

The Matching Method for Treatment Evaluation with Selective Participation and Ineligibles

Monica Costa Dias*

Hidehiko Ichimura[†]

Gerard J. van den Berg[‡]

May 2012

*Institute for Fiscal Studies, cef.up - Faculty of Economics at the University of Porto, and IZA. Address: IFS, 7 Ridgmount Street, London WC1E 7AE, UK. monica.d@ifs.org.uk

[†]Graduate School of Economics and Graduate School of Public Policy, University of Tokyo.

[‡]Alexander von Humboldt Professor of Econometrics and Empirical Economics, University of Mannheim; IFAU-Uppsala, VU University Amsterdam, IZA, CEPR, and IFS.

Keywords: propensity score, policy evaluation, treatment effect, regression discontinuity, selection, subsidized work, youth unemployment.

Acknowledgements: We thank the Editor, an anonymous Associate Editor, two anonymous Referees, Richard Blundell, Xavier de Luna, Barbara Sianesi, and Petra Todd, for useful comments. We also thank Louise Kennerberg and Barbara Sianesi for help with the Swedish data. We gratefully acknowledge financial support from the ESRC and from IFAU-Uppsala. Costa Dias is thankful to Fundacao para a Ciencia e Tecnologia and the European Social Fund for financial support. Ichimura thanks support from the JSPS Basic Research fund.

Abstract

The matching method for treatment evaluation does not balance selective unobserved differences between treated and non-treated. We derive a simple correction term if there is an instrument that shifts the treatment probability to zero in specific cases. Policies with eligibility restrictions, where treatment is impossible if some variable exceeds a certain value, provide a natural application. In an empirical analysis, we exploit the age eligibility restriction in the Swedish Youth Practice subsidized work program for young unemployed, where compliance is imperfect among the young. Adjusting the matching estimator for selectivity changes the results towards making subsidized work detrimental in moving individuals into employment. We also consider the eligibility change induced by the introduction of the program.

1 Introduction

The matching method for treatment evaluation compares outcomes of treated and non-treated subjects, conditioning on observed individual and environment characteristics. Basically, the average treatment effect on the treated (ATT) is estimated by averaging observed outcome differences over the treated. The main assumption is that the conditioning ensures that the assigned treatment status is conditionally mean independent from the potential outcomes (this is usually known as “the Conditional Independence Assumption” or, in short, CIA, although in fact it concerns mean independence).¹

The method is intuitive, as it mimics randomized experiments: the distributions of behavioral determinants and indicators are balanced as closely as possible over treated and non-treated, using observational data. The use of the method has improved the policy evaluation practice by clarifying the importance of common support restrictions for the distribution of conditioning variables. By now, it is a common tool for the analysis of active labor market policies (ALMP) and programs (see e.g. the survey in Kluve, 2006). However, matching has the well-recognized limitation that it does not ensure the balancing of the distribution of unobservable determinants of both treatment assignment and outcomes among treated and non-treated. When incapable to balance unobservables, matching may produce biased estimates of the treatment effects.

The first contribution of this paper deals with this problem by developing an estimation method for the average treatment on the treated that is robust to violations in the conditional independence assumption justifying matching. The idea is to correct the matching estimate with a measure of the bias due to selection on unobservables. Key to the estimation of such correction term is the availability of an instrument capable of driving participation to zero at certain of its (possibly limiting) values while keeping the selection mechanism partly unexplained at other parts of its distribution. Like the matching methods, our ap-

¹See e.g. Cochrane and Rubin (1973), Rosenbaum and Rubin (1983), and Heckman, Ichimura, and Todd (1998).

proach matches the distribution of observed variables between treated and non-treated groups, thus effectively combining matching with the exogenous variation provided by an instrument to balance unobservables.

Alternative approaches in order to correct matching estimators for selection problems typically assume that the relevant unobserved variables have additive effects on the potential outcomes (see Heckman and Robb, 1985, and Andrews and Schafgans, 1998). The popular conditional difference-in-differences estimator (Heckman, Ichimura, Smith, and Todd, 1998) is also based on this. By contrast, our approach does not require additivity.

Within the same framework we also suggest a new test of the CIA. In the presence of a valid instrument, satisfying the condition introduced above, the CIA holds if and only if the correction term is zero. Thus, testing the validity of the CIA is equivalent to test the statistical significance of the correction term.

The second contribution of this paper is to show that there are important empirical applications for this method. Consider, for instance, the case of treatment evaluation in the context of ALMP for unemployed workers. It has been recognized that individual characteristics and employment history may not capture the full range of skills and motivation that explain both treatment participation and employment-related outcomes.² However, many programs use clear eligibility rules based on observed variables and often involving boundary restrictions. Also, novel programs are introduced and others are terminated, creating variation in eligibility over calendar time. Eligibility variation can be exploited to construct an instrument capable of moving subjects in and out of treatment while otherwise being unrelated to the potential outcome(s) of interest.

The ideal setting for the application of our method is created by non-mandatory programs in the presence on boundary eligibility restrictions on personal characteristics or

²For example, Card and Sullivan (1988), Gritz (1993), Bonnal, Fougère and Sérandon (1997) and Richardson and Van den Berg (2012) argue that this can be expected to play a major role in the empirical evaluation of ALMP, and their estimation results confirm this. Van den Berg, Van der Klaauw and Van Ours (2004) contain similar findings for the effect of punitive sanctions for welfare recipients.

time. In such cases, full non-participation is observed at certain values of the instrument while compliance is imperfect at other values.³ This is a relevant setting for ALMP. It is a common feature of ALMP to restrict eligibility to individuals aged above or below a certain age, or to individuals with a certain minimum or maximum amount of education, and/or to individuals with a certain minimum amount of labor market experience (see e.g. Kluve, 2006). The latter includes eligibility based on profiling systems that determine potential treatments as a deterministic function of a set of individual characteristics. Many policies only apply to certain regions, cantons or states, and as noted above, the introduction and abolition of policies leads to eligibility changes as well. (In Section 3 of the paper we give concrete examples of studies that exploit eligibility variation in instrumental variable settings.) In many of these cases, compliance among the eligible individuals is imperfect, and actual participation is selective. Individuals may influence participation, or the case worker may use her discretionary power to assign individuals based on individual characteristics that are unobserved to the researcher. The same problems arise with random experiments if compliance to the treatment is imperfect. In all these cases, the matching approach can not be used. We propose overcoming this limitation by exploiting the eligibility boundary restriction within the matching framework.

Our approach is related to Battistin and Rettore (2008), who consider a specific partially fuzzy discontinuity design where eligibility rules preclude participation on one side of a threshold for a certain variable and allow - but do not impose - it on the opposite side (so non-compliance affects outcomes on one side of the threshold). To identify a Local Average Treatment Effect (LATE), they need the continuity assumption that is characteristic of sharp regression discontinuity (RD) designs. They note the similarities with the “Bloom

³Here, the word “compliance” is used in a statistical sense, meaning that some of the individuals who, according to the policy design, are eligible for treatment end up in the non-treated subpopulation.

setting”: a fully experimental setting with non-compliance on the treated side only.⁴ We derive a similar estimator, but our derivations are from a matching perspective, and, accordingly, our quantity of interest is the average treatment effect on the treated (ATT). The underlying assumptions are not identical and, as a result, the applicability also differs. In Section 2 we discuss general differences between our approach and the LATE estimator of Imbens and Angrist (1994) and Angrist, Imbens and Rubin (1996), and we consider other combinations of matching and instrumental variable estimation, notably the approach of Baiocchi et al. (2010). Both our estimator and the Battistin and Rettore (2008) estimator lead to a test of the CIA. When deriving our proposed testing procedure (Subsection 2.2) we point out in what sense they differ. Our empirical application (Section 3) illustrates how our approach can be applied in a case where the RD approach is not appropriate.

We empirically assess our approach by evaluating a major Swedish program aimed at helping young unemployed individuals to find work, the Youth Practice (YP).⁵ YP is a subsidized work program designed for short-term unemployed individuals aged below 25. The program is not compulsory. As a result, compliance is imperfect on the lower side of the age-eligibility threshold. We may therefore apply our selectivity-adjusted matching estimator using age as the instrument. In Section 3 we argue that the exclusion restriction is valid. The subpopulation of non-treated includes those below 25 who do not participate as well as those 25 and above. Participation is not sharply discontinuous at age 25 but declines gradually before age 25. This is not a problem for our method but could complicate the application of regression-discontinuity methods. The non-compulsory nature of the program among eligibles may raise difficulties for matching to balance unobservables. We use our

⁴The idea of exploiting one-sided compliance to deal with selective participation has some history in the analysis of treatment effects on duration outcomes in Mixed Proportional Hazard types of models with endogenous treatments. See Bijwaard and Ridder (2005) and Abbring and Van den Berg (2005).

⁵There is an increasing awareness that youth unemployment may be a serious problem for society despite the fact that youth unemployment durations are relatively short. This is because of the prevalence of psychological and labor-market scarring effects which may have long-run implications for the productivity of those affected (see e.g. Burgess et al., 2003).

correction factor to assess whether this is in fact the case and to eliminate the potential selection bias.

YP was introduced in July 1992. Availability of data prior to its introduction allows us to examine an alternative instrumental variable, based on calendar time. In this case, the subpopulation of non-treated includes individuals in the age eligibility group (below 25) who either opt out of YP or flow into unemployment some time before YP becomes available. Using time variation as the instrument, we re-estimate the effect of the treatment and we assess the validity of the CIA.

The Swedish YP has been evaluated before (see e.g. White and Knight, 2002, Larsson, 2003, Forslund and Nordström Skans, 2006, for results). It is of particular interest that existing YP evaluations are based on the matching approach. We find that adjusting the matching estimator for selectivity changes the results to become negative when the outcome of interest is outflow into employment.

