sample moments of the estimators do not exist.

In Table 5.1, inference is based on the nonparametric bootstrap. The bootstrap proceeded by drawing with replacement from the original sample with the 66713 observations and repeating the entire estimation process. For the estimators *without* covariates *X*, i.e. the Fuller estimator and 2SLS, we use a nested bootstrap with 999 times 999 replications in order to bootstrap the t-statistic, which is likely to lead to more reliable inference arising from asymptotic refinements, see Hall (1986), Beran (1987), Loh (1987) and Cameron and Trivedi (2005, p. 374). More precisely, in the outer bootstrap loop we estimate the aggregated effect and use the inner bootstrap loop to estimate the standard error of this effect in order to calculate the t-statistic. (This approach may not be valid for the earnings estimates of the 2SLS estimator since the finite sample standard errors may not exist, depending on the true distribution of the error terms. However, inference on the percentiles of the estimates, which always exist, lead to similar inference results: The earnings estimates are always very noisy and never statistically significantly different from zero.)

For the CIV estimator, which incorporates the covariates *X*, a nested bootstrap approach is not feasible due to an excessive demand on computation time. We therefore use 999 replications to bootstrap the aggregated estimate. In the following table we also report the bootstrap standard error of the estimate, but we base our inference on the percentiles of the estimates because the finite-sample standard error of the earnings estimate may not exist, whereas the percentiles always exist. (This is of course no concern for the employment estimates, which are constrained to lie within the interval 0 to 1.)

Table 5.1: Nonparametric estimates of local average treatment effects (with range restrictions)

	CIV (including covariates)					IV (without covariates)				
	Compliers		Treated compliers			Fuller 1		2SLS		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Employ	*Earnings*	*Employ*	*Earnings*	w_c *100	*Employ*	*Earning*	*Employ*	*Earning*	w_c *100
Aggregated effects										
Θ_c	** 0.18 (0.09)	827 [a]	** 0.18 (0.09)	906 [a]	-	*** 0.15 (0.06)	589 (388)	*** 0.16 (0.07)	601 [b]	-
$\Theta_{c,trim}$	** 0.21 (0.09)	313 [a]	** 0.16 (0.10)	394 (2143)	-	** 0.13 (0.07)	141 (399)	* 0.12 (0.08)	371 [b]	-
Θ_{obs}	*** 0.21 (0.11)	-1361 [a]	*** 0.19 (0.10)	- 424 [a]	-	0.02 (0.08)	202 (519)	-0.03 (0.10)	3015 [b]	-
$\Theta_{obs,trim}$	** 0.23 (0.12)	- 376 [a]	** 0.19 (0.12)	- 1460 [a]	-	0.11 (0.08)	- 65 (547)	0.13 (0.10)	441 [b]	-
Θ_{MD}	* 0.13 (0.07)	516 (517)	0.06 (0.07)	627 (494)	-	0.06 (0.05)	343 (327)	* 0.09 (0.06)	-	-
Effects for individual labour markets										
2	-0.74 (0.33)	- 1058 [a]	-0.58 (0.42)	-861 [a]	4.5	-1.00 (0.21)	- 3328 (1361)	-1.00 (0.22)	- 3481 [a]	3.9
7	-0.52 (0.57)	- 228 [a]	-0.86 (0.61)	-2054 [a]	1.8	-1.00 (0.45)	- 2308 (1653)	-1.00 (0.55)	- 2604 [b]	1.6
8	0.07 (0.52)	3681 [a]	0.15 (0.45)	1300 [a]	4.9	-0.82 (0.48)	- 334 (1407)	-1.00 (0.64)	0 [b]	2.4
9	0.47 (0.49)	1171 [b]	0.55 (0.53)	2660 (8598)	1.6	0.46 (0.23)	2586 (1712)	0.47 (0.30)	2988 [b]	1.3
10	0.00 (0.42)	4591 [b]	0.19 (0.44)	1306 [a]	0.7	0.27 (0.33)	3203 (2450)	0.44 (0.54)	7556 [b]	1.0
11	0.74 (0.27)	2907 [a]	0.77 (0.28)	2953 [a]	19.3	0.86 (0.18)	4392 (1317)	0.91 (0.18)	4640 [a]	16.8
12	0.40 (0.59)	2037 [b]	0.74 (0.57)	1686 [b]	5.1	0.46 (0.24)	153 (1082)	0.52 (0.28)	202 (2142)	6.0
13	-0.10 (0.12)	591 (648)	-0.10 (0.14)	747 (723)	17.9	0.03 (0.11)	1711 (690)	0.04 (0.11)	1749 (710)	17.7
15	-1.00 (0.71)	- 10282 [b]	-1.00 (0.66)	-4991 [a]	1.4	-1.00 (0.62)	- 6228 (4873)	-1.00 (0.80)	-12605	1.8
16	1.00 (0.64)	-27019 [b]	1.00 (0.68)	-11648 [b]	0.3	-0.14 (0.49)	506 (3994)	-1.00 (0.75)	53205 [b]	0.2
17	-0.37 (0.41)	- 1441 [a]	-0.23 (0.35)	356 [a]	2.5	-0.26 (0.24)	- 2268 (1755)	-0.22 (0.30)	- 2786 [b]	1.8
19	0.21 (0.22)	1158 (1445)	0.14 (0.22)	864 (1569)	14.5	0.07 (0.18)	-350 (1105)	0.08 (0.20)	- 357 (1188)	15.4
21	0.48 (0.25)	606 (1475)	0.47 (0.22)	849 (1575)	6.3	0.32 (0.21)	193 (1179)	0.34 (0.22)	219 (1281)	7.1
22	0.92 (0.55)	- 692 [a]	0.09 (0.51)	2729 [a]	1.7	-0.10 (0.32)	- 4736 (2531)	-0.15 (0.50)	- 6523 [b]	2.8
23	0.12(0.22)	149 (1779)	-0.03 (0.17)	294 (877)	11.7	0.04 (0.14)	- 1114 (865)	0.05 (0.14)	- 1137 (896)	12.2
24	0.23 (0.47)	- 419 (8871)	0.16 (0.47)	-283 [b]	2.7	0.57 (0.27)	2523 (1629)	0.66 (0.34)	3032 [b]	2.9
25	0.00 (0.25)	-12231 [a]	0.00 (0.27)	-14622 [b]	0.4	-0.13 (0.22)	366 (1637)	0.00 (0.21)	0 [b]	0.8
28	-0.20 (0.38)	- 920 (7740)	-0.24 (0.43)	-899 [a]	2.7	0.10 (0.23)	871 (1420)	0.11 (0.26)	941 (2009)	4.2

