

EXPLOITING REGIONAL TREATMENT INTENSITY
FOR THE EVALUATION OF LABOUR MARKET POLICIES

Markus Frölich
Michael Lechner

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP11/06

Exploiting Regional Treatment Intensity for the Evaluation of Labour Market Policies

Markus Frölich

*SIAW, University of St.Gallen,
IFAU Uppsala and IZA Bonn*

Michael Lechner

*SIAW, University of St. Gallen,
CEPR, ZEW, PSI and IZA Bonn*

Discussion Paper No. 2144
May 2006

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Exploiting Regional Treatment Intensity for the Evaluation of Labour Market Policies^{*}

We estimate the effects of active labour market policies (ALMP) on subsequent employment by nonparametric instrumental variables and matching estimators. Very informative administrative Swiss data with detailed regional information are combined with exogenous regional variation in programme participation probabilities, which generate an instrument within well-defined local labour markets. This allows pursuing instrumental variable as well as matching estimation strategies. A specific combination of those methods identifies a new type of effect heterogeneity. We find that ALMP increases individual employment probabilities by about 15% in the short term for unemployed that may be called 'marginal' participants. The effects seem to be considerably smaller for those unemployed not marginal to the participation decision.

JEL Classification: J68, C14, C21

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

Corresponding author:

Markus Frölich
SIAW
University of St. Gallen
Bodanstr. 8
CH-9000 St. Gallen
Switzerland
Email: Markus.Froelich@unisg.ch

^{*} 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 for helpful comments. The usual disclaimer applies.

1 Introduction*

In the 1990's many continental 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 for example by Fay (1996), Heckman, LaLonde, and Smith (1999), Kluve (2006), Kluve and Schmidt (2002), and Martin and Grubb (2001) which attempts to answer the important question whether such policies actually benefited those unemployed who participated in these programmes. Particular examples for Switzerland are Lalive, van Ours and Zweimüller (2000) and Gerfin and Lechner (2002).

Lacking a consensus about the impact of these policies on the employment chances of the programme participants, together with the almost complete absence of random experiments in Europe brings the issue of identification of such effects from observational data to the forefront of the discussion. It is probably fair to say that the vast majority of European studies either use matching estimation, based on the assumption that the data are informative enough to control for any selection bias, 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

* 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 for helpful comments. The usual disclaimer applies.

the credibility of both assumptions depends on the institutional setting and the data available, they were frequently attacked as not being plausible.

In this paper we want to add to the knowledge about the effects of ALMP by considering a different route to identification. We use exogenous differences of 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. Gerfin and Lechner (2002) argued that this database allows to control for all variables that jointly influence programme participation and labour market outcomes, such that a selection on observables assumption becomes plausible. Here, we use this approach as well and combine it with the IV estimation. Thereby, we can identify also the effects for the groups not responsive to this instrument and investigate the effect heterogeneity.

The idea of exploiting geographic borders as an instrumental variable to uncover the effects of policy interventions has been used in other studies. Card and Krueger (1994), Holmes (1998), Black (1999) and Pence (2003) use the U.S. state or district borders to estimate the effects of the minimum wage on employment, of right-to-work laws on manufacturing activity, of school quality on housing prices and of foreclosure laws on mortgage loan size, respectively. 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 in specifying linear IV models and thus to implicitly aggregate these different heterogeneous local effects into a single parameter. Since the homogeneity and functional form assumptions implicit in this approach are undesirable and most likely lead to inconsistent estimates, we introduce a *non-parametric* instrumental variable approach that allows for differences in observable characteristics across the local borders and propose specific aggregating schemes for the local effects. We find that the nonpara-

metric IV estimates and the estimates based on the parametric specification are fairly different (though both have the same sign), and that the magnitude of the nonparametric estimates is clearly more plausible.

To be more specific about the application, we exploit that Switzerland is a small country with autonomous administrative regions, and runs an extensive active labour market policy to counteract unemployment. With short commuting times and a good 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. To our knowledge, such an approach is new to the evaluation of labour market programmes.

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 (2006a). 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. We propose aggregation schemes that have desirable interpretations for such cases. Fourth, we show that by combining conditional IV and matching estimation, we identify treatment

effects for the groups of individuals that do not react to a change in the instrument, the so-called non-compliers. Subsequently, we try to obtain further insights into treatment effect heterogeneity, in particular whether we observe similar effects for those who would never participate even when the instrument changes compared to those who would always participate. Fifth, we find that ALMP increases individual employment probabilities by about 15% in the short term for the compliers. The effects seem to be much smaller for the never-participants and also for the always-participants. This casts doubts on the belief that Swiss case workers send those unemployed into the programmes who benefit the most from it.

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 our econometric approach for unconditional and conditional IV estimation as well as matching. Section 4 discusses the particular implementation for Switzerland. Section 5 gives the estimation results, and Section 6 concludes. An appendix contains material concerning the properties of the estimators. Additional background material is available on the internet at www.siaaw.unisg.ch/lechner/iv_ma.html.

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 the USA, UK and Germany as well. Programmes can be grouped into three categories: a) *Training* programmes range from basic skills courses, language courses, computer courses to specific work-related training (e.g. business and trade training, manufacturing and technical training), with a usual duration of 1 week to 3 months and are carried out by private education providers; b) *Employment programmes* are temporary job creation schemes and con-

sist of provisional or project work for about 3-6 months in the public administration or other public institutions (hospitals, old people's homes, nursing homes, schools, and kindergartens) or in private not-for-profit institutions (e.g. charities, cultural, environmental, recycling organisations). c) *Temporary wage subsidies* (Zwischenverdienst) is a programme rather unique to Switzerland and is a subsidy for temporary jobs in the regular labour market. Whereas employment programmes should not compete with regular jobs in the private or public sector, temporary wage subsidies are for jobs in the regular labour market. The subsidy consists of 80% of the pay difference to the previous earnings and is paid to the unemployed person. The subsidy can be granted for up to one year, with an average duration of about 4 months. During participation in ALMP the unemployed must continue job search activities and accept any reasonable job offer.¹

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 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 (REO), supervised by 26 cantonal centres. These centres, located in the cantonal capital (see Figure 2.1) contract private and public organisations for providing programmes, compile a catalogue of courses and programmes offered by the contracted providers and seek to ensure that a sufficient number of programme places can be offered as demanded by the REO. The REO are geographically organised, each REO serving several municipalities. For each unemployed there is one unique REO defined by place of residence. They cannot change their assigned REO other than by moving to another municipality. Exceptions are the city centres of Zurich and Geneva, which are served by several REO.

¹ 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).

Figure 2.1: The cantonal borders of Switzerland



Note: http://www.kyte.de/mygeo/karten/switzerland/landkarte_schweiz.jpg.

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 25'000 year-places (each representing 220 programme days) and was distributed across the cantons according to the formula

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

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 for Switzerland.

Table 2.1: Minimum quotas and number of unemployed

Canton	Minimum quota ^a		Number of unemployed	Quota per unemployed ^b	Quota per unemployed ^c	Realised places ^d	Number of unemployed	Places per unemployed ^e
	1997	1998	Jan 1998	Jan 1998	Dec 1998	total 1998	average 1998	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: Zurich, BE: Bern, LU: Lucerne, 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-Outer-Rhodes, AI: AppenzellInner-Rhodes, SG: StGall, GR: Grisons (Graubünden), AG: Argovia (Argau), TG: Thurgovia (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.

The costs of active labour market programmes and of their administration generally are borne by the federal unemployment insurance funds.² The cantons pay a very small lump sum contribution of 3000 Swiss Francs (CHF) per year-place for their assigned minimum quota. They can reduce this lump sum payment by up to

² Art. 92 AVIG (Arbeitslosenversicherungsgesetz), Art. 122a, 122b AVIV (Arbeitslosenversicherungsverordnung).

25% if the average unit costs of the purchased programme slots are below the national average within defined programme categories. No financial contribution has to be paid for places filled beyond the required minimum.³ 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.⁴ 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, as given in Table 2.1, were proclaimed.⁵ This formula for the computation of the minimum quota induced regional variation in treatment (participation) intention. Relative to the number of unemployed persons, 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.

Consider the situation of the management of the cantonal and the regional employment offices in the beginning of 1998 for planning their strategy in providing active labour market programmes. In the fourth and fifth column in 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 StGall (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 active labour market 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. Probably though, the management did not assume that the number of unemployed would remain constant throughout 1998 and their

³ Art. 72c (AVIG) and Art. 98b (AVIV).

⁴ Art. 72a (AVIG) and Art. 98b (AVIV).

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 at the middle or 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 persons to labour market 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 with local labour markets in which 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 strat-

⁵ Art. 72b (AVIG), Art. 99, 99a (AVIV). Regulation in *Amtliche Sammlung AS 1996 3071, AS 1997 2415*. See also Stauffer, 1998.

egy is described in the next section. Before that, we give some details on the data, as they might be relevant for judging the plausibility of the identifying assumptions in Section 3.