In Section 2 we develop a formal framework for the analysis. We define the objects of interest and we derive the selectivity-adjusted matching estimator. In Section 3 we discuss the Swedish YP program, estimation details, data and estimates. Section 4 concludes.

2 A correction term to matching

2.1 Identification of the ATT in case of selective participation and ineligibles

In what follows, we adopt standard counterfactual notation where Y_0 and Y_1 are individual potential outcomes associated with being assigned to non-treatment and treatment, respectively. The binary indicator, D , denotes the actual treatment status, where we use the terms “participation” and “treatment” to denote $D = 1$ and “non-participation” and “control” to denote $D = 0$. The finite-dimensional vector X contains conditioning variables. The actual outcome Y satisfies $Y = DY_1 + (1 - D)Y_0$.

We are interested in the Average Treatment Effect on the Treated (ATT):

$$\text{ATT} = \text{E}[Y_1 - Y_0 \mid D = 1].$$

Clearly,

$$\begin{aligned} \text{ATT} &= \text{E}_{X|D=1} \text{E}[Y_1 - Y_0 \mid X, D = 1] = \\ &\text{E}_{X|D=1} \text{E}[Y_1 \mid X, D = 1] - \text{E}_{X|D=1} \text{E}[Y_0 \mid X, D = 1] \end{aligned} \quad (1)$$

where the expectations $\text{E}_{X|D=1}$ are taken over the distribution of X among the treated. Under the unconfoundedness assumption or Conditional Independence Assumption (CIA) stating that $Y_0 \perp D \mid X$, the ATT is identified and can be estimated using a matching method. We do *not* make such an assumption because we do not rule out that, conditional on observed confounders X , the actual treatment assignment is related to the individual potential outcomes. The most obvious reason to expect violation of the CIA is that there may be individual characteristics that affect both the treatment status and the potential outcome in case of non-participation in the program, where some of these characteristics may be unobserved to the researcher. We therefore refer to violation of the CIA as “selection on unobservables”. Instead of adopting the CIA, we assume that there exists a variable Z with the following two features,

1. $Y_0 \perp Z \mid X$;
2. There exists a set of points $\{z^*, z^{**}\}$ in the domain of Z where

$$\text{P}[D = 1 \mid X, Z = z^*] = 0 \quad \text{and} \quad 0 < \text{P}[D = 1 \mid X, Z = z^{**}] \leq 1$$

for all X .

Assumption 1 states that Z does not explain Y_0 when conditioning on the explanatory variables X . It is a common instrumental variables exclusion assumption, but notice that it is imposed on untreated outcomes only and that it is conditional on observed confounders. In contrast, the LATE estimator further requires that Z has no impact on treated outcomes

and is exogenous in the assignment rule (Imbens and Angrist, 1994; Angrist, Imbens and Rubin, 1996).

Assumption 2 states that D is a non-trivial function of Z after conditioning on X . In particular, it ensures that participation can be driven to zero at certain parts of the distribution of Z . If the participation probability is zero then we call the individual ineligible. The assumption is stronger than the corresponding LATE assumptions, which require the instrument to drive participation (informative), and which rule out “defiers” (i.e., which impose monotonicity).

These assumptions do not rule out that participation is selective. In particular, if $Z = z^{**}$, then D may depend on Y_0 even after conditioning on X . Assumptions 1 and 2 can be called an exclusion restriction and an “informative instrument” assumption, so it is natural to call Z an instrumental variable. Notice that Assumption 2 can be verified empirically, whereas Assumption 1 requires an external justification. A common point of concern for exclusion restrictions like Assumption 1 is that there may be characteristics that affect both the instrument Z and the potential outcome Y_0 , where these characteristics may be unobserved to the researcher. If the instrument is a threshold value of a personal characteristic, one needs to address whether the threshold may represent additional differences between the set of agents on one side of the threshold and the set of agents on the other side. For example, the threshold may be indicative of a stratification of the labor market. Another concern with exclusion restrictions is that individuals may act upon knowledge of their personal value of Z before the treatment status is determined, and this behavior may affect the potential outcomes, leading to a violation of the restriction (Van den Berg, 2007). In Section 3 we discuss the justification of Assumption 2 in our empirical application and in similar potential applications.

Identification of the ATT also hinges on standard assumptions like SUTVA (Stable Unit Treatment Value Assumption) and common support. SUTVA requires the potential outcomes of each individual i to be invariant to the assignment of treatment in the rest of the population (Rubin, 1980 and 1990). Under SUTVA there is no interference between

units or individuals in the population. It effectively rules out effects of treatment arising through market adjustment mechanisms or social interactions.

To discuss the common support assumption, let us define $P^*(X)$ and $P^{**}(X)$ as the density functions of X on the sub-populations of ineligible ($D = 0, Z = z^*$) and untreated eligibles ($D = 0, Z = z^{**}$), respectively. The common support assumption can be stated as

$$0 < P^*(X) \quad \text{and} \quad 0 < P^{**}(X) \quad \text{for all} \quad X \in \text{Supp}(X|D = 1)$$

where $\text{Supp}(X|D = 1)$ is the domain of covariates X among the treated. Empirical applications of matching explicitly impose an overlapping support condition, thus excluding non-overlapping regions. In the latter case, our proposed method will identify the average impact of treatment on treated represented among both non-treated *and* ineligible. This follows standard matching results.

In expression (1) for ATT, the term $E_{X|D=1}E[Y_1 | X, D = 1]$ is directly identified from the mean observed outcome among the treated. The challenge is to identify the mean counterfactual outcome, $E_{X|D=1}E[Y_0 | X, D = 1]$.

Under Assumption 1,

$$\begin{aligned} E[Y_0 | X] &= E[Y_0 | X, Z] \\ &= E[Y_0 | X, Z, D = 0] P[D = 0 | X, Z] + \\ &\quad E[Y_0 | X, Z, D = 1] P[D = 1 | X, Z]. \end{aligned} \tag{2}$$

Since this relationship holds for all possible values of Z , and in particular for $Z = z^*$, Assumption 2 ensures that

$$E[Y_0 | X] = E[Y_0 | X, Z = z^*, D = 0]. \tag{3}$$

On the other hand, the following decomposition always yields,

$$\begin{aligned} E[Y_0 | X] &= E[Y_0 | X, D = 0] P[D = 0 | X] + \\ &\quad E[Y_0 | X, D = 1] P[D = 1 | X] \end{aligned}$$

implying

$$\begin{aligned}
& \mathbb{E}[Y_0 \mid X, D = 1] \\
&= \frac{\mathbb{E}[Y_0 \mid X] - \mathbb{E}[Y_0 \mid X, D = 0] \mathbb{P}[D = 0 \mid X]}{\mathbb{P}[D = 1 \mid X]} \\
&= \frac{\mathbb{E}[Y_0 \mid X, Z = z^*, D = 0] - \mathbb{E}[Y_0 \mid X, D = 0] \mathbb{P}[D = 0 \mid X]}{\mathbb{P}[D = 1 \mid X]} \\
&= \mathbb{E}[Y_0 \mid X, D = 0] + \frac{\mathbb{E}[Y_0 \mid X, Z = z^*, D = 0] - \mathbb{E}[Y_0 \mid X, D = 0]}{1 - \mathbb{P}[D = 0 \mid X]}. \tag{4}
\end{aligned}$$

Equation (4) is an expression for the mean counterfactual outcome $\mathbb{E}[Y_0 \mid X, D = 1]$ given X . The mean counterfactual outcome given X that is used in standard matching estimation, $\mathbb{E}[Y_0 \mid X, D = 0]$, is corrected for individual selection on unobservables by the second term in line four of the equation.

The terms $\mathbb{E}[Y_0 \mid X, D = 0]$ and $\mathbb{E}[Y_0 \mid X, Z = z^*, D = 0]$ in the right-hand side of equation (4) are identified from the mean observed outcome among the controls at given X and the mean observed outcome among the ineligible controls at given X , respectively. Taken together, this implies that the mean counterfactual outcome $\mathbb{E}[Y_0 \mid X, D = 1]$ given X is identified from equation (4). In turn, by the law of iterated expectations, the mean counterfactual outcome $\mathbb{E}_{X|D=1}\mathbb{E}[Y_0 \mid X, D = 1]$ unconditional on X is identified by averaging over the observable distribution of X given $D = 1$. Hence, the ATT is identified. Notice that identification does not require any additivity assumption on the relationships between outcome, treatment, and instrument. Also, identification does not require the instrument to be discrete or to be continuous. In the next subsection we discuss, in some detail, the implementation of the estimator suggested by the above constructive identification proof.

An alternative but similar approach to identification and inference is based on the fact that, in the absence of selection on unobservables, we can discard the ineligible controls and instead use only the eligible controls to obtain the mean counterfactual outcome for the treated. In general, we can express the ATT as

$$\text{ATT} = \mathbb{E}_{[X, Z=z^{**}|D=1]} \mathbb{E}[Y_1 - Y_0 \mid X, Z = z^{**}, D = 1]$$

where z^{**} stands for all possible values of z^{**} satisfying Assumption 2. This expression for the ATT follows from the fact that $D = 1$ automatically implies that $Z = z^{**}$ for some z^{**} satisfying Assumption 2. We will now follow the above derivation of the identification of $E[Y_0 | X, D = 1]$, where we now condition on $Z = z^{**}$ as well as on X .