Note: In the first five rows the aggregated estimates are given. Θ_c is based on weights w_c that are proportional to the number of compliers (=fraction of compliers multiplied by number of observations). $\Theta_{c,trim}$ is based on the same weights, but using only the labour markets where no range restrictions had to be imposed. Θ_{obs} and $\Theta_{obs,trim}$ are based on the number of observations. Θ_{MD} uses minimum distance weights. ***, **, * indicates significance at the 1%, 5% and 10% level, respectively. Inference is based on the bootstrap, using the percentile method for the CIV estimator and a nested bootstrap of the T-statistic for the Fuller and the 2SLS estimator. There are 59 regressors 32634 observations.

a) The bootstrap standard error is larger than 10000.

b) The bootstrap standard error is larger than 100000.

We replicate the entire estimation process within each bootstrap replication, including the estimation of the "propensity score". 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 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.

Our preferred specification, the rather precisely estimated *compliers-weighted* LATE Θ_c, suggests that employment increased between 0.15 and 0.18 due to the Swiss active labour market policies. This represents about two months of additional employment within the subsequent 12 months. For the IV estimation without covariates, this result appears to be stable across estimators. For the CIV specification, it appears not to be sensitive to the distribution of the covariates between participating and non-participating compliers. The earnings effect shares the same features as the employment effects, but the estimates are very noisy. We can neither exclude that the effect on monthly earnings equals 570 CHF, which would be expected if the additional employment is paid at the average wage (because $0.15 \cdot 3800$ CHF = 570 CHF), nor that the earnings effect is zero. (In the Appendix, we conduct a number of sensitivity analyses, which all confirm a positive employment effect.)