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 comes 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 1999. They were obtained from databases of the unemployment insurance system (AVAM/ASAL) and the social security records (AHV). The personal information is about 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. The evaluation studies of Gerfin and Lechner (2002) and Gerfin, Lechner, and Steiger (2005) used these very informative data to estimate average treatment effects using matching estimators based on a selection on observables assumption. Based on these prior experiences, we first analyze the identifying power of the instrument and then add a selection on observables assumption as in the two studies mentioned above.

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

In this section, we discuss two nonparametric identification strategies for identifying and estimating the effects of participation in ALMP on employment. The first strategy 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 (2006a). The second strategy exploits the very informative data for which a selection on

observable strategy appears to be plausible as well. Finally, the information obtained under those two types of assumptions is combined.

3.1 *Instrumental variables with homogeneous effects*

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 labour market 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 labour market programmes (see 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. Let Z denote the instrument *quota per unemployed*. The data consist of a large random sample $\{y_i, d_i, z_i, x_i\}_{i=1}^N$, where X represents individual characteristics as explained later. Allowing for a short-cut notation, a conventional instrumental variable analysis proceeds by specifying a linear model, such as

$$Y = \bar{\alpha} + \bar{\beta}D + U, \quad E(U | Z) = 0, \quad P(D = 1) \neq P(D = 1 | Z), \quad (1)$$

and estimating it, for example, by 2SLS using Z as the instrumental variable. 2SLS is consistent for the coefficients $\bar{\alpha}$ and $\bar{\beta}$ if Z is uncorrelated with U and correlated with D . As argued above this exclusion restriction is unlikely to be satisfied in this simple model, though, since the definition of Z is related to the cantonal unemployment rate in 1996.

In confined regions along the internal administrative boundaries, however, the exclusion restriction should be valid. Thus, we are going to identify economically *integrated local labour markets* that are divided by a cantonal border. A local labour market is considered as integrated if the value of different job opportunities does not depend on the location of residence. In other words, all relevant employment opportunities can be

reached within convenient commuting distance (e.g. half an hour) from both sides of the border, such that the choice of workplace location and the choice of residence are not immediately tied. There should then be no opportunities for wage arbitrage by moving residence. Switzerland, with its numerous winding administrative borders (see Figure 2.1) and a very good commuting infrastructure, is a candidate country for finding such local labour markets. Nevertheless, the methods developed below apply more generally, because in many countries administrative borders pass through densely populated areas.

Based on the previous considerations, we might want to add local labour market fixed effects for individual i living in labour market m to the linear model (and eliminate all individuals not living in any such local labour market). Let $\alpha = (\alpha_1, \dots, \alpha_M)$ represent the vector of all labour market fixed effects and let L be the random vector indicating the labour market in which individual i lives. (I.e. L is a column vector of zeros with an entry of one at the location where individual i lives.) The model then is

$$Y = \alpha L + \bar{\beta} D + U, \quad E(U | Z, L) = 0, \quad P(D = 1 | L) \neq P(D = 1 | Z, L). \quad (2)$$

The fixed effects account to some extent for differences in regional industry structure that may be related to unemployment in 1996.

Even after including labour market fixed effects, there may be further differences between the cantons, e.g. in tax rates, social assistance, policies supporting families, children allowances, which attract different types of individuals to move to the one or the other side of the border. Hence, we condition also on characteristics X , such as family situation, education, profession, earnings and employment history, to compare only individuals that are identical in most respects except for that they live on different sides of the border:

$$Y = \alpha L + \bar{\beta} D + \bar{\gamma} X + U, \quad E(U | Z, L, X) = 0, \quad P(D = 1 | L, X) \neq P(D = 1 | Z, L, X). \quad (3)$$

In this model, the expected outcomes may vary across local labour markets through $\alpha = (\alpha_1, \dots, \alpha_M)$, but the treatment effect $\bar{\beta}$ is assumed to be homogenous. This homogeneity assumption may not be plausible out of, at least, two reasons: First, the treatment effect may depend on the local unemployment rate and the industry structure. Second, the types of ALMP may differ between the labour markets. Thus, the treatment

effects may vary between the local labour markets, and one might want to add a vector of treatment effects $\beta = (\beta_1, \dots, \beta_M)$ to the above model and also to permit the effects of X to differ across labour markets.

With all regressors interacted with a full set of local labour market dummies, this is identical to estimating 2SLS separately in each labour market. At this point, we might prefer to switch to a fully *nonparametric* approach in each labour market to avoid imposing a specific functional form between Y and the control variables X . Another advantage would be that nonparametric regression permits endogenous control variables X to some extent, whereas linear regression does not (see Frölich, 2006b). This matters, because some of the X variables, e.g. past employment histories, may be correlated with U , which is not of concern for the nonparametric approach, though. Most importantly, the nonparametric approach permits individual treatment effect heterogeneity: The treatment effect may not only vary across local labour markets but also between individuals within the same labour market. Finally, a bounded support of some outcome variables can be naturally incorporated into the nonparametric approach, as discussed below. As we are going to identify the potential outcomes separately, we can ensure that earnings cannot be negative and that the number of months employed in a year lies between 0 to 12 months.

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 3.3.

3.2 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 $Z \in \{\bar{z}, \underline{z}\}$, where \bar{z} is observed on the one side of the border and \underline{z} on the other side. Suppose \bar{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 re-integration strategies, which are partly influenced by the minimum quota the canton has to fulfil. REO in cantons with an ambitious quota per unemployed will assign earlier and more persons to programmes than cantons with a lower quota. This identification strategy is related to a regression-discontinuity design as in Hahn, Todd and van der Klaauw (2001).

Let D_z denote the potential participation status of an individual i if the level of the instrument were externally set to z . With the instrument taking only two different values, the potential participation variable D_z defines four different types of individuals denoted by $T \in \{a, n, c, d\}$. Following the literature, we call the different 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 \bar{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 $Z \in \{\bar{z}, \underline{z}\}$.⁶ 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 is 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.

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

⁶ With a single instrument at hand and without imposing further structure, only the effect of the active labour market policy as a mix of programmes is identified, but not the effects of single sub-programmes, like employment programmes or training courses. In matching, it is possible to isolate the effect of a sub-programme by deleting all observations that participated in any other type of labour market programmes (see Lechner, 2001). For identification by IV, however, we should delete all observations that would participate in other types of programmes if the value of their instrument were changed. However, it is impossible to observe who these unemployed are.

$$E[Y^1 - Y^0 | T = c] = \frac{E[Y | Z = \bar{z}] - E[Y | Z = \underline{z}]}{E[D | Z = \bar{z}] - E[D | Z = \underline{z}]} \quad (4)$$

This local average treatment 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 being 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 (2006a) extended those results to incorporate additional control variables X . The following assumptions are taken from Frölich (2006a), which condition on X . If the set of control variables X is *empty*, we will refer to them as the unconditional IV assumptions (IV.1) to (IV.4). These (IV.1) to (IV.4) assumptions represent a somewhat weaker version of the original assumptions of Imbens and Angrist (1994).⁷

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

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

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

$$P(T = t | X = x, Z = \underline{z}) = P(T = t | X = x, Z = \bar{z}) \quad \text{for } t \in \{a, n, c\}$$

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

$$\begin{aligned} E[Y^0 | X = x, Z = \underline{z}, T = t] &= E[Y^0 | X = x, Z = \bar{z}, T = t] && \text{for } t \in \{n, c\} \\ E[Y^1 | X = x, Z = \underline{z}, T = t] &= E[Y^1 | X = x, Z = \bar{z}, T = t] && \text{for } t \in \{a, c\} \end{aligned}$$

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

⁷ Their assumptions were (Exclusion restriction) $Y_{i,z}^d = Y_{i,z''}^d$ for all d, z', z'' , (Existence of instruments) $Y_i^0, Y_i^1, D_{i,z} \perp\!\!\!\perp Z_i$ for all $z \in \text{Supp}(Z)$ and $E[D | Z = z]$ is a nontrivial function of z , and (Monotonicity) for all pairs (z, z') either $D_i(z) \geq D_i(z')$ for all i or vice versa $D_i(z) \leq D_i(z')$ for all i .

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 organisational changes. Although it cannot be ruled out that an increase in the quota would also have changed the patterns of people being assigned to programmes, 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. $P(T = c | X)$ can be zero for some X , because values of X where $P(T = c | X) = 0$ 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 realised 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 both sides of the border do not differ systematically).

With respect to 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 also control for individual characteristics X .