The mean no-treatment outcome at a specific point $(X, Z = z^{**})$ with a non-zero probability of treatment is

$$\begin{aligned} E[Y_0 | X, Z = z^{**}] &= E[Y_0 | X, Z = z^{**}, D = 0] P[D = 0 | X, Z = z^{**}] \\ &\quad + E[Y_0 | X, Z = z^{**}, D = 1] P[D = 1 | X, Z = z^{**}] \end{aligned}$$

while Assumptions 1 and 2 ensure that

$$\begin{aligned} E[Y_0 | X, Z = z^{**}] &= E[Y_0 | X, Z = z^*] \\ &= E[Y_0 | X, Z = z^*, D = 0]. \end{aligned}$$

But then, the counterpart of (4) if conditioning on Z is

$$\begin{aligned} E[Y_0 | X, Z = z^{**}, D = 1] &= E[Y_0 | X, Z = z^{**}, D = 0] \\ &\quad + \frac{E[Y_0 | X, Z = z^*, D = 0] - E[Y_0 | X, Z = z^{**}, D = 0]}{P[D = 1 | X, Z = z^{**}]}. \end{aligned} \tag{5}$$

The terms $E[Y_0 | X, Z = z^{**}, D = 0]$ and $E[Y_0 | X, Z = z^*, D = 0]$ in the right-hand side of equation (5) are identified from the corresponding observed outcomes. This implies that the mean counterfactual outcome $E[Y_0 | X, Z = z^{**}, D = 1]$ at given X and $Z = z^{**}$ is identified from equation (5). In turn, the mean counterfactual outcome $E_{X, Z = z^{**} | D = 1} E[Y_0 | X, Z = z^{**}, D = 1]$ unconditional on X and Z is identified by averaging over the observable distribution of X, Z given $D = 1$. Again, the ATT follows. Notice that if there is no selection on unobservables then the correction term in (5) vanishes, so only controls used to estimate the ATT are the non-treated eligibles. In this sense, the alternative approach subsumes the instrument Z in the set of conditioning variables X .⁶ With selection on unobservables, of

⁶See Heckman and Lozano (2004) for a discussion of the selection of covariates in matching.

course, the ineligible controls with $Z = z^*$ are also used in the alternative approach to obtain the correction term.

Although we rely on an instrumental variable to correct matching from selection on unobservables, our approach can still recover the ATT. This is in contrast to the approach in the LATE literature initiated by Imbens and Angrist (1994), which imposes a monotonicity assumption to determine the impact of treatment on “compliers” (see also Angrist, Imbens and Rubin, 1996). Instead, we rely on the alternative premise that the instrument can drive participation in a program to zero. By ruling out the existence of “defiers” and “always-takers” and by making all treated “compliers”, our assumption is stronger than (and thus implies) the LATE monotonicity assumption. However, as we have seen, it allows us to relax LATE independence assumptions on the treated outcomes and participation rule.

Similarly, our exclusion restriction (Assumption 1) implies Battistin and Rettore (2008)’s continuity assumption around the threshold point. At the same time, we avoid the local discontinuity in participation that they use for identification and for their LATE estimator.

Baiocchi et al. (2010) develop a treatment evaluation estimator that also combines matching and instrumental variables. Their setting is more closely related to the LATE setting than to our setting. Indeed, it specializes to a LATE setting under the usual LATE interpretation and assumptions. In the first stage of their estimation procedure, they match individuals with low values of the instrumental variable to individuals with high values. Subsequently, they examine the effect of the instrument on the treatment and the effect of the instrument on the outcome. Combining this results in an estimator that can be interpreted as a LATE estimator. As in our case, controlling for covariates makes it less likely that the exclusion restriction is violated by common unobserved determinants of the instrument and the potential outcomes.

2.2 Inference

Our estimation method for the ATT closely follows the above identification proofs. For the sake of brevity we focus on the method for the first approach discussed above, where we use equation (4) to obtain the mean counterfactual outcome among the treated $E(Y_0 | D = 1)$. Equation (4) is conditional on X and $D = 0$, but we need to average it over the observable distribution of X given $D = 1$ to obtain $E(Y_0 | D = 1)$. For this purpose we estimate a propensity score for $P(D = 1 | X)$, using the full sample. Next, we match each treated individual to non-treated individuals, using propensity-score kernel-matching. However, contrary to the standard matching approach to treatment evaluation, we do not take the difference of the outcome of the treated and the matched (weighted mean) outcome of the controls, but we take the difference of the outcome of the treated and the matched (weighted mean) value of the right-hand side of equation (4). In the right-hand side, the separate terms are kernel-smoothed for this purpose, using propensity scores as well. Notice that in the alternative approach based on equation (5), the main propensity score does not only depend on X but also on Z for values $Z = z^{**}$.⁷ In Subsection 3.4 we discuss practical implementation issues for our estimator, in the context of the evaluation of the Swedish Youth Practice (YP) program.

The standard errors are estimated with bootstrapping. Each replication samples from the original data, before matching is performed, and replicates the whole estimation routine. This procedure accounts for all steps in estimation, including estimation of the propensity score. Standard errors are computed directly on each of the estimated parameters, not as a combination standard errors of different terms that may form the ATT.⁸ Because the

⁷The ATT estimates suggested by the two alternative approaches are not necessarily identical. This provides scope for the construction of a general specification test. However, it remains to be seen whether such a test has satisfactory power, as the underlying estimates are driven by outcomes from overlapping subsamples.

⁸This is important as various terms, such as the standard matching estimate and the correction terms, may be correlated.

number of matches increases with sample size and the estimator is asymptotically linear as shown in Heckman, Ichimura and Todd (1998), bootstrapping is expected to provide correct inference when applied to kernel matching (this is at odds with the results for nearest neighbor matching derived in Abadie and Imbens, 2008). It is also much simpler to implement than the asymptotic variance derived in Heckman, Ichimura and Todd (1998).

Under Assumptions 1 and 2 (together with common support and SUTVA), the estimators implied by (4) and (5) are equally consistent for ATT. However, when the population of ineligibles is comparatively large, the denominator in (4) may become excessively small and lead to imprecise estimates of the correction term. This is similar to the classical weak instruments problem. In the context of our estimator, it can be tackled by comparing treated with other eligibles only as in equation (5) if the sizes of treated and non-treated groups among eligibles are comparable.

We may also use the results of the previous subsection to design tests of the usual CIA assumption that $Y_0 \perp D \mid X$, provided that Assumptions 1 and 2 apply. The standard matching method assumes CIA, which implies that the first term of the right-hand side of (4) captures $E(Y_0 \mid X, D = 1)$. As already pointed out, the second term can be labelled a correction term due to selection on unobservables. Thus, the usual CIA assumption holds iff the correction term is zero for any possible X , so iff

$$E[Y_0 \mid X, Z = z^*, D = 0] = E[Y_0 \mid X, D = 0] \tag{6}$$

for any X . In the alternative approach (see (5)), this is replaced by

$$E[Y_0 \mid X, Z = z^*, D = 0] = E[Y_0 \mid X, Z = z^{**}, D = 0]. \tag{7}$$

for any possible X and z^{**} . This can again be aggregated over X and z^{**} . These equalities can be used to test the usual CIA assumption in standard matching estimation. Alternatively, we may test directly whether the correction terms are zero, because these are a by-product of the ATT estimation.

Battistin and Rettore (2008) propose a selection test based on the bias term

$$E[Y_0 | X, Z, D = 0] - E[Y_0 | X, Z, D = 1] \quad (8)$$

defined in regions of Z where participation is not deterministic. Under their RD design with one-sided imperfect compliance, the bias term in (8) can be computed at the eligibility cutoff point. The statistical significance of this term at that specific point provides some information of what may happen elsewhere. In contrast, the matching setup that we explore allows us to directly test the CIA on a larger part of the domain of Z , and therefore on a larger population. This is not empirically irrelevant as sample sizes often preclude meaningful analysis in local discontinuity estimation.

3 Empirical Application: Evaluating Youth Practice

We study the impact of a Swedish youth employment program, the Youth Practice (YP), on the employment probability of young men. The aim of this program is to ease the flow of young unemployed into work by providing work experience. The main focus of our evaluation is its impact on transitions into employment. In what follows we discuss the program, the data, the choice of instruments, the estimation procedure, and the results.

3.1 The program

YP is a Swedish large-scale subsidized-work program targeted at the 18-24 years old unemployed. It was launched in July 1992, in response to the adverse labor market conditions of the early 1990s recession in Sweden. In October 1995 it was subsumed into an extended policy program for youth unemployment.

The YP program was primarily intended for unemployed individuals with a high school diploma. Officially, two rules determined eligibility: (1) that the individual is aged 18 to 24 at the time of enrollment into YP, and (2) that prior to enrollment she/he has been

registered with the employment office for a minimum duration of 4 months for the 20-24 years old and 8 weeks for the 18-19 years old. Empirical data show that the second eligibility requirement was not respected: almost 20 percent of 20-24 years old participants enter YP within 1 month of registering, and over 60 percent enter before completing the first 4 months. In contrast, the age eligibility rule is strictly respected: participants are always 24 or younger at the moment of enrolling into YP.

The treatment consisted of a job placement in the private or public sector for 6 months with a possible extension to 12 months. While at work, YP participants received an allowance below the current wage rate. The employer paid at most a small fraction of the allowance. The job was supposed to be supplementary, in the sense that it should not displace regular employment, and to include a mixture of work and training leading to human capital accumulation. In addition to work, participants were also expected to spend four to eight hours per week actively seeking regular employment at the employment office. In practice, however, the latter guidelines may not have been strictly adhered to. No-displacement and job-search guidelines seem to have been ignored regularly. Most job search requirements were simple, and the proportion of time in training was negligible (see Larsson, 2003, and references therein).