We aggregated various subgroups of labour markets that are fairly homogeneous with respect to certain regional measures to better understand whether the considerable heterogeneity visible in the estimates for the singe labour markets is related to particular features of those local labour markets. The key finding we obtain for the employment effects is that the programmes seem to be more effective in labour markets that include larger cities (8, 11, 13, 19, 21, 22, 23). For the larger cities the complier weighted employment effect is 0.29 (standard error: 0.11) for the compliers and 0.25 (0.10) for the treated compliers, whereas in the remaining labour markets the corresponding effects are -0.14 (0.18) and -0.06 (0.19).

The finding that the CIV estimates are rather similar to the IV estimates without covariates may indicate that potential worries about confoundedness of the instrument turned out to be of less concern: There are differences in the populations living on the two sides of the border but these do not seem to be systematic with respect to the outcome variables.

Table 5.2 shows the expected potential outcomes separately for the always-participants, compliers and the never-participants, obtained from the CIV estimator and averaged over the 18 local labour markets. Although we cannot identify the treatment effects for the always- and the never-participants, we observe that the non-treatment outcome for the never-participants $E[Y^0 \mid T = n]$ is much higher than the non-treatment outcome for the compliers $E[Y^0 \mid T = c]$. At the same time, the treatment outcome $E[Y^1 \mid T = a]$ for the always-participants is much lower than the treatment outcome for the compliers $E[Y^1 \mid T = c]$. Hence, it seems that the group of never-participants is the group with highest re-employment chances, whereas the group of always-participants has a lower re-employment rate. The compliers are in between these two groups. Our interpretation of these estimates is that the never-participants are least likely to participate in ALMP because caseworkers may expect them to find a job on their own, whereas the always-participants are the difficult cases who are considered most in need of assistance. The compliers are an intermediate group. This is important to keep in mind when interpreting the overall positive effects of ALMP found here.

Table 5.2: Aggregated estimated outcomes of always- and never-participants (CIV estimator)

Employment				*Earnings*			
$E(Y^1 \mid T=c)$	$E(Y^0 \mid T=c)$	$E(Y^1 \mid T=a)$	$E(Y^0 \mid T=n)$	$E(Y^1 \mid T=c)$	$E(Y^0 \mid T=c)$	$E(Y^1 \mid T=a)$	$E(Y^0 \mid T=n)$
0.70	0.51	0.61	0.65	2874	2047	2258	2541

Note: Weighted average for the 18 local labour markets, weighted by the numbers of compliers, always- and never-participants.

6 Discussion and conclusions

In this paper, we estimated the effects of participation in active labour market policies (ALMP) on employment and earnings using nonparametric IV methods. The instrument is based on a mandated quota, which stipulated the minimum number of places a canton had to provide. This quota introduced exogenous variation in the likelihood of being assigned to ALMP. Since the quota differs even within homogenous labour markets, it is used as a local instrument within small local labour markets. To counter concerns about potential confoundedness of this instrument, a large number of observed individual characteristics were controlled for. The results were robust to the inclusion of these covariates, as well as to a large number of additional robustness checks that have been performed. Thus, confoundedness and fragility of the final estimates are not a serious concern.

We showed not only how to identify and estimate local average treatment effects in such specific labour markets, but also how to identify potential outcome distributions for always- and never-participants. Since the nonparametric estimates within each labour market were rather noisy, we proposed different aggregation schemes with desirable interpretations. They permit us to summarize the information in the local estimates in a concise way.

Our main results suggest that employment increased for the compliers by about *two months* per year due to participation in ALMP. These relatively large effects may be related to the evolution of the business cycle: While the unemployment rate was still very high during the first half of 1998, it decreased to only 1.5% in the year 2001. This is in line with the findings in Lechner and Wunsch (2009), who found for West Germany, that the effects tend to be larger when unemployed entered in ALMP during recession periods rather than during boom periods.

Compared to previous research for Switzerland, the size of the estimated effects is comparable to the results of Gerfin and Lechner (2002), Gerfin, Lechner, and Steiger (2005) and Lalive, van Ours, and Zweimüller (2000). One difference to those studies is the aggregation of the active labour market programmes, as they distinguish the effects for different types of programs. In this paper, all labour market programmes are aggregated into one group, because disaggregated effects by programme type are not identified with this instrumental variable strategy.