With respect to 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.⁸ Nevertheless, to be on the safe

⁸ The internet appendix (Table IC.3) details the composition of the ALMP in the REO in the local labour markets. It shows the allocation of the treated to training, employment programmes, temporary wage subsidies and other programmes. When regressing the quota per unemployed of Table 2.1 on the average ALMP compositions in the cantons from Table IC.3, all coefficients are insignificant. A regression of the quota on the share in training yields a coefficient of 0.12 with a t-statistic of 1.3. A regression of the quota on the share in employment programmes gives -0.05 (t-statistic 0.4), and a regression on the share in temporary wage subsidies gives -0.15 (1.1). When regressing on the training share *and* the employment programme share, the t-statistics are 1.4 and 0.8. Similarly, the t-statistics are 0.8 and 0.5 when training share and temporary wages subsidy share are included as regressors, and 0.6 and 1.1 for employment programme share and temporary wages subsidy share.

side, we are going to 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, that the populations residing on the two sides of the border are identical in terms of their employability. The clear-cut 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 publicised 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 1999.⁹ 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, children 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 therefore 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

⁹ For example, a low unemployment rate in 1996 and 1998 would have resulted in a rather high quota per unemployed. If the good employment prospects continued into 1999, the instrument and the employment outcomes were positively correlated leading to upward biased estimates. However, other factors could also have induced a negative correlation, for example if the higher educated are more likely to migrate between cantons. Because the value of the instrument depends on the population and unemployment *shares* in 1996 divided by the *number* of unemployed in 1998, different cantonal population growth rates could have generated a negative correlation. In-migration would have reduced the quota per unemployed, because of an increase in the *number* of unemployed persons in 1998. Analogously, out-migration would have increased the quota per unemployed. If the higher educated (which enjoy

up differences in other characteristics. (Differences in local tax rates, for example, certainly would induce a sorting by incomes, but having included earnings and family composition in X , the impact of differential tax rates is controlled for.)

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. We do not want to control for variables that are already causally affected by the instrument, though. To be on the safe side, in Section 5 we control for a large number of covariates in the CIV estimator (defined below). Nevertheless, since it turns out that the results with and without X do not differ very much, confoundedness of the instrument might not really have been of much relevance.

Assumption (CIV.5) requires a common support of the characteristics X across the border, which is an assumption that we verify in Section 4.

3.3 *Conditional instrumental variables (CIV)*

The previous instrumental variable 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. Nevertheless, one might still be doubtful whether the populations on both sides of the border are comparable and might therefore want to condition on a number of individual characteristics X . Without covariates X , the complier effect was identified by (4). When including X , Frölich (2006a) shows that the treatment effect for the compliers $E[Y^1 - Y^0 | T = c]$ is nonparametrically identified as:

$$E[Y^1 - Y^0 | T = c] = \frac{\int (E[Y | X, Z = \bar{z}] - E[Y | X, Z = \underline{z}]) dF_X}{\int (E[D | X, Z = \bar{z}] - E[D | X, Z = \underline{z}]) dF_X}, \quad (5)$$

and can be estimated by a ratio of two matching estimators:

better employment prospects) are over-represented among the migrants, the instrument and the employment outcomes would be negatively correlated.

$$\overbrace{E[Y^1 - Y^0 | T = c]} = \frac{\sum_{i:z_i=\bar{z}} (y_i - \hat{m}_{\bar{z}}(x_i)) - \sum_{i:z_i=\underline{z}} (y_i - \hat{m}_{\underline{z}}(x_i))}{\sum_{i:z_i=\bar{z}} (d_i - \hat{\mu}_{\bar{z}}(x_i)) - \sum_{i:z_i=\underline{z}} (d_i - \hat{\mu}_{\underline{z}}(x_i))}. \quad (6)$$

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

With a similar reasoning, the treatment effect on the treated compliers is identified as well:

$$E[Y^1 - Y^0 | D = 1, T = c] = \frac{\int (E[Y | X, Z = \bar{z}] - E[Y | X, Z = \underline{z}]) \cdot \pi(X) \cdot dF_X}{\int (E[D | X, Z = \bar{z}] - E[D | X, Z = \underline{z}]) \cdot \pi(X) \cdot dF_X}, \quad (7)$$

where $\pi(x) = P(Z = \bar{z} | X = x)$, see proof in Appendix A.

Similarly, we identify the potential outcomes of the complier population separately by noting that:¹⁰

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

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

We obtain a similar expression for the treated compliers:

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

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

The corresponding matching estimators are analogous to equation (6).

¹⁰ These relationships are directly derived by obvious changes in the proofs in Frölich (2006a) and are omitted.

Estimating these expected potential outcomes separately permits 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 $\pi(x) = \Pr(Z = \bar{z} | X = x)$ and let $\hat{\pi}_i$ be a consistent estimator of $\pi_i = P(Z = \bar{z} | X = x_i)$, then the propensity score based matching estimator

$$\overbrace{E[Y^1 - Y^0 | T = c]} = \frac{\sum_{i:z_i=\bar{z}} (y_i - \hat{m}_{\underline{z}}(\hat{\pi}_i)) - \sum_{i:z_i=\underline{z}} (y_i - \hat{m}_{\bar{z}}(\hat{\pi}_i))}{\sum_{i:z_i=\bar{z}} (d_i - \hat{\mu}_{\underline{z}}(\hat{\pi}_i)) - \sum_{i:z_i=\underline{z}} (d_i - \hat{\mu}_{\bar{z}}(\hat{\pi}_i))}, \quad (8)$$

where $m_z(\rho) = E[Y | \pi(X) = \rho, Z = z]$ and $\mu_z(\rho) = E[D | \pi(X) = \rho, Z = z]$, is a consistent estimator of LATE, as shown in Frölich (2006a). Compared to (6) 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 $E[Y^1 | T = c]$ and $E[Y^0 | T = c]$ and of $E[Y^1 | D = 1, T = c]$ and $E[Y^0 | D = 1, T = c]$ can be obtained.

Besides the potential outcomes for the compliers, we identify the fractions of compliers, always-participants and never-participants 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:

$$\Pr(T = c) = \int (E[D | X, Z = \bar{z}] - E[D | X, Z = \underline{z}]) dF_X,$$

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

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

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

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

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

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

3.4 Matching and CIV

The CIV estimator is to matching estimators (Heckman, Ichimura and Todd 1998, Imbens 2004): It a matching estimator for the subpopulation of compliers. Matching estimators are valid in a selection on observables framework, where the key assumption is known as ignorable treatment assignment (Rosenbaum and Rubin 1983) or conditional independence assumption (CIA, Lechner 1999):

$$Y^1, Y^0 \perp D | X, \quad (9)$$

where $A \perp B | C$ denotes (mean) independence of A and B conditional on C. Therefore, we obtain $E[Y^1 | D, X] = E[Y^1 | X]$ and $E[Y^0 | D, X] = E[Y^0 | X]$. Under this assumption the average treatment effect (ATE) and the average treatment effect on the treated (ATET) are identified as

$$E[Y^1 - Y^0] = \int (E[Y | X, D = 1] - E[Y | X, D = 0]) dF_X, \quad (10)$$

$$E[Y^1 - Y^0 | D = 1] = E[Y | D = 1] - \int E[Y | X, D = 0] dF_{X|D=1}, \quad (11)$$

provided there is common support. These average effects can be estimated in every labour market, assuming that the X variables capture all heterogeneity that affects treatment choice and potential outcomes.

To better understand the relationship between CIV and matching estimation, let us assume selection on observables in the complier population:

$$Y^1, Y^0 \perp D | X, T = c. \quad \text{CIV.4'}$$

Developing a matching estimator for the complier treatment effect analogous to (10) and (11) would give

$$E[Y^1 - Y^0 | T = c] = \int (E[Y | X, D = 1, T = c] - E[Y | X, D = 0, T = c]) dF_{X|T=c}. \quad (12)$$

In contrast to ATE and ATET, however, this is not directly identified, because the type (a, n, c) is unobserved. Nevertheless, Appendix A shows this expression to be equivalent to (5). Therefore, the CIV estimator is a matching estimator for the compliers. It thus requires conditional independence for the compliers, whereas identification of the ATE requires conditional independence to hold in the entire population.

In fact, assumptions CIV.4' and CIV.4 for the compliers are equivalent, given the direct correspondence between D and Z for the compliers. Hence, CIV.4' can be thought of as the core assumption. The CIV estimator requires additionally the assumptions CIV.1, CIV.2, CIV.3, CIV.5 and CIV.4 for the never- and always-participants, whereas, loosely speaking, matching estimators of ATE and ATET require the conditional independence to hold for always- and never-participants as well (and for defiers, if they exist). The CIA for the compliers may often be more plausible than for the always- and never-participants, as they are at the margin of changing participation status, i.e. their participation status may be more or less random.

Given the richness of the data available in this application, we may be willing to assume conditional independence also for the always- and never-participants. If, in addition to the CIV assumptions, we assume

$$Y^0 \perp D | X, \quad (13)$$

the expected potential outcome $E[Y^0 | T = a]$ is identified. By noting that

$$E[Y^0] = E[Y^0 | T = a]P(T = a) + E[Y^0 | T = c]P(T = c) + E[Y^0 | T = n]P(T = n),$$

it follows that

$$E[Y^0 | T = a] = \frac{E[Y^0] - E[Y^0 | T = c]P(T = c) - E[Y^0 | T = n]P(T = n)}{P(T = a)}. \quad (14)$$

A similar decomposition of $E[Y^0 | D = 1]$ leads to:

$$E[Y^0 | T = a] = \frac{E[Y^0 | D = 1] - E[Y^0 | T = c, D = 1]P(T = c | D = 1)}{P(T = a | D = 1)}. \quad (15)$$

Since all terms on the right-hand side are identified, also noting that $P(a | D = 1) = P(a) / P(D = 1)$ by Bayes' formula, the outcome $E[Y^0 | T = a]$ can be estimated identified. With the corresponding outcome $E[Y^1 | T = a]$ already having been identified in Section 3.3, the treatment effect for the always-participants can be estimated. To identify the effect for the never-participants, we assume that

$$Y^1 \perp D | X, \quad (16)$$

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

$$E[Y^1 | T = n] = \frac{E[Y^1] - E[Y^1 | T = c]P(T = c) - E[Y^1 | T = a]P(T = a)}{P(T = n)}, \quad (17)$$

$$E[Y^1 | T = n] = \frac{E[Y^1 | D = 0] - E[Y^1 | T = c, D = 0]P(T = c | D = 0)}{P(T = n | D = 0)}. \quad (18)$$

Thus, the effect for the never-participants is identified as well.