Participation was not compulsory. YP was one among several non-compulsory treatments that agents could enter. Notably, agents could try to participate in Labor Market Training. This is an expensive program that mostly consists of vocational training. However, this program was primarily intended for displaced workers in need of a new type of occupation (Richardson and Van den Berg, 2012). Indeed, YP was by far the most common treatment among young unemployed individuals. In over 22% of the new registration spells of eligible individuals aged 20 to 24 during the YP period, YP is the first reported event after registration at the employment office, whereas the other possible treatments amount to only 16%, of which just over a third concerns Labor Market Training.⁹

⁹We consider exits from the first unemployment period after registration with the employment office. The reported figures refer to individuals aged 20 to 24 when first registering with the employment office

Assignment to YP follows a complex selection mechanism. Similarly to all alternative treatments offered by the Swedish employment offices at the time, assignment to YP was decided in mutual agreement between the unemployed and the employment officer. According to the accounts of job seekers and employment officers alike, YP was generally regarded as providing real work experience (see Larsson, 2003, Forslund and Nordström Skans, 2006). Moreover, candidates interested in YP were encouraged to find a placement on their own to build initiative, job search and representation skills. Together, these two conditions support the view that participants may have been better equipped to find a job than non-participants. In some occasions, however, employment officers offered YP placements to job seekers. The allocation of such treatments may be driven by officers prioritizing candidates in special need of help or close to benefits eligibility exhaustion. Both processes challenge the ability of matching to deal effectively with selection. Our proposed method can be used to assess the presence of residual selection after matching and to correct it when a valid instrument is available.

3.2 Data

3.2.1 Register and sample

We use the Swedish unemployment register called *Händel*. This is an administrative dataset that comprises information from August 1991 onwards on registration spells with the employment offices, labor market status of registered individuals (individuals in low-paid jobs may register), program participation and the subsequent labor market status of those who deregister (e.g. employment, education or inactivity). All individuals with registration spells since 1991 are included in the dataset and their registration and unemployment history can be followed over time. *Händel* also includes demographic information on age, gender, citizenship, area of residence and education.

The *Händel* dataset required considerable cleaning and selection work, mainly due to

between July 1992 and September 1994.

the high incidence of negative and overlapping spells. The criteria applied to construct the final dataset are described in the appendix.

We select observations for young men to estimate the impact of treatment on individuals aged 20 to 24 when first registering with the employment office between August 1991 and September 1994 (or some subgroup thereof). Men aged 18-19 at registration are excluded due to a range of differences with the policy regime for those below 20 (see Forslund and Nordström Skans, 2006, and the other references in Section 1, for details on YP and youth unemployment in Sweden). Observations for older men (age 25 to 29) or men registering prior to the onset of YP are used to construct control groups. Estimation relies on either all new registration spells, independently of the employment status of the new applicant, or the subsample of new registration spells classified as open unemployment for comparison purposes.¹⁰ We consider a single registration spell per individual during the period of interest, namely the earliest one. Registration spells starting after September 1994 are disregarded, as YP take-up slowed substantially from then onwards until extinction in October 1995.¹¹

Two sources of variation will be explored to define our instruments, age and calendar time. Depending on the instrument, data will be sliced in different ways. More details on the construction of treatment and comparison groups, sample sizes and properties, together with the instruments can be found below (see sections 3.5 and 3.6 for results on age and calendar time, respectively).

3.2.2 Eligibility status, treatment status, and outcomes of interest

We aim to measure the impact of YP on subsequent individual labor market outcomes. The potential participants, or *eligibles*, are men starting a new registration spell with the

¹⁰Employed individuals looking for a new job may register with the employment office; they account for less than 4% of all new spells for the population we are considering.

¹¹Among eligibles, YP occurred in only 3% of registration spells starting after September 1994 and under 1% of registration spells starting after January 1995.

employment office between July 1992 and September 1994 while aged 20 to 24. In contrast, *ineligibles* are men who either start a new registration spell when aged 25 or above or prior to the onset of YP in July 1992. *Treatment* is defined as enrolling into YP as the first destination after registration, if enrollment occurs during the first 3 or during the first 6 months of the spell (we define “time to treatment” to be the duration of the registration spell before the enrollment in YP). Obviously, enrollment is only possible for eligibles. The alternative sample of non-eligibles having registered before July 1992 is restricted to those who flowed in sufficiently early to experience all their time to potential treatment before the introduction of YP.

Dynamic selection¹² is beyond the scope of our study. Thus, our definition of treatment status is *unconditional* on time to treatment other than through the time window requirement described above. Likewise, non-treated eligibles and non-eligibles are not selected on time in unemployment before moving into some alternative treatment or deregistering. The severity of any resulting bias from disregarding time to treatment will depend on the time window allowed before enrollment. We consider only relatively short durations prior to enrollment into treatment, of up to 3 or 6 months, and compare results to assess the importance of our choices.¹³ By excluding YP participants enrolling later in their registration spell, we eliminate selection due to the imminent exhaustion of entitlement to unemployment benefits (which occurs after 14 months in the claiming count). We notice that the treated, or participants, do move quickly into YP after registration, with just under 50% and 83% doing so during the first 3 and 6 months of the new spell, respectively.

To assess the impact of enrolling into YP, we measure employment status after registration. The discussion below focuses on the take-up of regular employment, or, alternatively,

¹²Controlling for the duration of unemployment prior to enrollment into treatment would call for a dynamic framework, which has problems of its own, in particular since we do not rule out that there is selection on unobservables; see Abbring and Van den Berg (2003, 2005). See Fredriksson and Johansson (2008) and Crépon et al. (2009), for the use of matching methods that deal explicitly with dynamic enrollment.

¹³Sample sizes become forbiddingly small when further tightening the enrollment window.

on having deregistered to move to any destination, after 12 or after 24 months of registration.¹⁴

3.3 Instrumental variables

For the purpose of illustrating the empirical relevance our method, we consider two alternative instruments. The first is age. It explores the clear cutoff point in eligibility at the moment the individual turns 25, by comparing individuals at each side of the threshold. The second is time and explores the introduction of YP in July 1992 to compare individuals under different policy regimes.

When using age, the exclusion restriction entails that the potential outcome in case of non-treatment should not systematically depend on whether the individual is above or below 25 years old at the moment at which the treatment status is realized, conditional on X . We provide some arguments in support of this. First, notice that although employers may find it relevant whether an applicant for a regular job is above or below 25 years old, they are less likely to be interested in whether the individual was above or below 25 at the time of potential enrollment into YP. As we shall see, we use samples with rather narrow age intervals around 25, so at the time at which the outcomes are realized, many treated individuals are older than 25. Notice also that individuals who are aged around 25 and who enter unemployment often have a relatively low level of education. Such individuals have not yet achieved a high degree of specialized work experience. Hence, their baseline position in the job market is not likely to be strongly dependent on the exact age within the age interval we consider. More in general, the age range within which individuals compete with each other in the job market is probably broader than just a few years.

The use of age eligibility thresholds as instrumental variables is widespread in the empirical evaluation literature. For example, Angrist and Krueger (1991) use age eligibility

¹⁴Since we do not follow individuals throughout their out-of-the-registar periods, all we know is their destination upon leaving. When considering the employment status, we implicitly assume it has not changed after registration.

thresholds for compulsory schooling to estimate the impact of education on earnings, while Gelbach (2002) uses it to estimate the impact of children’s enrollment in public education on mother’s labor supply. Stancanelli and Van Soest (2011) use retirement age thresholds to study the effect of the partner’s retirement on home production. Dickens, Riley and Wilkinson (2010), Olssen (2011) and others use age discontinuities of mandatory minimum wages to study effects of youth minimum wages on individual labor market outcomes. Calendar time thresholds (usually involving the introduction of a new policy) are also widely used as instrumental variables in evaluation studies. Examples are Blundell et al. (2004) and Van den Berg, Bozio and Costa Dias (2010), who use the introduction of a job search assistance program to study participation effects on individual labor market outcomes.

In our setting, the age threshold is our preferred instrument, for two main reasons. First, it allows for a richer set of covariates to be used in matching. Specifically, the use of calendar time forces the exclusion of past (un)employment history from the set of conditioning variables given the coincidental shortly distanced starting dates for data coverage and YP.¹⁵ And second, estimates relying on age comparisons are less vulnerable to changing macro-economic conditions than estimates obtained from comparisons over time. YP was introduced at a difficult time in Sweden, when the national product was contracting and employment prospects were deteriorating. These adverse conditions could, conceivably, bias results relying on time comparisons. In particular, we show that the validity of Assumption 1 is not as clearly established when the instrument is calendar time. Notice that in the presence of heterogeneity in the impact of treatment, the two instruments may identify different parameters as they implicitly define different subsamples of treated. Yet, and despite the specific drawbacks of calendar time as instrument, the results are fairly aligned to those obtained when using age.

¹⁵The use of variables describing past (un)employment history in matching is widely regarded as good practice in labor economics as these are possibly the best conveyors of information on unobserved ability and preferences for work (see Heckman, Ichimura and Todd, 1997).

3.4 Estimation procedure

Irrespective of the instrument being used, the application of our method requires two control groups to be defined. When the counterfactual of interest is as described in equation (4), the *first* control group is the standard matching one. It is drawn from the population of non-participants ($D = 0$) to reproduce the distribution of the matching variables X among the treated. Since the instrument Z is not in X , non-participants comprise both non-eligible individuals ($Z = z^*$, for whom $D = 0$ always) and eligible individuals that opted out of YP as their first activity after registration within the considered unemployment duration ($Z = z^{**}, D = 0$). The *second* control group is required to compute the correction term and draws exclusively from the population of ineligibles ($Z = z^*$), again reproducing the distribution of the matching variables X among the treated.

The alternative estimator described in equation (5) includes the instrument in the set of matching variables when computing the standard matching counterfactual. In this case, the *first* control group is that of eligible non-participants ($Z = z^{**}, D = 0$) and is drawn exclusively from the population of eligibles who did not enroll in YP as their first activity shortly after registration. The *second* control group is as defined above.