References

Abbring, J.H., and G. van den Berg (2003), "The Nonparametric Identification of Treatment Effects in Duration Models," *Econometrica*, 71, 1491-1517.

Beran, R. (1987), "Prepivoting to reduce level error of confidence sets," *Biometrika*, 74, 457-468.

Black, S. (1999), "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114, 577-599.

Cameron, C., and P. Trivedi (2005), *Microeconometrics: Methods and applications*, Cambridge University Press, Cambridge.

Card, D., and A. Krueger (1994), "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review*, 84, 772-793.

Frölich, M. (2004), "Finite Sample Properties of Propensity-Score Matching and Weighting Estimators," *Review of Economics and Statistics*, 86(1):, 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 (2), 214-227.

Frölich, M., and M. Lechner (2006), "Exploiting regional treatment intensity for the evaluation of labour market policies". IZA Discussion paper 2144. CEPR Discussion paper 5728. This discussion paper is accompanied by an internet appendix available on *www.sew.unisg.ch/lechner*.

Fuller, W. (1977), "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939-953.

Gerfin, M., and M. Lechner (2002), "Microeconometric Evaluation of the Active Labour Market Policy in Switzerland," *The Economic Journal*, 112 (482), 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, 2005.

Hahn, J., P. Todd, and W. van der Klaauw (2001), "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201-209.

Hall, P. (1986), "On the bootstrap and confidence intervals," *Annals of Statistics*, 14, 1431-1452.

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

Holmes, T. (1998), "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders," *Journal of Political Economy*, 106, 667-705.

Imbens, G. W. (2004), "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, 86(1), 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 (2000), "The Impact of Active Labour Market Programs on the Duration of Unemployment," IEW Working Paper No. 41, University of Zurich.

Lechner, M., and C. Wunsch (2009), "Are Training Programs More Effective When Unemployment is High?", *Journal of Labor Economics*, 27, 653-692, 2009,.

Loh, W.-Y. (1987), "Calibrating confidence coefficients," *Journal of the American Statistical Association*, 82, 155-162.

Pence, K. (2006), "Foreclosing 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.

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.

Wald, A. (1940), "The Fitting of Straight Lines if Both Variables are Subject to Error," *Annals of Mathematical Statistics*, 11, 284-300.

Appendix: Sensitivity analysis

In this section, we examine the sensitivity of the results presented above by checking a variety of alternative specifications. (We thank the referees for many helpful suggestions.)

One concern may arise regarding the reliability of the bootstrap inference. As mentioned in Section 3, we cap the estimates of the conditional means of the outcome variables at the boundaries of their logical support. The employment variable has to lie in the interval 0 to 1, whereas the earnings variable has to be non-negative. Negative estimates are set to zero and, for the employment outcomes, estimates larger than one are set to one. These range restrictions thus represent a non-differentiable transformation of the original estimate. To examine whether this non-differentiability has any effect on the finite sample distribution of the bootstrap, we explore an alternative approach to impose the logical range restrictions in a smooth, i.e. differentiable, way. Let β be the unconstrained estimate, e.g. of $E[Y^1 \mid T = c]$ or $E[Y^0 \mid T = c]$. For the outcome variable employment, which has to lie within 0 and 1, we impose the *smooth* range restrictions by calculating the constrained estimate as $\Phi(3 \bullet (\beta - 0.5))$, where Φ is the Gaussian cdf. When the unconstrained estimate β lies within 0.1 and 0.9, the transformation is nearly linear and essentially reproduces β. When the unconstrained estimate β is smaller than 0.1, including negative values, or larger than 0.9, the transformed estimate converges smoothly to 0 and 1, respectively.

(For the outcome variable earnings, which have to be non-negative, we apply the transformation $\Phi(\beta) \cdot \beta + \phi(\beta)$, where Φ is the Gaussian cdf and ϕ is the Gaussian pdf, which corresponds to the expected value of the Tobit model. In any case, the earnings estimates remain very noisy, irrespective of smoothing, and the confidence intervals are always very wide and cover zero, such that the earnings estimates are not discussed any further in the following.)