4 Implementation of the evaluation of Swiss active labour market policies

4.1 Data and sample selection

The population for the microeconomic evaluation are all individuals who were unemployed on January 1, 1998, for at most one year. Persons who were unemployed for more than one year are excluded because they entered in unemployment before the reform was enforced in January 1997 and were thus subject to different rules and regulations at the entry in unemployment. For a random sample of 81'399 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 are legal restrictions to their mobility. In addition, persons with very low earnings (monthly earnings in last job below 1000 CHF, \approx 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 (details in Table IB.1 in the internet appendix). 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

To see the dynamics of the effects, we follow the individual labour market situation over 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 the year 1999, divided by 12. (Employment in a subsidized job, e.g. temporary wage subsidies, is not counted as regular employment.) *Earnings* is defined as total earnings from employment or self-employment during 1999, 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 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 last rows contain eight additional control variables which we include when estimating ATE and ATET. 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.

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 foreign languages (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, %)	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
Additional control variables (used later with the matching estimators)				
Employability rating:				
unknown	4	5	3	3
does not need assistance	5	7	2	2
good	16	16	17	17
intermediate	56	55	57	56
Participated in employment programme in 1997	9	3	7	3
in a temporary wage subsidy in 1997	31	19	31	19
in training in 1997	4	3	5	5
Treatment started on 1.1.1998	9	0	9	0

Note: 1 Swiss Franc (CHF) \approx 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. Participations 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 REO* that satisfies the following criteria: 1) The REO are spread over 2 cantons; 2) commuting times by car between these REO are 30 minutes or shorter; 3) the same language (French, German or Italian) is spoken in the areas belonging to the REO; 4) the ALMP composition is similar in the REO. With the first criterion, we identify local labour markets pair-wise between cantons. 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, since in Switzerland about 50% of the working population commute less than 15 minutes and about 80% commute 30 minutes or less (one way; Bundesamt für Statistik, 2003). 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 where to search for jobs. According to the criteria one to three, 30 local labour markets are identified. See the internet appendix for more details.

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 some 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. Again, the details are relegated to the internet appendix.

Table 4.2: Local labour markets divided by administrative border

#	Cantons	Regional employment offices (REO)		Number of observations		% Treated		Complier ^b	Diff. in instrument Table 2.1 ^c
				N ₁	N ₂	(7)	(8)		
(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, Münchenstein, Binningen	Basel (3x)	934	2'081	67	53	14.2	2.1
15	LU-NWOW	Luzern, Emmen, Emmenbrücke, Kriens	Hergiswil (2x)	1'607	265	64	62	2.4	-10.4
16	LU-ZG	Luzern, Emmen, Emmenbrücke, 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.

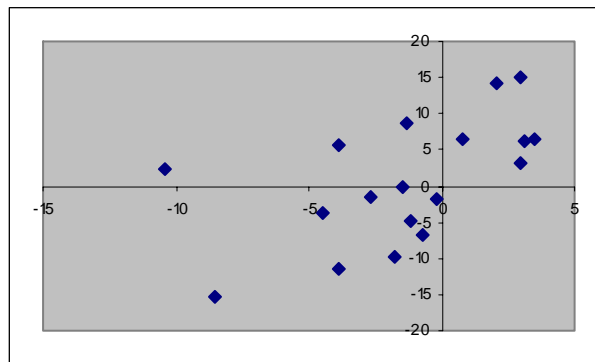
^b The estimate of the fraction of compliers is the difference between the previous two columns.

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

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 REO belonging to this labour market (on the two sides of the border). Columns (5) and (6) give the number of observations in the sample.

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) of Table 4.2 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 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.

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).

4.4 Descriptive evidence relating to the plausibility of the identifying assumptions

Having defined the various local labour markets, we examine whether there are differences in the individual characteristics X within a labour market. If the distribution of X seems to be balanced on the two sides of the border, there might not be any need to control for covariates. On the other hand, if differences in X are related to the quota per unemployed Z , we might want to control for these characteristics in the CIV estimator. (Nevertheless, in Section 5 we will see that the estimates with and without X turn out to be similar.)

The Table 4.3 provides evidence for the four largest relevant labour markets, where size is defined in terms of the number of compliers (i.e. estimated complier fraction multiplied with the number of observations). This table gives the estimated coefficients of a probit regression of the quota Z (which takes only two different values within each labour market) on the characteristics X . They will be the basis of the "propensity

score" based CIV estimator as implemented in the next section. The ease the reading of the table, only those coefficients significant at the 5% level are shown. From the table it is obvious that the distribution of X is not perfectly balanced within the labour markets. However, since there is only one variable that is significant in all four comparisons (preferred future job is the same as the last job), it is hard to see any systematic differences that question our basic IV assumptions.

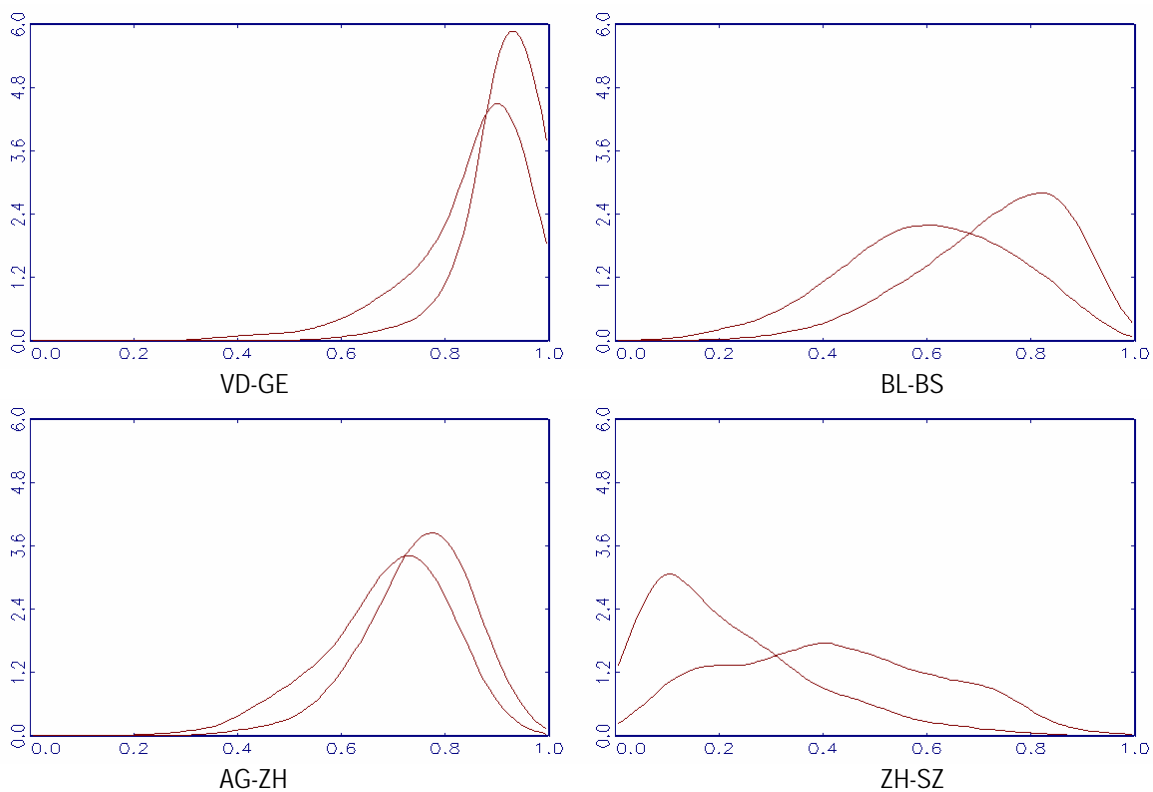
Table 4.3: Probit estimates for the four largest labour markets – significant coefficient only

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

Note: 59 regressors plus a constant. For the list of all control variables, see Table 4.1. Asymptotic t-statistics are in brackets.

Figure 4.2 summarizes the association between the characteristics X and the quota within the four largest labour markets by showing the distribution of the estimated "propensity scores" $\pi(x) = \Pr(Z = \bar{z} | X = x)$ on the two sides of the border. These graphs summarize the dissimilarity of the characteristics across the border in a single index. (The graphs for all 18 labour markets are given in the internet appendix.) These figures indicate that there do not seem to be large differences in the characteristics and that there is substantial overlap. We therefore expect the estimates with and without control variables to not be very different.