Estimation uses propensity score matching with Epanechnikov kernel weights. The propensity score is estimated on all observable characteristics apart from the instrumental variable, namely citizenship, education and region of residence. When using age as the instrument, quarter of entry and labor market history during the year preceding the start of the unemployment spell are also controlled for. If using calendar time, month of entry is included. Moreover, the instrument Z (either age in years or period of enrollment - before or after the introduction of YP) is included in the conditioning set to estimate the standard matching counterfactual using the estimator in equation (5). Matching is performed with replacement in all cases, so each observation in the control group can be used to match multiple treated observations. The choice of bandwidth is based on a global cross-validation exercise. All standard errors are bootstrapped using 200 replications.

In what follows we discuss the estimation details and results for each instrument separately.

3.5 Empirical analysis with instrument 1: age

3.5.1 Descriptives

Age is our preferred instrument. For ease of exposition, we summarize the empirical setting. We explore the cutoff point in eligibility at the 25th birthday and consider the instrument Z to be 0 or 1 depending on whether the individual is past his 25th birthday at registration or not, respectively. The analysis is restricted to men starting a new registration spell during the strong YP period, from its inception in July 1992 to the start of its phasing down period in September 1994. Only one registration spell is considered for each individual, the first to be observed starting on or after July 1992. So eligibles are men registering for the first time with the employment office during the observation window of July 1992 to September 1994 while aged 24 or younger. For them, Z is 1. Non-eligibles are men registering for the first time with the employment office during the observation window of July 1992 to September 1994 while aged 25 or older. For them, Z is 0.

The *treated group* is composed of eligibles who select into YP as their first activity after registering. We consider alternative treatment groups depending on two dimensions:

1. duration of registration spell prior to enrollment into the YP: up to 3 and 6 months;
2. and distance in days to 25th birthday at registration - up to 6 months, 1 year and 2 years. We decided not to tighten this requirement given the small number of treated observations close to the age cutoff point.¹⁶

¹⁶We also estimated the impact of treatment on the sample of individuals as far as 5 years away from their 25th birthday but the ensuing increase in the sample size causes the procedure to become forbiddingly slow when it comes to estimate the precision of the effect. It is also conceivable that our exclusion restriction does not hold for very wide age groups.

Table 1 reports sample sizes by eligibility and participation status for different age groups. Participants are individuals moving into YP as the first activity after registration, irrespective of time to treatment. Column 3 shows that the number of program participants increases more than proportionally with distance to 25th birthday. Among individuals within 3 months of turning 25 at registration, only 81, or 1.5% of the eligibles in this group, become participants; this proportion rises to 24.5% among eligibles within 5 years of their 25th birthday. This happens despite the whole population of treated being used and despite the comparatively high take-up rate among eligibles. The explanation may be a mechanical assignment issue. Although YP is the most popular treatment among young individuals in the registrar, eligibles at the verge of completing 25 years of age at inflow have a short time to enroll into the program. On the contrary, younger agents have comparatively more time, and therefore better chances, to be allocated a place.

This variation in participation rates by age is shown in figure 1. It displays the rate of transition into YP by time since registration among individuals aged 24 at the moment they register and depending on whether they are at more (red curve) or less (green curve) than 4 months from completing 25 years of age. The figure shows that participation rates for the youngest cohort is steadily above zero straight from inflow, peaks at 4 months and starts declining after that. It also provides further detail to the pattern described in table 1, showing that the older cohort participation is concentrated over the first months in unemployment and is never as high as for younger cohorts. As a result, the overall hazard rates are much lower for the whole population of 24 years old at inflow (blue curve) than among those younger than 4 months from completing 25 years of age.

This pattern of participation by age creates a gradual decline in participation rates with age at inflow. Figure 2 depicts it. There is no visible discontinuity to be explored. This is not an ideal empirical setting for an application of regression discontinuity. It may also raise concerns that the selection process depends on distance to 25th birthday at registration. In the empirical analysis below we show that our identification assumptions hold and that the estimated effects of YP are robust to changes in the considered age interval and to the

exclusion of individuals close to the age cutoff point.

3.5.2 Common support

Recovering the ATT and testing the CIA requires the support of X among the treated to be represented among non-treated ($D = 0$) and non-eligibles ($Z = z^*$). If such an extended common support assumption does not hold, the identified effects will represent the treatment effects in the overlapping support region only. To empirically check the common support assumption, figure 3 plots the distribution of the predicted propensity scores by treatment and eligibility status when age is excluded from the covariates set. The population being depicted is that of 24 and 25 years old at registration with the employment office, where treated are individuals moving into YP in the first 6 months of the new registration spell.

Enrollment into treatment seems to be partly dependent on the observable characteristics but the distribution of the propensity score exhibits very little dependence on the eligibility status. In fact, the covariates are relatively balanced between the treated and alternative non-treated groups, even before matching, with a maximum bias of 22%. Matching on the propensity score succeeds in improving balancing for all observables, reducing the bias very substantially in most cases and to a maximum of below 4%.¹⁷

Estimation excluded observations lying below the highest 5th percentile and above the lowest 99th percentile of the distributions of the propensity scores among treated and comparison groups. This selection procedure restricts attention to the overlapping support while moving away from the lower part of the distribution of the propensity score. Equation (4) justifies this asymmetric trimming of the distribution as the estimates of the correction term can be very imprecise for very low values of the propensity score.

¹⁷For the discrete dummy covariates, this test ensures almost perfect alignment between treated and non-treated or non-eligible groups. For continuous covariates, graphical inspection shows that the common support assumption holds for each of them taken independently and that all the post-matching distributions are closer to the distributions among treated. Results available from the authors.

3.5.3 Some empirical evidence concerning Assumption 1 (exclusion restriction)

Conditional on the matching variables, X , Assumption 1 requires age to have no impact on the untreated outcome. While this is an untestable assumption since only treated outcomes are observed among the treated, we may use observations for non-eligibles to test whether this assumption holds in the absence of YP. We explore two alternative comparisons. The first uses information on spells starting during August 1991, the earliest data period, and compares outflows into all destinations and regular employment after 10 months of registration. This is the latest we observe these individuals before YP is introduced and justifies our choice. The second contrasts outflows after 12 and 24 months of registration for older ineligible individuals during the YP period. We compare men close to either their 26th or 27th birthday. Table 2 shows the results of t-tests for the comparison of means (columns 1 and 2) and Kolmogorov-Smirnov tests for the comparison of the entire distributions of untreated outcomes.¹⁸ All statistics were computed after matching on the set of characteristics used to estimate the treatment effect.

Overall there is no evidence that Assumption 1 is violated in this application, at least when considering populations not exposed to the YP. All p-values are above the standard 5% significance level and most are well above. However, these results do not exclude the possibility of Assumption 1 being violated once YP is introduced given its age-eligibility rule that in practice means younger individuals can wait longer before enrolling. We explore this possibility below with some sensitivity analysis of our results to different choices of comparison groups.

¹⁸We omit test results on outflows after 12 months of registration as they are qualitatively identical to the results on outflows after 24 months. They can be requested from the authors.

3.5.4 Results

Our preferred sample comprises males registering with the employment office within 1 year of their 25th birthday while YP is operating in full (between July 1992 and September 1994). This sample is large enough to support precise estimates while ensuring that age differences between eligibles and ineligibles do not compromise the validity of Assumption 1. Among eligibles at inflow, just over 2% (511 observations) flow into YP within 1 month, 5.5% (1,182 observations) within 3 months and almost 9% (1,887 observations) within 6 months of registering with the employment office. Our main estimates use the latter group of participants for the sake of sample size. However, we will also present alternative comparisons using different age groups and unemployment durations before enrolling into YP to ensure our results are not driven by dynamic selection mechanisms related with age.

Table 3 displays the estimates of the ATT on the probability of finding a regular job within 12 and 24 months of registering with the employment office.¹⁹

Row 1 in the table displays the main set of estimates, based on individuals aged 24

¹⁹Estimates for males only. Sample selection criteria varies by row as detailed in row titles. All estimates based on sample of new registrations with the employment office. “Treatment” in rows 1 to 5 stands for flowing into YP within 6 months of registering with employment office as first destination after registration. Row 1 compares treated aged 24 at registration with non-treated aged 24 or 25 at registration. Row 2 restricts the sample to those registering as open unemployed. Row 3 restricts the sample of eligibles to 24 years old at more than 4 months from their 25th birthday at registration. Row 4 restricts the control group in standard matching to the eligibles (aged 24 at registration). Finally, row 5 redefines “treatment” as flowing into YP as first destination within 3 months of registration and compares treated aged 24 at inflow with non-treated aged 24 or 25. The impact of treatment is estimated on the probability of moving into employment within 12 months (columns 1 to 3) and 24 months (columns 4 to 6) of registration. Columns 1 and 4 display standard matching estimates. Columns 2 and 5 display the correction term as specified in the right-hand side of equation (4) or, for row 4, of equation (5). Columns 3 and 6 display the corrected matching estimates using the counterfactuals as specified in equation (4) or, for row 4, in equation (5). The number of observations in column 7 are for the treated group only. Matching on the propensity score using kernel Epanechnikov weights with a bandwidth of 0.02 for a probability ranging in the unit interval. Bootstrapped standard errors based on 200 replications in brackets below the estimate.

and 25 at registration and defining treatment as flowing into YP during the first 6 months as the first destination after registration. Standard matching estimates suggest YP has a null effect on the probability of moving into employment within 12 (column 1) and 24 (column 4) months of registration. The corrected matching estimates corroborate this result when applied to the 12 months' outcome. However, the figure regarding outflows within 24 months of registration is significantly different. The correction term suggests that treated are not randomly selected once observables have been controlled for. Instead, the treated seem to be comparatively better positioned to find a job in the absence of treatment than similar non-treated. This is in line with the fact that treated are encouraged to find their own job positions and that such selection mechanism might dominate in the beginning of the registration spells. The consequence is the large and significantly negative effect of treatment on outflows to employment identified by the corrected matching estimator.