The first row of Table A.1 reproduces the employment estimates of Table 5.1, when using the non-smooth range restrictions. The second row of Table A.1 gives the estimates when using the smooth range restrictions. The estimated effects are nearly identical to the previous ones, and the bootstrap inference produces essentially the same results. The bootstrap distribution of the smoothed estimator is very similar to the bootstrap distribution of the non-smoothed original estimator.

The third row of Table A.1 gives the employment effects when no range restrictions are imposed. That is, negative estimates of $E[Y^1 | T = c]$ or $E[Y^0 | T = c]$ as well as estimates larger than one enter unconstrained into the aggregation procedure. It turns out that the aggregated effects are still similar to the previous ones, yet the standard errors become huge since some single estimates are very large (in absolute value). (It may well be that the finite sample standard errors are infinite, depending on the true distribution of the error terms. Therefore, we mainly examine the quantiles (which always exist) of the bootstrap distribution.)

The following sensitivity analyses examine the robustness of the estimation results to various changes of the population. We examine how the aggregated effect changes when some of the 18 local labour markets are dropped. First, we examine the estimates when we drop each one of the 18 markets in turn to see how much the aggregated effect is driven by a single labour market only. We obtain the following 18 estimates for Θ_c when we drop each labour market in turn: 0.226, 0.196, 0.189, 0.178, 0.184, 0.050, 0.171, 0.245, 0.200, 0.181, 0.197, 0.179, 0.163, 0.171, 0.192, 0.181, 0.184, 0.225. For Θ_{obs} we obtain: 0.258, 0.229, 0.228, 0.211, 0.217, 0.107, 0.202, 0.240, 0.277, 0.164, 0.221, 0.214, 0.200, 0.178, 0.218, 0.212, 0.221, 0.246. Hence, whichever labour market we drop, the estimated employment effects remain positive in any case. They obtain their lowest value (Θ_c = 0.050, Θ_{obs} =0.107), when labour market 11 is dropped. On the other hand, the largest estimate for Θ_c is 0.245 and for Θ_{obs} is 0.277.

(Note that these were the results with the non-smooth range restrictions. When we impose the smooth range restrictions, for Θ_c we obtain the estimates: 0.221, 0.191, 0.184, 0.174, 0.180, 0.048, 0.165, 0.241, 0.195, 0.176, 0.191, 0.176, 0.160, 0.167, 0.184, 0.177, 0.179, 0.189 and for Θ_{obs} we obtain: 0.251, 0.222, 0.220, 0.205, 0.210, 0.102, 0.192, 0.234, 0.270, 0.157, 0.213, 0.208, 0.194, 0.175, 0.210, 0.206, 0.214, 0.217.)

Next, we examine the aggregated effects when we drop several labour markets at the same time. The subsequent rows of Table A.1 give estimation results for various alternative specifications. Rows 4 to 21 give the results for various subsets of the 18 local labour markets of the main specification. Only the employment estimates are shown since the earnings estimates were always very noisy (as before). The results are given with and without smooth range restrictions. Rows 4 and 5 present the estimates without local labour markets (2, 15, 17), where the estimated fraction of compliers had the wrong sign, i.e., where the estimate of the right-hand side of equation (10) was negative. In principle, we would expect all labour markets to be located in the upper-right or lower-left quadrant of Figure 4.1. Those three labour markets, however, are located in the upper-left quadrant, which would indicate that an increase in the quota is estimated to lower treatment probability. (However, this is statistically significantly only the case for labour market 2. The confidence intervals for the estimated fraction of compliers in labour markets 15 and 17 cover negative as well as positive values.) The estimates in Table A.1 remain positive and are somewhat (though not statistically significantly) larger.

Next, we also delete those labour markets where the difference in the quota (last column of Table 4.2) is very small (8, 11, 12). In those markets, we expect the instrument to be very weak because the difference in the quota is so small that it may not have affected the two neighbouring cantonal administrations differently. The estimates become smaller but remain positive.