Figure 4.2: Distribution of the estimated instrument propensity scores in the four largest labour markets



Note: Kernel density plot of distributions of estimated propensity scores $\pi(x) = \Pr(Z = \bar{z} | X = x)$ for the four largest local labour markets conditional on \bar{z} and \underline{z} , bandwidth 0.1.

5 The effects of active labour market policies in Switzerland

In this section, we first exploit the cantonal quota as an instrumental variable to overcome selection on unobservables and identify local average treatment effects. Furthermore, we will then use the richness of the data to impose a selection on observables assumption to identify average effects and average effects on the treated. Finally, we combine both assumptions to extract a maximum amount of information about the ef-

fectiveness of active labour market policies from the data and identify average effects for those groups of unemployed whose participation decision is not influenced by the observed variation in the instrument.

All estimates are based on those 32634 individuals who live in the 18 labour markets defined in the previous section. The same sample is also used for the matching estimator to ease the comparison of the results for the different estimators. We estimate the effects for the two outcome variables, employment (in months, divided by 12, *employ*) and monthly gross earnings (in CHF, *earn*). We start with a parametric analysis and subsequently relax the assumptions to use nonparametric IV and matching.

5.1 Parametric specifications

Table 5.1 shows the results of the parametric specifications defined in Section 3. Columns (1) to (4) contain the results of OLS. Columns (5) to (8) contain the results of 2SLS and columns (9) to (12) the results of the Fuller estimator. In all cases, inference is based on the asymptotic distribution. (Bootstrapped t-values are very similar to the asymptotic ones). For the IV estimators, the lower part of Table 5.1 also contains the first stage regression (identical for both IV estimators). A variety of test statistics examined did not indicate any problems of weak instruments in these parametric settings. Since functional forms matter for the parametric specifications, we consider log earnings as an additional outcome variable.

First, we compare the different estimators across specifications. We find that for all models, OLS and the IV estimators are drastically different. OLS suggests a small negative effect of ALMP, while the IV estimators suggest a positive effect, although the latter is too large to be credible. Comparing the 2SLS and Fuller estimators, we do not find substantial estimator-specific differences.¹¹ (The reasons for the negative OLS estimates are discussed in Section 5.4.)

¹¹ Nevertheless, the inference of Fuller (1) might be more reliable, see e.g. Hahn and Hausman (2003) or Hahn, Hausman, and Kürsteiner (2004).

Table 5.1: Parametric estimates of the effects of ALMP

<i>2nd stage</i>		OLS				2SLS				Fuller (1)			
<i>Effect on ...</i>		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
<i>Employ</i>		-0.03**	-0.03**	-0.02**	-0.02**	0.55**	0.45**	0.39**	0.49**	0.54**	0.44**	0.38**	0.49**
<i>Earn</i>		-197**	-200**	-198**	-190**	1517**	2635**	2011**	1598**	1509**	2555**	1944**	1588**
<i>Log earn</i>		-0.03**	-0.03**	-0.02**	-0.02**	0.697**	0.89**	0.74**	0.64**	0.69**	0.86**	0.72**	0.64**
<i>1st stage</i>													
Quota per capita	--	--	--	--	--	0.016*	0.017*	0.017*	0.015*	0.016*	0.017*	0.017*	0.015*
t-value of instrument	--	--	--	--	--	14.5	5.8	5.6	13.3	14.5	5.8	5.6	13.2
Partial R ²						0.0064	0.0010	0.0010	0.0054	0.0064	0.0010	0.0010	0.0054
Partial F-stat						210	34	31	175	210	34	32	175
Total R ²						0.0064	0.0137	0.0482	0.0412	0.0064	0.0137	0.0482	0.0412
F-stat						210	21	21	23	210	22	21	23
<i>Specification</i>													
Labour market fixed effects α_m	no	yes	yes	no	no	no	yes	yes	no	no	yes	yes	no
Individual controls (X)	no	no	yes	yes	no	no	no	yes	yes	no	no	yes	yes

Note: **, * indicates significance at the 5% and 10% level, respectively. Significance at the 1% level and 0.1% level is indicated in **bold face** and **bold face in italics**. Inference is based on the asymptotic distribution. 32634 observations (37401 according to Table 4.2 minus 4767 that are counted twice because they appear in two markets at the same time). The regressions with individual controls contain 59 characteristics. The regressions with labour market fixed effects contain 18 local labour market dummies as well as three labour office dummies for those observations where local labour markets overlap (i.e. which are in two local labour markets). In the IV regressions, D is instrumented by quota p.c.

Second, with respect to models, it appears that the OLS coefficient estimates and their precision do not change much when moving from the most restrictive specification (column 1) to the most flexible specification with labour market fixed effects and individual controls (column 3). A similar feature holds for the IV estimators. Although, making the specification more flexible reduces the employment effects as expected (this is consistent with our previous expectation that the simple 2SLS estimates would be upward biased due to the relationship with the past unemployment rate), it increases the earnings effects, which appears strange. Not surprisingly, although all effects remain significant, their precision is drastically reduced (as are the partial F-statistics of the first stage), when labour market fixed effects are allowed for, suggesting that they absorb much of the variation in the instrument.

To relax the parametric assumptions and allow for full flexibility with respect to treatment heterogeneity and the relation of the potential outcomes to the covariates, we examine nonparametric estimators in the following sections.

5.2 Implementation of the nonparametric estimators

5.2.1 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 (4) 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 in a previous version of this paper where the Wald estimator often seemed to have an infinite variance. In that Monte Carlo study, we examined several alternative estimators and found the Fuller (1977) estimator to perform best. This is also in line with the evidence in Hahn and Hausman (2003) and Hahn, Hausman and Kürsteiner (2004), who cautioned against the use of estimators without finite sample moments. The Fuller estimator is a k -class estimator (Theil, 1958 and Nagar, 1959) and has finite sample moments even without overidentification. Further details on this estimator are given in Appendix A.2.

5.2.2 Estimates conditional on covariates

The previous unconditional IV estimator relies on unconfoundedness of the instrument. However, since the populations living on the two sides of the border seem to differ in their characteristics as witnessed in Section 4 (although the differences do not seem to be very large), we might want to incorporate these differences in the IV estimator as discussed in Section 3. 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, 2006a): (i) The probability $\pi(x) = P(Z = \bar{z} | X = x)$ 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

$$E[D | \hat{\pi}(X) = \rho, Z = \underline{z}], \quad E[D | \hat{\pi}(X) = \rho, Z = \bar{z}], \quad E[YD | \hat{\pi}(X) = \rho, Z = \underline{z}],$$

$E[YD | \hat{\pi}(X) = \rho, Z = \bar{z}]$, $E[Y(1-D) | \hat{\pi}(X) = \rho, Z = \underline{z}]$ and $E[Y(1-D) | \hat{\pi}(X) = \rho, Z = \bar{z}]$. Bandwidths are chosen from the expanding grid with 10 values: $\{1/100, 1.9/100, 1.9^2/100, \dots, 1.9^8/100, \infty\}$. (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.¹² Given a sample of N observations on $(x_i, y_i) \in \mathfrak{R} \times \mathfrak{R}$ and a bandwidth value h , the ridge regression estimate at location x is defined as

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

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

This procedure estimates $E[Y^1 | T = c]$ and $E[Y^0 | T = c]$ 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.

In Section 5.4, matching estimators are used to estimate the average treatment effect (ATE) and the average treatment effect on the treated (ATET). For ease of comparison, we choose an implementation of these estimators that is very similar to the CIV estimator: they are estimated separately for each local labour market (exact match on labour market). The matching estimator of ATE is identical to the numerator of the CIV estimator with Z replaced by D throughout. The propensity score version of this matching estimator now refers to the conditional probability of D , i.e. the propensity score $p(x) = P(D = 1 | X = x)$ is the probability of participating among all individuals within the same labour market. The same set of control variables X is used for the CIV and for the ATE/ATET matching estimators. With an estimate of $p(x)$, the ATE and ATET are estimated as

¹² Conventional local linear regression estimators often perform poorly. For curve estimation, this is observed by

$$\widehat{E[Y^1 - Y^0]} = \sum_{i:d_i=1} (y_i - \hat{m}_0(\hat{p}_i)) - \sum_{i:d_i=0} (y_i - \hat{m}_1(\hat{p}_i)), \quad (19)$$

$$\widehat{E[Y^1 - Y^0 | D = 1]} = \sum_{i:d_i=1} (y_i - \hat{m}_0(\hat{p}_i)), \quad (20)$$

where $m_d(\rho) = E[Y | p(X) = \rho, D = d]$. Estimation of the conditional expectation function $m_d(\rho)$ is analogous to the CIV estimator.

5.3 Results of the instrumental variable estimators

Table 5.2 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.¹³ Let $\hat{\Theta} = w'\hat{\theta}$, where $\hat{\theta}$ is the vector of the 18 local

Seifert and Gasser (1996) and for matching estimation by Frölich (2004).