To assess the robustness of this result, we tried several alternative comparisons. Some of the results are displayed in the other rows of Table 3. We restrict the sample to those registering as open unemployed in row 2. We exclude eligibles at less than 4 months of completing 25 years of age in row 3 to control for potential bias arising from older individuals in the eligible group to be rushed into YP while still eligible. And we restrict the control group in standard matching to be composed only of non-treated eligibles in row 4. All results are consistent with those shown in row 1. Only in row 4 are the corrected estimates after 24 months of registering not statistically significant at 5% significance level, but the exhibited pattern is similar to all other cases.²⁰

The last row of Table 3 considers treatment to be 'starting YP within first 3 months after registering'. If dynamic selection issues were important at these relatively short durations, we would expect the results to show some response to such change in the definition

²⁰We also considered using other groups: restricting the sample to Swedish citizens; focusing on individuals with vocational training only (the largest educational group with registration spells); and including exits to registered employment as a positive outcome. All results are consistent with those displayed in Table 3 and are available from the authors under request.

of treatment. However, estimates in row 5 are very similar in size and pattern to those displayed in the other rows of the same table. The robustness of these results suggests our preferred time window is sufficiently narrow to keep time of treatment exogenous in this analysis.

We further investigate the sensitivity of these results to age in Table 4 by varying the width of the age interval around the 25th birthday at registration. Columns 1 to 3 display estimates of the effect of YP on the odds of finding a job within 24 months of registration. For comparison purposes, the first row repeats the last three columns in the first row of Table 3. The following two rows display results for the population of men up to 2 years (row 2) and half year (row 3) away from their 25th birthday at registration. Neither widening or narrowing the age interval changes the pattern of the results. However, results in row 3 are substantially larger but very imprecisely estimated given the small sample size.

Columns 4 to 6 display results on an alternative outcome, deregistration within 24 months of first registering. The classical matching estimate for 24-25 years old (row 1, column 4) suggests a negative overall impact of the program, maybe due to an extended lock-in effect or to the extension of eligibility to benefits as a consequence of treatment take-up.²¹ A similar result holds for 23-26 (row 2). In both cases, however, the correction points to the opposite direction and the resulting effect is found to be positive and statistically significant when the larger age group is used (row 2). Again here, sample size precludes a clear pattern to emerge from the analysis of the narrower age group (row 3).²²

Results for both outcomes are considerably stable across age groups. Such lack of variation is consistent with an homogeneous effect of treatment by age for the interval being considered.

²¹The Swedish welfare system provides unemployment insurance for a limited amount of time after a transition from employment into unemployment. However, this period can be extended by participation in the programs made available by the employment offices, of which YP is one example. Repeated participation would, in principle, allow the unemployed to remain out of work and on benefits indefinitely.

²²We have estimated effects on all outflows on other samples as in table 3 and applying the alternative estimator as defined in equation (5). Results are similar to those discussed here.

Overall, both tables suggest that standard matching may not be identifying the correct causal effect of interest (i.e., the ATT). Standard-matching results suggest that the program has no effect on the probability of finding a job and a small negative effect on the overall odds of leaving unemployment. Correcting for the potential selection bias in matching changes the picture quite substantially. The program seems to strongly reduce employment take-up in the medium run, after 24 months of registration. With regard to the overall impact on the odds of leaving unemployment, our estimation strategy suggests YP has either a zero or a small positive effect, within the same time frame. Analysis of other outcomes suggest that the possible positive effect of YP on deregistration is driven by exits into formal education (these estimates are available under request from the authors).²³

3.6 Empirical analysis with instrument 2: calendar time

We explore the introduction of YP in July 1992 and compare pre- and post-YP periods to construct a new set of estimates of the impact of treatment on the probabilities of employment and deregistration. To minimize bias from changing macro conditions and seasonal variation, we select the closest possible periods allowing for the construction of a non-eligible, pre-YP group over the same calendar months as the post-YP group. Specif-

²³Swedish subsidized work programs have been the focus of other studies. In particular, Sianesi (2004) analyzes the overall impact of the Swedish ALMP system and the differential impact of each of the numerous available treatments for adults (so this excludes YP). She finds that subsidized employment is the best performer in terms of moving unemployed back into work, and that the positive effect of subsidized employment seems to last. All other programs have either a zero or a negative impact, possibly arising through the renewed eligibility to benefits as a consequence of program participation. Larsson (2003) studies specifically the effects of YP on exits to employment and finds negative effects 12 months after treatment using standard matching techniques. More generally, youth programs have often shown disappointing results (see surveys by Heckman, LaLonde and Smith, 1999, and Bergemann and Van den Berg, 2008). A noticeable positive exception are programs that mix improved job-search assistance and tougher job-search monitoring such as the British New Deal for Young People (e.g. Blundell et al., 2004, De Giorgi, 2005, Anderson, 2000, Van den Berg et al., 2004, and Van den Berg and Van der Klaauw, 2006).

ically, the instrument Z now describes a discretization of calendar time, being 0 or 1 for registration spells starting between August and December 1991 or 1992, respectively.²⁴ As before, the treatment consists of enrolling in YP as the first activity after registration when this occurs in the first 3 or 6 months of the new spell. Estimation uses observations for men aged 24 or younger at registration only, all satisfying the age-eligibility rule. Thus eligibility is exactly defined by the instrument: eligibles (ineligibles) are those who register in 1992 (1991). The treated are eligibles who move into YP before any other alternative in the first 3 or 6 months after registration. Sample sizes are presented in table 5.

Empirical evidence is not as favorable to the use of calendar time as it is to age. Relying on older never-eligible individuals, we check Assumption 1 in the absence of treatment. Table 6 displays the p-values of tests comparing means and distributions of older ineligible registering in 1991 and 1992. There is strong evidence that registration and employment status 24 months after inflow are markedly different by year of registration (rows 3 and 4). A comparison of employment rates reflects the deteriorating macro conditions over this period. The tests fail to reject the equality of means and distributions of deregistration probabilities after 12 months of registration (rows 1 and 2, columns 2 and 4), but do not support conditional independence of time variation and the probability of employment (rows 1 and 2, columns 2 and 4). We proceed by focusing on 12 months' outcomes but interpret any results with extreme care while noticing they serve illustrative purposes on the use of a widely available instrument.

Table 7 shows the estimated effects and corrections. Matching yields negative and significant effects of YP on deregistration and employment after 12 months of inflow into the spells of interest among 24 years old (row 1). The correction term is positive, suggesting participants are drawn from a group with lower attachment to unemployment, but non-significant. It leads to a non-significant corrected estimate. These results are more pessimistic but broadly in line with those found using age as instrument.

Ineligibles face at least 6 additional months of pre-YP period after registration but are

²⁴August 1991 is the first month in Handel.

then exposed to the possibility of enrolling into YP if still young enough.²⁵ The latter is in contrast with the conditions facing ineligibles as defined by the age instrument but does not per se compromise the quality of calendar time as an instrument given our definition of treatment. Nevertheless, we computed estimates for those at less than 6 months from their 25th birthday, who are never eligible if registering in 1991. Results are displayed in row 2 but the small number of observations precludes any statistically significant conclusion to be drawn.

Row 3 extends the analysis to all those aged 20 to 24 at registration. The pattern is now better defined, suggesting that treatment may reduce the odds of employment (column 3) and deregistration (column 6) by more than implied by standard matching (columns 1 and 4) due to selection of treated on unobserved characteristics (columns 2 and 5). However, the feeble evidence supporting Assumption 1 cannot exclude the alternative interpretation that time trends are driving the negative results. Restricting the comparison group in standard matching to eligible non-treated individuals as in row 4 eliminates the time effects from the estimates in columns 1 and 4 and may accentuate the correction in columns 2 and 5. Yet, we find only mild evidence of this pattern, with standard matching estimates being still significantly negative and correction terms being of similar order of magnitude as those in row 3.

4 Conclusion

We have developed and applied an evaluation method for the effects of program participation (or policy exposure) on individual outcomes, if participation is selective but individuals are ineligible in case of a certain value of some observed instrumental variable. From a practical point of view this is a common setting, in particular for active labor market policies for young individuals. In those cases, participation may be selective because individuals

²⁵Given the definition of treatment adopted here and the calendar time periods we propose to compare, participation is possible among eligibles only.

can choose between different programs and/or because the duration until enrollment is not deterministically set. Program participation is only possible if the individual is aged below a certain age. With selective participation, if the CIA is violated, matching cannot be used. For the same reason, one cannot simply compare those below the threshold who are treated to those above the threshold (who are all non-treated). However, our novel method, which exploits the eligibility boundary restriction within the matching framework, provides consistent estimates of the average treatment effect on those who are treated.

Our approach relies critically on the availability of an instrument satisfying Assumptions 1 and 2 in Section 2. Assumption 2 is automatically satisfied in our preferred practical application of a policy that allows for selective participation only on certain values of an observed variable. To obtain precise estimates of our correction term, however, we also require the program to generate a reasonable number of participants to avoid dividing by a number close to zero.