In the next rows, we delete labour markets where the difference in the quota has been rather large. Very large differences in the quota could in principle induce migration effects in that some people might decide to change their canton of residence in order to increase (or decrease) their probability of being sent to ALMP if they happen to become unemployed. In rows 8 and 9 we delete labour markets 15 and 23 where the difference in the quota (last column of Table 4.2) is larger than eight. In rows 10 and 11, we delete all labour markets where the difference in the quota is larger than four. Finally, in rows 12 and 13 we delete all cantons where the quota itself is very high. In particular, we delete those cantons where the quota in column 5 of Table 2.1 is above 20. These cantons (Uri, Schwyz, Ob/Nidwalden, Appenzell-Inner/Ausserrhoden) may be preferred by individuals who want to live in a region where their treatment probability is largest. Deleting these cantons implies deleting labour markets 15, 17, 23 and 28. In all these cases, the estimates in Table A.1 remain positive and similar to the main results. (Note that we did not delete cantons where the quota is very low since we cannot identify clear outliers in column 5 of Table 2.1. Cantons with a low quota include all regions of Switzerland: German, French, and Italian parts, urban as well as rural areas. On the other hand, those cantons with a high quota are all small cantons in the German speaking mountainous regions of Central and Eastern Switzerland.)

In rows 14 to 21 we examine subsets of labour markets, where the employment effects are presumably more reliably estimated, such that the aggregated employment effect would tend to be more reliable. As observed from Table 5.1, several of the estimated treatment effects for the separate labour markets are very noisy and unreasonably large. These are usually those labour markets where the number of observations and/or the fraction of compliers is very small. The aggregated treatment effects may thus be prone to some components consisting of noisily estimated outliers. In the standard specification, we addressed this problem with the minimum-distance weights, which considerably down-weight those labour markets. Rows 14 to 21 in Table A.1 examine alternative approaches to purge the aggregated effect from single noisy estimates. In rows 14 and 15, we delete the labour markets 15 and 16 where the estimated treatment effect in Table 5.1 had been 1 and –1, respectively. In rows 16 and 17, we delete all labour markets (2, 7, 11, 15, 16, 22) where the estimated treatment effect was larger than 0.50 in absolute value in Table 5.1. In rows 18 and 19, we include only those labour markets with the largest *number* of compliers (w_c > 10% in column 5 of Table 5.1). These are labour markets 11, 13, 19 and 23. Finally, in rows 20 and 21, we dropped labour market 11 where the estimated treatment effect is unreasonably large. (This specification represents the subset of labour markets where the number of compliers w_c is larger than 10% and the estimated treatment effect is smaller than 0.50 in absolute value.) In all these scenarios, the estimated treatment effects remain positive.

In rows 22 to 29, we examine various subpopulations of the main population. As a sensitivity analysis of our identifying assumptions, we drop the groups of unemployed with high and low previous earnings from our population. We cannot rule out that (some of the) high earners might have chosen residence based on differences in taxation between cantons. Such differences should however be of little relevance to low- and medium-income earners as the Swiss tax system is progressive. In rows 22 and 23 we delete all individuals with insured monthly earnings larger than 6000 Swiss Franc. In the subsequent row, we also drop all individuals with earnings larger than 5000 Swiss Franc. Thereby, we lose 10% and 21%, respectively, of the observations. In the next rows, we drop also all low-income individuals with insured earnings below 2000 Swiss Francs, which implies a loss of another 9% of the observations. Here, one might have been concerned about regional differences in social assistance or other policies for low-income people. In all cases, the estimates remain positive and similar to the main specification (though less precise due to the smaller sample size).

Table A.1: Sensitivity Analysis: Employment effects on compliers (CIV)