¹³ To examine treatment effect heterogeneity we also regressed the 18 different estimates on specific characteristics of the local labour markets. For the composition of ALMP, for example, the regressions did not provide precise results, and we leave a further analysis for future research.

estimates and w a vector of weights. The first row of Table 5.2 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 compliers) 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. However, the Θ_{MD} weights are difficult to interpret in our application, 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 5.2, inference is based on the nonparametric bootstrap. Significance levels rely on the percentiles of the estimates (9999 bootstrap replications), because the finite sample moments do not exist for the unrestricted 2SLS estimator or for earnings with the restricted 2SLS estimator or the CIV estimator.¹⁴

¹⁴ The bootstrap proceeded by drawing with replacement from the original sample with the 66713 observations and repeating the entire estimation process. The probits are estimated by maximum likelihood augmented with the following features to deal with collinearity problems that might occur during the bootstrapping. 1) All regressors without variation are dropped. 2) All regressors that cause local multicollinearity are dropped. For detecting (nearly) linear dependencies in the regressor matrix, the pivotal orthogonal-triangular (QR) decomposition is used, see Judd (1998, p. 58f) or Press, Flannery, Teukolsky, and Vetterling (1986, p. 357ff). This decomposition decomposes a regressor or moment matrix into an orthogonal matrix Q and an upper 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. Additionally,

Table 5.2: 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)
	<i>Employ</i>	<i>Earn</i>	<i>Employ</i>	<i>Earn</i>	$w_c * 100$	<i>Employ</i>	<i>Earn</i>	<i>Employ</i>	<i>Earn</i>	$w_c * 100$
Aggregated effects										
Θ_c	0.18**	827	0.18**	906	-	0.15**	589	0.16**	601	-
$\Theta_{c,trim}$	0.21**	313	0.16**	394	-	0.13**	141	0.12*	371	-
Θ_{obs}	0.21**	-1361	0.19**	-424	-	0.02	202	-0.03	3015	-
$\Theta_{obs,trim}$	0.23**	-376	0.19**	-1460	-	0.11	-65	0.13	441	-
Θ_{MD}	0.13	516	0.06	627	-	0.06	343	0.09*	-	-
Effects for individual labour markets										
2	-0.74	-1058	-0.58	-861	4.5	-1.00	-3328	-1.00	-3481	3.9
7	-0.52	-228	-0.86	-2054	1.8	-1.00	-2308	-1.00	-2604	1.6
8	0.07	3681	0.15	1300	4.9	-0.82	-334	-1.00	0	2.4
9	0.47	1171	0.55	2660	1.6	0.46	2586	0.47	2988	1.3
10	0.00	4591	0.19	1306	0.7	0.27	3203	0.44	7556	1.0
11	0.74	2907	0.77	2953	19.3	0.86	4392	0.91	4640	16.8
12	0.40	2037	0.74	1686	5.1	0.46	153	0.52	202	6.0
13	-0.10	591	-0.10	747	17.9	0.03	1711	0.04	1749	17.7
15	-1.00	-0282	-1.00	-4991	1.4	-1.00	-6228	-1.00	-12605	1.8
16	1.00	-27019	1.00	-11648	0.3	-0.14	506	-1.00	53205	0.2
17	-0.37	-1441	-0.23	356	2.5	-0.26	-2268	-0.22	-2786	1.8
19	0.21	1158	0.14	864	14.5	0.07	-350	0.08	-357	15.4
21	0.48	606	0.47	849	6.3	0.32	193	0.34	219	7.1
22	0.92	-692	0.09	2729	1.7	-0.10	-4736	-0.15	-6523	2.8
23	0.12	149	-0.03	294	11.7	0.04	-1114	0.05	-1137	12.2
24	0.23	-419	0.16	-283	2.7	0.57	2523	0.66	3032	2.9
25	0.00	-12231	0.00	-14622	0.4	-0.13	366	0.00	0	0.8
28	-0.20	-920	-0.24	-899	2.7	0.10	871	0.11	941	4.2

Note: In the first three 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 5% and 10% level, respectively. Significance at the 1% level is indicated in **bold face**. Inference is based on the bootstrap, for details see the previous footnote. 59 regressors. 32634 observations used.

Our most trusted results, the rather precisely estimated *compliers-weighted* LATE Θ_c , suggest 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. This is quite a large effect, which is however much more reasonable than the effect previously estimated with the parametric IV specifications. For the IV estimation without covariates, this result appears to be stable across estimators. For the

for the Fuller estimator and 2SLS (without covariates) values are bootstrapped, as this is likely to lead to more precise inference arising from asymptotic refinements. A nested bootstrap with 999 times 999 replications is used, see Hall (1986), Beran (1987), Loh (1987) and Cameron and Trivedi (2005, p. 374). Because of an excessive demand on computation time, this is not possible for the CIV estimator where inference is based on 499 replications only.

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. However, its precision is too low to exclude a zero effect. 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.

To examine the sensitivity of the results, we checked a variety of alternative specifications: First, a finer bandwidth grid with 30 values $\{1/100, 1.2/100, 1.2^2/100, \dots, 1.2^{28}/100, \infty\}$, 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 huge 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 (see previous footnote). 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. Fourth, to judge the relevance of imposing the range restrictions, estimates without such restrictions were computed. Without these restrictions, the estimated employment effects may be outside of the interval $[0, 1]$ and the earnings effects are more variable. Overall, the Fuller estimates are similar to those of Table 5.2, but the 2SLS estimates are strongly affected by one outlier.¹⁵

Based on the CIV estimator, Table 5.3 shows the estimated proportions of always-participants, complier and never-participants in the 18 local labour markets as well as their expected potential outcomes.¹⁶ The expected potential outcomes for the compliers vary considerably between labour markets, while they are more stable for the always- and never-participants. Note that whenever the estimated outcomes for the compliers are restricted to the boundary of their range, the number of compliers is very low (below 120 for

¹⁵ For details, see the internet appendix.

¹⁶ Due to space constraints, the detailed results for all 18 labour markets are relegated to the internet appendix.

earnings and below 40 for employment). The non-treatment outcomes for the never-participants $E[Y^0 | T = n]$ are usually higher than the treatment outcomes $E[Y^1 | T = a]$ for the always-participants. Since IV does not identify the treatment effects for these two groups, we cannot exclude the possibility of a very negative treatment effect for both groups. In light of the previous evidence and of the results of the following section, this does not appear likely, though. Rather, it might indicate differences in the average employability of these two groups. The never-participants are least likely to participate in ALMP as caseworkers may expect them to find a job on their own, whereas the always-participants are those who are considered most in need of assistance.

Table 5.3: Aggregated estimated outcomes of always- and never-participants (CIV estimator)

<i>Employment</i>				<i>Earnings</i>			
$E(Y^1 T=c)$	$E(Y^0 T=c)$	$E(Y^1 T=a)$	$E(Y^0 T=n)$	$E(Y^1 T=c)$	$E(Y^0 T=c)$	$E(Y^1 T=a)$	$E(Y^0 T=n)$
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.

5.4 Results of the matching estimators

The previous results showed the effects of participating in ALMP anytime during the year 1998 for the compliers based on the CIV assumptions. As discussed in Section 3, it is interesting to compare these results with the results from matching estimation, in particular since such a comparison allows to distinguish between treatment effects on compliers, always-participants and never-participants.

For applying the matching estimator, we need the CIA to hold in the entire population within each labour market. Therefore, we need to extend the set of control variables and the definition of the treatment. The additional control variables (see the last eight rows of Table 4.1) capture the caseworkers' assessment of the employability of the unemployed person, which also reflects personality and character traits and social skills. In addition, previous treatment participation is controlled for. This may be relevant, because repeated participation in ALMP may have different effects than the first ALMP received. Furthermore, the fact that somebody participated in ALMP, but is still (or again) unemployed may reflect difficulties in finding or keeping a job. With these additional regressors, the set of control variables is similar to those of Gerfin and

Lechner (2002) and Gerfin, Lechner and Steiger (2005) who argued that the conditional independence assumption is plausible.¹⁷

Apart from adding some more control variables for plausibility of the CIA, it also appears necessary to change the definition of treatment. So far, a person was considered treated if being assigned to ALMP any time during the year 1998. The quota increased the treatment probability and the timing of treatment may have depended not only on the date of unemployment registration, but as well on local waiting times and on other seasonality factors. The observations compared were those with different values of the quota Z . When comparing treated and non-treated in the matching estimators of ATE and ATET, the definition of D is not innocuous anymore. If one considers treatment assignment D and finding a job as two competing processes, the group of non-treated contains disproportionately many observations who found a job before treatment started, whereas the group of treated contains more individuals who would have found a job rather late and treatment happened to start before it. Hence, individuals with good labour market chances (which may be partly unobservable) are over-represented in the $D=0$ group and under-represented in the $D=1$ group. In our population everybody is unemployed on January 1, 1998. Therefore, assignment on that day would not be subject to that problem. Of course, defining as treated only those who started exactly on a specific day would lead to extremely small sample sizes. A treatment window of two or three months is a reasonable trade-off between sample sizes and bias.¹⁸ Thus, a person is defined as treated if assigned to ALMP during January to March 1998, otherwise this person is considered as untreated. For robustness, we examine also the period January to February 1998 as the treatment window (see the internet appendix).