The application to the Swedish Youth Practice program shows that our method can deliver evaluation results that differ from those based on standard matching methods. The standard matching estimates for the effect on re-employment are zero, whereas the estimates based on our method can be strongly negative. Using our preferred instrument age, the difference between the estimates is systematically significant when the outcome of interest is “finding a job within 24 months of becoming unemployed”. The effects on the overall exit probabilities out of unemployment are invariably estimated to be smaller than those based on matching, although the differences here are not significant. Our results confirm suspicions that treated differ systematically from comparable non-treated, suggesting they are drawn from a more employable group. As a result, we are more pessimistic about the effect of subsidized work on the rate of finding work than if we had incorrectly based ourselves on the matching estimates, while overall exit rates from unemployment are less negative (even possibly positive) than if we had relied on standard matching alone. The latter exits mostly involve outflows into formal education. From a policy point of view, our results suggest that perhaps the optimism about the use of subsidized work programs to

bring unemployed youth back to work should be tempered.

References

- Abadie, A. and G.W. Imbens (2008), “On the failure of the bootstrap for matching estimators”, *Econometrica*, 76, 1537–1557.
- Abbring, J.H. and G.J. van den Berg (2003), “The non-parametric identification of treatment effects in duration models”, *Econometrica*, 71, 1491–1517.
- Abbring, J.H. and G.J. van den Berg (2005), “Social experiments and instrumental variables with duration outcomes”, Working paper, IZA, Bonn.
- Anderson, P. (2000), “Monitoring and Assisting Active Job Search”, OECD Proceedings, *Labour Market Policies and the Public Employment Service*
- Andrews, D. and M. Schafgans (1998), “Semiparametric Estimation of the Intercept of a Sample Selection Model,” *Review of Economic Studies*, 65, 497–517
- Angrist, J.D., G.W. Imbens, and D.B. Rubin (1996), “Identification of causal effects using instrumental variables”, *Journal of the American Statistical Association*, 91, 444–455.
- Angrist, J.D. and A.B. Krueger (1991), “Does compulsory school attendance affect schooling and earnings?”, *Quarterly Journal of Economics*, 106, 979–1014.
- Baiocchi, M., D.S. Small, S. Lorch, and P.R. Rosenbaum (2010), “Building a stronger instrument in an observational study of perinatal care for premature infants”, *Journal of the American Statistical Association*, 105, 1285–1296.
- Battistin, E. and E. Rettore (2008), “Ineligibles and Eligible Non-Participants as a Double Comparison Group in Regression Discontinuity Designs”, *Journal of Econometrics*, 142, 715–730
- Bergemann, A.H. and G.J. van den Berg (2008), “Active labor market policy effects for women in Europe – a survey”, *Annales d’Économie et de Statistique*, 91/92, 385–408
- Bijwaard, G. and G. Ridder (2005), “Correcting for selective compliance in a re-employment bonus experiment”, *Journal of Econometrics*, 125, 77–111.

- Blundell, R., M. Costa Dias, C. Meghir and J. Van Reenen (2004), “Evaluating the Employment Impact of a Mandatory Job Search Program”, *Journal of the European Economic Association*, 2(4), 569–606
- Bonnal, L., D. Fougère, and A. Sérandon (1997), “Evaluating the Impact of French Employment Policies on Individual Labour Market Histories,” *Review of Economic Studies*, 64, 683–713
- Burgess, S., C. Propper, H. Rees and A. Shearer (2003), “The Class of 1981: The Effects of Early Career Unemployment on Subsequent Unemployment Experiences”, *Labour Economics*, 10, 291–309
- Card, D. and D. Sullivan (1988), “Measuring the Effect of Subsidized Training Programs on Movements in and out of Employment,” *Econometrica*, 56, 497–530
- Cochrane, W. and D. Rubin (1973), “Controlling Bias in Observational Studies,” *Sankhya*, 35, 417–446
- Crépon, B., M. Ferracci, G. Jolivet and G.J. van den Berg (2009), “Active labor market policy effects in a dynamic setting”, *Journal of the European Economic Association*, 7, 595–605
- De Giorgi, G. (2005), “Long-term effects of a mandatory multistage program: the New Deal for Young People in the UK”, Working paper, IFS, London
- Dickens, R., R. Riley and D. Wilkinson (2010), “The impact on employment of the age related increases in the National Minimum Wage”, Working paper, LSE, London.
- Forslund, A. and O. Nordström Skans (2006), “Swedish youth labour market policies revisited”, Working paper, IFAU, Uppsala
- Fredriksson P. and P. Johansson (2008), “Dynamic Treatment Assignment - The Consequences for Evaluations using Observational Data”, *Journal of Business and Economics Statistics*, 26, 435–455.

- Gelbach, J.B. (2002), “Public schooling for young children and maternal labor supply”, *American Economic Review*, 92, 307–322.
- Gritz, R. (1993), “The Impact of Training on the Frequency and Duration of Employment,” *Journal of Econometrics*, 57, 21–51
- Heckman, J., H. Ichimura, and P. Todd, (1998) “Matching as an Econometric Evaluation Estimator,” *Review of Economic Studies*, 65, 261–294
- Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998) “Characterization of Selection Bias Using Experimental Data,” *Econometrica*, 66, 1017–1098
- 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 Labor Economics*, Volume 3, North-Holland, Amsterdam
- Heckman, J., and S. Lozano (2004), “Using matching, instrumental variables and control functions to estimate economic choice models”, *Review of Economics and Statistics*, 86, 30–57
- Heckman, J. and R. Robb (1985) “Alternative Methods for Evaluating the Impact of Interventions,” in J. Heckman and B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*, Cambridge University Press, New York
- Imbens, G.W. and J.D. Angrist (1994), “Identification and estimation of local average treatment effects”, *Econometrica*, 62, 467–475.
- Kluve, J. (2006), “The Effectiveness of European Active Labor Market Policy”, Working paper, RIW, Essen
- Larsson, L. (2003), “Evaluation of Swedish Youth Labor Market Programs”, *Journal of Human Resources*, 38, 891–927
- Olssen, A. (2011), “The short run effects of age based youth minimum wages in Australia: a regression discontinuity approach”, Working paper, Melbourne Institute, Australia.

- Richardson, K. and G.J. van den Berg (2012), “Duration dependence versus unobserved heterogeneity in treatment effects: Swedish labor market training and the transition rate to employment”, *Journal of Applied Econometrics*, forthcoming.
- Rosenbaum, P. and D. Rubin (1983) “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika*, 70, 41–55
- Rubin, D. (1980), Comment on “Randomization Analysis of Experimental Data: The Fisher Randomization Test,” by D. Basu, *Journal of the American Statistical Association*, 75: 591-93.
- Rubin, D. (1990), Comment on “Neyman (1923) and Causal Inference in Experiments and Observational Studies,” *Statistical Science*, 5: 472-80.
- Sianesi, B. (2004), “An Evaluation of the Swedish System of Active Labour Market Programmes in the 1990s”, *Review of Economics and Statistics*, 86, 1, 133–155
- Stancanelli, E.G.F. and A. van Soest (2011), “Retirement and home production: a regression discontinuity approach”, Working paper, IZA, Bonn.
- Van den Berg, G.J., B. van der Klaauw, and J.C. van Ours (2004), “Punitive Sanctions and the Transition Rate from Welfare to Work,” *Journal of Labor Economics*, 22, 211–241
- Van den Berg, G.J. and B. van der Klaauw (2006), “Counselling and Monitoring of Unemployed Workers: Theory and Evidence from a Controlled Social Experiment, *International Economic Review*, 47, 895–936
- Van den Berg, G.J. (2007), An economic analysis of exclusion restrictions for instrumental variable estimation, Working paper, IZA, Bonn.
- Van den Berg, G.J., A. Bozio and M. Costa Dias (2010), Policy discontinuity and duration outcomes, Working paper, University of Mannheim.
- White, M. and G. Knight (2002), “Benchmarking the effectiveness of NDYP: A review of European and US literature on the microeconomic effects of labour market programmes for young people”, Working paper, PSI, London

Appendix: Data cleaning and selection

Händel is an administrative dataset comprising information on all registered unemployment spells from August 1991 onwards. It details longitudinal information on the whole population of registered spells, including any undertaken treatments, the history of earned subsidies, destination on leaving the registrar and demographics such as age, citizenship, education and usual occupation.

The main obstacle in using Händel is the frequency of negative and overlapping spells. We have dealt with these occurrences in a conservative way to minimize any resulting bias introduced by data handling.

To start with, we created a condensed variable describing labor market status while in registrar. The four broad categories considered are: unemployment, registered employment, YP, all other possible treatments. Using these, we collapsed all overlapping spells in the same broad category. Spells in different broad categories overlapping by no more than 2 weeks were corrected by setting the exit date of the earliest equal to the entry date of the latest as exit dates are generally more imprecise. Zero duration spells were discarded. At last, individual histories with a remaining error were censored from the time of the error onwards and a censoring indicator was created to correct estimates for the possibility of censored histories.

Data selection followed a number of criteria. First, we used only males. Then we selected individuals starting a new registered *unemployment* spell during the period YP was more popular, between July 1992 and September 1994. Of all the selected spells, we kept only the first one and followed the corresponding individuals over time to find out about treatment take up and labor market outcomes. We considered individuals aged between 20 and 29 at the time of registration and classified as eligibles those aged 24 or younger.