		Θ_c	$\Theta_{c,trim}$	Θ_{obs}	$\Theta_{obs,trim}$
1) Main specification:	non-smooth range restrictions	** 0.18 (0.09)	** 0.21 (0.09)	*** 0.21 (0.11)	** 0.23 (0.12)
2)	smooth range restrictions	** 0.18 (0.09)	-	*** 0.21 (0.11)	-
3)	no range restrictions	0.15 (16.50)	-	0.25 (17.46)	-
Subset of labour markets:					
4) without # 2, 15, 17	non-smooth range restrictions	*** 0.26 (0.10)	*** 0.26 (0.10)	*** 0.34 (0.11)	*** 0.30 (0.12)
5)	smooth range restrictions	*** 0.25 (0.09)	-	*** 0.33 (0.11)	-
6) without # 2,8,11,12,15,17	non-smooth range restrictions	0.12 (0.10)	0.10 (0.10)	*** 0.26 (0.12)	0.14 (0.12)
7)	smooth range restrictions	0.11 (0.10)	-	*** 0.25 (0.12)	-
8) without # 15, 23	non-smooth range restrictions	*** 0.21 (0.09)	** 0.23 (0.10)	*** 0.29 (0.11)	** 0.24 (0.12)
9)	smooth range restrictions	*** 0.20 (0.09)	-	*** 0.28 (0.11)	-
10) without # 15, 22, 23	non-smooth range restrictions	** 0.20 (0.10)	** 0.23 (0.10)	*** 0.25 (0.11)	** 0.24 (0.12)
11)	smooth range restrictions	** 0.19(0.09)	-	*** 0.25 (0.11)	-
12) without # 15, 17, 23, 28	non-smooth range restrictions	*** 0.24 (0.10)	*** 0.24 (0.10)	*** 0.31 (0.12)	** 0.26 (0.13)
13)	smooth range restrictions	*** 0.23 (0.10)	-	*** 0.30 (0.11)	-
14) without # 15, 16	non-smooth range restrictions	** 0.20 (0.09)	** 0.21 (0.10)	** 0.23 (0.11)	** 0.23 (0.12)
15)	smooth range restrictions	** 0.19 (0.09)	-	** 0.22 (0.10)	-
16) without # 2,7,11,15,16,22	non-smooth range restrictions	0.11 (0.10)	0.12 (0.11)	0.13 (0.12)	0.16 (0.14)
17)	smooth range restrictions	0.11 (0.10)	-	0.13 (0.12)	-
18) only # 11, 13, 19, 23	non-smooth range restrictions	** 0.27 (0.11)	** 0.27 (0.11)	*** 0.34 (0.13)	*** 0.34 (0.13)
19)	smooth range restrictions	*** 0.26 (0.11)	-	*** 0.33 (0.12)	-
20) only # 13, 19, 23	non-smooth range restrictions	0.06 (0.11)	0.06 (0.11)	0.10 (0.13)	0.10 (0.13)
21)	smooth range restrictions	0.06 (0.11)	-	0.10 (0.13)	-
Subpopulation of unemployed:					
22) income < 6000 CHF	non-smooth range restrictions	*** 0.21 (0.09)	*** 0.23 (0.09)	* 0.15 (0.11)	*** 0.27 (0.11)
23)	smooth range restrictions	*** 0.19 (0.08)	-	0.13 (0.10)	-
24) income < 5000 CHF	non-smooth range restrictions	** 0.17 (0.10)	** 0.22 (0.12)	* 0.16 (0.12)	*** 0.30 (0.15)
25)	smooth range restrictions	** 0.16 (0.10)	-	* 0.15 (0.11)	-
26) income >2000 and < 6000	non-smooth range restrictions	*** 0.22 (0.10)	*** 0.23 (0.11)	*** 0.24 (0.11)	** 0.25 (0.13)
27)	smooth range restrictions	*** 0.22 (0.09)	-	*** 0.25 (0.10)	-
28) income >2000 and < 5000	non-smooth range restrictions	** 0.17 (0.11)	** 0.20 (0.13)	*** 0.20 (0.12)	*** 0.23 (0.13)
29)	smooth range restrictions	** 0.17 (0.11)	-	*** 0.20 (0.11)	-

Note: See note below Table 5.1.

Various alternative sensitivity analyses were conducted in the discussion paper Frölich and Lechner (2006) and its accompanying appendix. We summarize the basic results here: First, a finer bandwidth grid with 30 values $\{1/100, 1.2/100, 1.2^2/100, ...,\cdot 1.2^{28}/100, \infty\}$, instead of 10, was used when choosing the bandwidths by cross-validation. The results were very similar. Second, alternative sets of control variables were tried, including a very large set with 106 variables, a few intermediate sets and finally a smaller set, which contained only those 42 variables that did not create any collinearity or perfect prediction problems for the probit estimator during the bootstrap. Generally, the results for the estimated Θ were similar and always positive. Third, as another alternative we used only the 8 local labour markets with the most similar ALMP compositions. The complier-weighted average effects were somewhat larger and about 0.24 for employment and 800 CHF for earnings with similar significance levels.