¹⁷ We did not want to include these additional regressors in the previous IV estimations as they might be partly affected by the quota and may thus be masking a part of the effect. Employment offices that struggled to fill their quota were likely to rate more of their unemployed as being in need of additional services, and may have resorted to ALMP more frequently in the previous year (as the quota is highly correlated over time). Hence, conditioning on these variables may eliminate part of the impact that the quota had on the participation probabilities and may therefore reduce the magnitude of the CIV estimates. *Not conditioning* on these variables, however, is likely to lead to biased estimates of ATE and ATET as the conditional independence assumption may not be plausible.

Table 5.4 shows the matching estimates of ATE and ATET for the specification based on the January to March definition of participation. The estimated employment and earnings effects are positive, albeit small, and precisely estimated. It is noteworthy that compared to IV estimation, the results for the particular labour markets are much more precise and not close to the boundaries of the supports of the effects.

Table 5.4: Aggregated matching estimates of ATE and ATET based on population weights

OLS (with controls; coefficient of D)				Matching				
Labour market fixed effects		No labour market fix. effects		ATE: $E[Y^1 - Y^0]$		ATET: $E[Y^1 - Y^0 D=1]$		
<i>Employ</i>	<i>Earn</i>	<i>Employ</i>	<i>Earn</i>	<i>Employ</i>	<i>Earn</i>	<i>Employ</i>	<i>Earn</i>	
<i>0.021**</i>	<i>72**</i>	<i>0.025**</i>	<i>80**</i>	Θ_c	0.020**	49	0.016**	44
				Θ_{obs}	0.022**	73**	0.017**	50

Note: An unemployed is defined as treated if entering in ALMP between January to March 1998. **, * indicates significance at the 5% and 10% level, respectively. Significance at the 1% level is indicated in **bold face**. Inference is obtained from bootstrapping the estimate (percentile method). The number of replications is 499. 67 regressors. 32634 observations.

The OLS estimates are close to the matching estimates for ATE, which are somewhat larger than for ATET. The OLS coefficients are small as are those of Table 5.1, but the sign of the effects changed as could have been expected from the previous discussion about the appropriate time window for defining treatment.

5.5 Combining instrumental variable and matching estimators

The next step is to combine the knowledge obtained from matching and CIV estimation. Before we can do that, we should base the CIV estimator on the same footing as the matching estimator, i.e. to define treatment in the same way and use the same control variables. As discussed further in the internet appendix, including the additional control variables reduces the employment effects by about one third, as we had expected if the caseworkers' stated employability ratings were partly affected by the quota. The results are nevertheless robust to the exact definition of the treatment window.

Combining the results of the CIV and the matching estimator, all potential outcomes are identified for the compliers, always- and never-participants, as shown in Section 3. Table 5.5 gives the subpopulation-weighted averages over all 18 labour markets for the mean potential outcomes and effects. A first finding is

¹⁸ Alternative ways to proceed are the random start date approach (as in Lechner, 1999) or a dynamic treatment model (as in Lechner and Miquel, 2005). However, these options are not used here to stay as closely as possible to the specification of the previous section. The same holds true for the choice of estimator.

that the potential outcomes do not seem to differ between always- and never-participants. This could also be related to the re-definition of treatment, compared to Section 5.3. Before, a never-participant was a person who would not take part in ALMP during the entire year 1998, whereas it is now a person who would not participate in ALMP in the first months of 1998 but perhaps later. The treatment effects for these two subpopulations are small and barely significantly different from zero. The effects may be slightly larger for the always- than for the never-participants, but are too imprecise to confirm this suspicion.

Table 5.5: Aggregated estimates of mean potential outcomes and effects based on population weights

	Always-participants			Complier				Never-participants		
	$E[Y^1 \cdot]$	$E[Y^0 \cdot]$	$E[\Delta \cdot]$	$E[Y^1 \cdot]$	$E[Y^0 \cdot]$	$E[\Delta \cdot]$	$E[\Delta \cdot, D=1]$	$E[Y^1 \cdot]$	$E[Y^0 \cdot]$	$E[\Delta \cdot]$
<i>Employ</i>	0.64	0.62	0.020	0.68	0.56	0.12**	0.16**	0.63	0.62	0.014*
<i>Earnings</i>	2419	2318	100	2788	2731	57	266	2281	2245	55

Note: Δ refers to the treatment effect $Y^1 - Y^0$. All results re-estimated with the specification used for matching (Table 5.4). An unemployed is defined as treated if entering in ALMP between January to March 1998.

** , * indicates significance at the 5% and 10% level, respectively. Significance at the 1% level is indicated in **bold face**.

The group of compliers seems to benefit more from treatment than the always- and never-participants.¹⁹ In particular, their employment outcome Y^0 is lower than for these other two groups. This suggests that always- and never-participants consist of unemployed who have better chances on the labour market, or accept more job offers, than the compliers. Therefore, the effect of the treatment is small for them and much larger for the compliers. Interestingly, the Y^0 outcomes for the compliers suggest that they may not be as successful in finding jobs, but if they find them, they are better paying than those for the two other groups. Alternatively, they may have higher reservation wages that led them rejecting (or not receiving) job offers that the two other groups accepted.

From the perspective of allocating the 'right' unemployed persons to the programmes, this finding suggests that sending the always-participants to the programmes is not really a good idea. These findings are con-

¹⁹ The results for earnings are too imprecise, but for employment, the effect is significantly different. A two-sided bootstrap test of the equality of the effect for compliers and for always-participants rejected at the 10% level, as did a test on equality of the effect for compliers and never-participants.

firmed by rough cost benefit calculations that suggest that for the compliers the programmes are probably cost effective, whereas for the other groups they are not.²⁰

6 Conclusions

In this paper, we estimated the treatment effects of participation in active labour market policies (ALMP) on employment prospects and earnings using parametric and nonparametric IV methods as well as matching estimators. The instrument is based on a mandated quota, which stipulated the minimum number of places a canton had to provide. This quota introduced an exogenous variation in the likelihood of being assigned to an ALMP. Since the quota differs even within homogenous labour markets, it is used as a local instrument within special, precisely defined local labour markets. To counter concerns about potential confoundedness of this instrument, a large number of observed individual characteristics were controlled for using parametric methods as well as nonparametric propensity-score conditional IV estimation. The results were robust to the inclusion of these covariates. Thus, confoundedness is 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 to summarise the information in the local estimates in a concise way. The estimation results from this nonparametric IV strategy appeared more credible than the results obtained from a standard parametric IV estimation with control variables and local labour market fixed effects. Finally, we showed that by combining conditional IV and matching estimation, we identify effects for the non-complying unemployed, i.e. for whom the differences in the quota do not induce them to participate (never-participants) or to not participate (always-participants).

Our most trusted results suggest that employment increased by about *two months* per year due to participation in ALMP. Given rough cost-benefit considerations, such an increase would be sufficient to make programme participation worthwhile. For the IV specification without covariates, this result is stable across

²⁰ See the internet appendix for more details on these calculations.

estimators. For the conditional IV specification, it is not sensitive to the distribution of the covariates between participating and non-participating compliers. The earnings effect shares the same features as the employment effect, but is too imprecise to exclude a zero effect. Interestingly, the employment effects for the compliers are larger than the treatment effects for the always- and the never-participants.

The size of the effects are roughly in line with previous research for Switzerland in 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 programmes. 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.

The more surprising feature is that the effects for the compliers appear to be larger than for the always-participants. A simple argument would postulate that the treatment effect for this marginal group should be smaller than the treatment effect on the treated if utility maximising individuals decide about participation or if caseworkers act on their behalf. In the case of Switzerland, this need not be the case. With the activation principle introduced by the reform, the caseworker can push an unemployed person into a labour market programme. Lechner and Smith (2006) analysed this assignment process and found that caseworkers were not very successful in assigning unemployed to their most beneficial programmes. The marginal group of compliers might therefore have benefited more from labour market programmes than others (at least in the short term as measured in this paper, when the economy was booming), and the local average treatment effect thus could have been higher than the treatment effect on the treated. Summarising this discussion succinctly, the Swiss active labour market programmes seem to have been effective for at least a part of the population, while they may not have been so for the population as a whole. This individual treatment effect heterogeneity may support the need for a better targeting of active labour market programmes.

References

- Abbring, J.H., and G. van den Berg (2003): "The Nonparametric Identification of Treatment Effects in Duration Models," *Econometrica*, 71, 1491-1517.
- Basmann, R.L. (1961): "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Overidentified Cases," *Journal of the American Statistical Association*, 56, 619-636.
- Basmann, R.L. (1963): "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in a Leading Three-Equation Case," *Journal of the American Statistical Association*, 58, 161-171.
- 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.
- Bundesamt für Statistik (2003): *Volkszählung 2000, Pendelverkehr - Neue Definition der Agglomerationen*, Pressemitteilung vom 15.5.2003, Neuchatel.
- Buse, A. (1992): "The Bias of Instrumental Variable Estimators," *Econometrica*, 60, 173-180.
- 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.
- Donald, S.G., and W.K. Newey (2001): "Choosing the Number of Instruments," *Econometrica*, 69, 1161-1191.
- Fay, R. G. (1996): "Enhancing the Effectiveness of Active Labour Market Policies: Evidence from Programme Evaluations in OECD countries," *Labour Market and Social Policy Occasional Papers*, 18, OECD.
- Frölich, M. (2004): "Finite Sample Properties of Propensity-Score Matching and Weighting Estimators," *Review of Economics and Statistics*, 86(1);, 77-90.
- Frölich, M. (2006a): "Nonparametric IV estimation of local average treatment effects with covariates," forthcoming in *Journal of Econometrics*.
- Frölich, M. (2006b): "A Note on Parametric and Nonparametric Regression in the Presence of Endogenous Control Variables," IZA Discussion paper 2126.

- Fuller, W. (1977): "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939-953.
- Gerfin, M., and M. Lechner (2002): "Microeconomic 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., and J. Hausman (2003): "Weak Instruments: Diagnosis and Cures in Empirical Econometrics," *American Economic Review, Papers and Proceedings*, 93, 118-125.
- Hahn, J., J. Hausman and G. Kürsteiner (2004): "Estimation with Weak Instruments: Accuracy of Higher-Order Bias and MSE Approximations," *Econometrics Journal*, 7, 272-306.
- 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., R. LaLonde and J. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," in: O. Ashenfelter and D. Card (eds.):, *Handbook of Labour Economics*, Vol. 3, 1865-2097, Amsterdam: North-Holland.
- Heckman, James, Hidehiko Ichimura, and Petra 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.

- Kabe, D.G. (1964): "On the Exact Distributions of the GCL Estimators in a Leading Three Equation Case," *Journal of the American Statistical Association*, 59, 881-894.
- Klueve J. (2006): "The Effectiveness of European Active Labor Market Policy," IZA DP No. 2018.
- Klueve, J., and C. Schmidt (2002): "Can training and employment subsidies combat European unemployment?," *Economic Policy*, 35, 411-448.
- Lalive, R., J. C. van Ours and J. Zweimüller (2000): "The Impact of Active Labor Market Programs on the Duration of Unemployment," IEW Working Paper No. 41, University of Zurich.
- Lechner, M. (1999): "Earnings and Employment Effects of Continuous Off-the-job Training in East Germany after Unification," *Journal of Business and Economic Statistics*, 17, 74-90.
- Lechner, M. (2001): "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption," in: M. Lechner and F. Pfeiffer (eds.): *Econometric Evaluation of Active Labour Market Policies*, 43-58, Heidelberg: Physica.
- Lechner, M., and J. A. Smith (2006): "What is the Value Added by Caseworkers?," forthcoming in *Labour Economics – An International Journal*.
- Lechner, M., R. Miquel, and C. Wunsch (2004): "Long-run effects of Public Sector Sponsored Training in West Germany," Discussion Paper 2004-19, Department of Economics, University of St. Gallen.
- Loh, W.-Y. (1987): "Calibrating confidence coefficients," *Journal of the American Statistical Association*, 82, 155-162.
- Maddala, G.S., and J. Jeong (1992): "On the Exact Small Sample Distribution of the Instrumental Variable Estimator," *Econometrica*, 60, 181-183.
- Mariano, R.S., and T. Sawa (1972): "The Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables," *Journal of the American Statistical Association*, 67, 159-163.
- Martin, J. P. and D. Grubb (2001): "What Works and for Whom: A Review of OECD Countries' experiences with active labour market policies," *Swedish Economic Policy Review*, 8(2):, 9-56.
- Nagar, A.L. (1959): "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," *Econometrica*, 27, 575-595.

- Nelson, C.R., and R. Startz (1990): "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967-976.
- Pence, K. (2003): "Foreclosing on Opportunity: State Laws and Mortgage Credit," *mimeo*, Board of Governors of the Federal Reserve System.
- Press, W., B. Flannery, S. Teukolsky, and W. Vetterling (1986): *Numerical Recipes*, Cambridge University Press, Cambridge.
- Richardson, D.H. (1968): "The Exact Distribution of a Structural Coefficient Estimator," *Journal of the American Statistical Association*, 63, 1214-11226.
- Rosenbaum, P.R. and D.B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-50.
- Sawa, T. (1969): "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators," *Journal of the American Statistical Association*, 64, 923-937.
- 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.
- Stauffer, H. (1998): *Rechtsprechung des Bundesgerichts zum Sozialversicherungsrecht, Bundesgesetz über die obligatorische Arbeitslosenversicherung und Insolvenzenschädigung*, 2nd ed. Zürich: Schulthess Polygraphischer Verlag.
- Theil, H. (1958): *Economic Forecasts and Policy*. Amsterdam: North Holland.
- Wald, A. (1940): "The Fitting of Straight Lines of Both Variables are Subject to Error," *Annals of Mathematical Statistics*, 11, 284-300.

Appendix A: Identification and estimation

A.1 Proofs of identification

Proof of (7): Notice first that in the subpopulation of compliers conditioning on $D=1$ is equivalent to $Z = \bar{z}$ and conditioning on $D=0$ is equivalent to $Z = \underline{z}$. It thus follows that

$$E[Y^1 - Y^0 | D = 1, T = c] = \int E[Y^1 - Y^0 | X, Z = \bar{z}, T = c] \cdot dF_{X|Z=\bar{z}, T=c} = \int E[Y^1 - Y^0 | X, T = c] \cdot dF_{X|Z=\bar{z}, T=c}$$

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

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

$$dF_{X|Z=\bar{z}, T=c} = \frac{\Pr(Z = \bar{z}, T = c | X) \cdot dF_X}{\Pr(Z = \bar{z}, T = c)} = \frac{\Pr(T = c | X, Z = \bar{z}) \cdot \pi(X) \cdot dF_X}{\int \Pr(Z = \bar{z}, T = c | X) dF_X} = \frac{\Pr(T = c | X) \pi(X) dF_X}{\int \Pr(T = c | X) \pi(X) dF_X}$$

by the unconfounded type assumption (CIV.3). We also notice that

$$E[D | X, Z = \bar{z}] - E[D | X, Z = \underline{z}] = \Pr(T = c | X),$$

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

see Frölich (2006a, Appendix A), to obtain

$$E[Y^1 - Y^0 | D = 1, T = c] = \frac{\int E[Y | X, Z = \bar{z}] - E[Y | X, Z = \underline{z}] \pi(X) dF_X}{\int E[D | X, Z = \bar{z}] - E[D | X, Z = \underline{z}] \pi(X) dF_X}. \quad (21)$$

Proof of equivalence between (12) and (5): Notice that for compliers conditioning on D and conditioning on Z are equivalent. Therefore, by exploiting the assumption of conditional independence for the compliers, the expression (12) can be written as

$$\begin{aligned} E[Y^1 - Y^0 | T = c] &= \int (E[Y^1 | X, D = 1, T = c] - E[Y^0 | X, D = 0, T = c]) dF_{X|T=c} \\ &= \int (E[Y^1 | X, Z = \bar{z}, T = c] - E[Y^0 | X, Z = \underline{z}, T = c]) dF_{X|T=c} = \int (E[Y^1 | X, T = c] - E[Y^0 | X, T = c]) dF_{X|T=c} \end{aligned}$$

by the exclusion restriction. The rest of the proof is identical to Frölich (2006a).

A.2 Some details on the Fuller estimator

The k -class estimators were introduced by Theil (1958) and Nagar (1959) and are of the following form:

$$\{D_N' [I_N(1-k) + kP_N] D_N\}^{-1} D_N' [I_N(1-k) + kP_N] Y_N, \quad (22)$$

where N is the sample size, Y_N is the data vector of the outcome variable, D_N is the data matrix of the endogenous variables (including a constant), I_N the identity matrix and $P_N = Z_N(Z_N'Z_N)^{-1}Z_N'$ the projection matrix of the data matrix of instruments Z_N (including a constant). The constant k defines the specific estimator: $k = 1$ gives the conventional 2SLS estimator. $k = 0$ corresponds to OLS. Choosing k as the smallest root of the determinantal equation $|\Xi' \Xi - k \Xi'(I_N - P_N) \Xi| = 0$, where $\Xi = [Y_N : D_N]$ is the horizontal concatenation of Y_N and D_N , gives the Limited Information Maximum Likelihood estimator (LIML). In the case of exact identification, LIML and 2SLS are identical. For normal errors (or errors with even fatter tails), LIML does not possess finite moments and 2SLS has moments only in the case of over-identification.²¹ Fuller (1977) proposed a modified LIML estimator with $k = k_{LIML} - \alpha / (N - L)$, where α is a positive constant and L is the number of instruments. The Fuller estimator has first and second moments in finite-samples. Choosing $\alpha = 1$ gives nearly unbiased estimates and performed best in our Monte Carlo analysis.

²¹ See Basmann (1961, 1963), Kabe (1964), Richardson (1968), Sawa (1969), Mariano and Sawa (1972), Nelson and Startz (1990), Buse (1992), Maddala and Jeong (1992) and others.