In running the estimation procedure, we also focused on more narrowly defined groups as defined by the following variables: (*i*) Distance to 25th birthday at inflow: less than 2 years, less than 1 year, between 12 and 4 months (this latter condition applies to eligibles

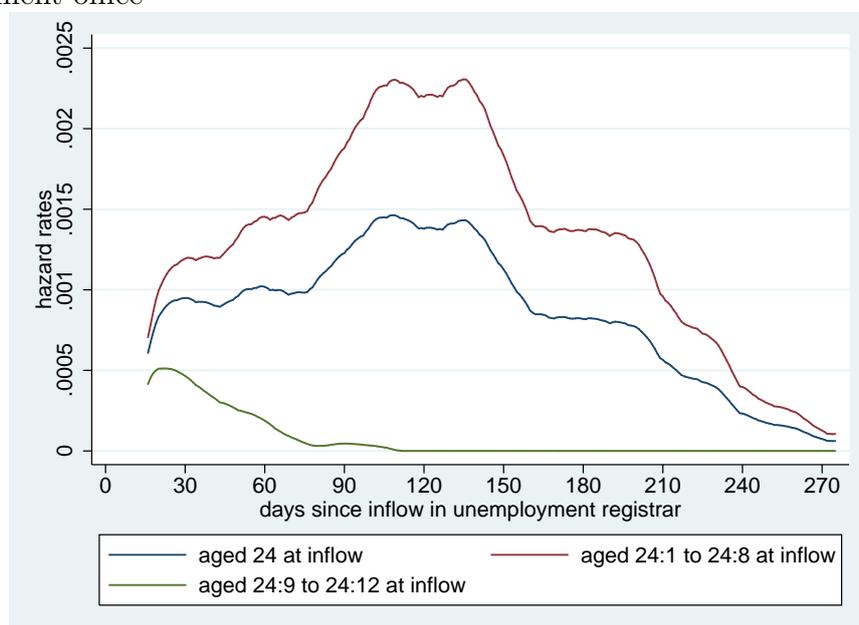
only, and compares with ineligibles less than 1 year away from 25th birthday at inflow); *(ii)* Nationality: Swedish nationals only or all new unemployed; *(iii)* Employment status at registration: whether or not registering as open unemployed; and *(iv)* Education attainment: vocational training or all levels of education.

Table 1: Number of observations by age group and eligibility/treatment status; instrument is age; men only

Distance to 25th birthday at inflow	ineligibles	eligibles (under 25)		Total
	(over 25)	non-participants	participants	
	(1)	(2)	(3)	(4)
(1) up to 3 months	5,444	5,240	81	10,765
(2) up to 6 months	11,029	10,157	448	21,634
(3) up to 1 year	21,950	19,428	2,029	43,407
(4) up to 2 years	43,683	37,118	6,064	86,865
(5) up to 5 years	102,450	112,501	32,528	247,479

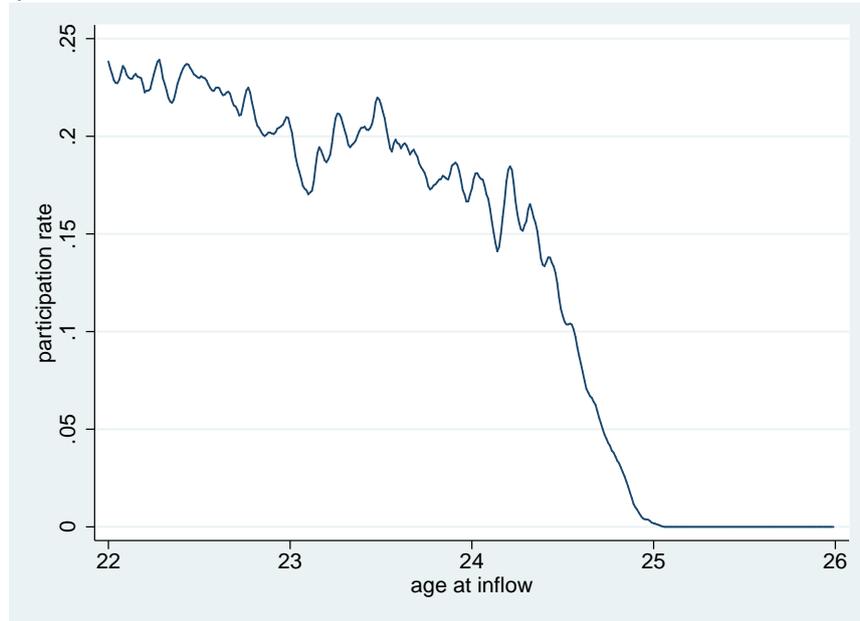
Notes: Population of males close to the 25th birthday when registering with the employment office between July 1992 and September 1994. Age groups defined by distance to 25th birthday when first registering within timeframe. Each individual is represented only once in the sample, when first registering. Eligibles (ineligibles) are those aged 24 and below (25 and above) at registration. Participants are those taking YP as the first event after registration.

Figure 1: Hazard rates into YP by duration of unemployment spell and age at registration with employment office



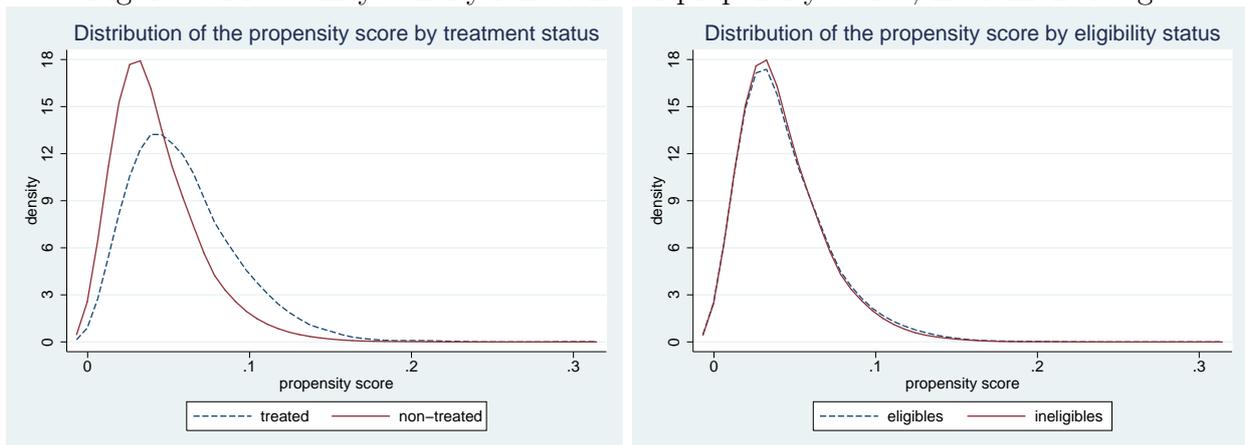
Notes: Plotted curves are smoothed Kaplan Meyer hazard rates using Local Linear Regression with a bandwidth of 15 days. Population of males aged 24 when registering with employment office for the first time between July 1992 and September 1994.

Figure 2: Probability of participation by age at inflow into new registered unemployment spell; men only



Notes: Population of males aged 22 to 26 when registering with the employment office for the first time between July 1992 and September 1994. 'Participation' means flowing into YP as first event after registration.

Figure 3: Probability density functions for propensity scores; instrument is age



Notes: Plotted curves are probability density functions of propensity scores estimated on the population of men aged 24 and 25 when registering with the employment office between July 1992 and September 1994. Treated are 24 years old moving into YP as first destination within 6 months of inflow. Non-treated are 24 and 25 years old not participating in YP as first event within 6 months of inflow. Eligibles (ineligibles) are those aged 24 (25) at registration.

Table 2: P-values for tests of Assumption 1; instrument is age

		equality of means		equality of distribution		
		deregistration	employment	deregistration	employment	observ.
		(1)	(2)	(3)	(4)	(5)
Outflows after 10 months of registration for spells starting during Aug 1991						
Comparing men aged 24 or younger with men aged 25 or older at registration						
(1)	24 versus 25	0.082	0.814	0.990	0.605	3,959
(2)	24:7-12 vs 25:1-6	0.826	0.516	0.938	0.843	1,928
(3)	24:1-8 vs 25	0.100	0.898	0.591	0.266	3,323
(4)	23-24 vs 25-26	0.091	0.910	0.948	0.988	7,913
Outflows after 24 months of registration for spells starting during YP period						
Comparing men aged close to 25 or 26 at registration						
(5)	25 versus 26	0.264	0.289	0.385	0.811	27,625
(6)	25:7-12 vs 26:1-6	0.138	0.904	0.961	0.401	13,921
(7)	25-26 vs 27-28	0.739	0.927	0.071	0.996	54,458

Notes: Columns 1 and 2 display the p-values for the difference in means. Columns 3 and 4 display the p-values for the Kolmogorov-Smirnov test of equality of distributions. Rows 1 to 4 use registration spells starting in August 1991 and follows them for 10 month, till just before the launch of YP. Rows 5 to 7 use registration spells of older individuals during the YP period. Bootstrapped standard errors using 200 replications.

Table 3: ATT on the outflows to regular employment

Outcome: employment							
12 months after registration			24 months after registration			nr of observ.	
classical matching	correction term	adjusted matching	classical matching	correction term	adjusted matching		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	
<i>Treated: 24 years old moving into YP within 6 months of registration</i>							
<i>Controls: ineligible (25 years old) and eligible (24 years old) non-treated</i>							
(1)	-0.009 (0.011)	-0.015 (0.045)	0.006 (0.047)	-0.008 (0.012)	0.096* (0.048)	-0.104* (0.050)	1,699
<i>Treated: 24 years old moving into YP within 6 months of registration - open unemployed</i>							
<i>Controls: ineligible (25 years old) and eligible (24 years old) non-treated - open unemployed</i>							
(2)	-0.012 (0.011)	-0.022 (0.045)	0.010 (0.048)	-0.010 (0.013)	0.100* (0.050)	-0.109* (0.051)	1,606
<i>Treated: 24:1 to 24:8 years old moving into YP within 6 months of registration</i>							
<i>Controls: ineligible (25 years old) and eligible (24:1 to 24:8 years old) non-treated</i>							
(3)	-0.011 (0.012)	-0.002 (0.036)	-0.008 (0.038)	-0.004 (0.014)	0.070* (0.035)	-0.073* (0.036)	1,579
<i>Treated: 24 years old moving into YP within 6 months of registration</i>							
<i>Controls: eligible (24 years old) non-treated</i>							
(4)	-0.015 (0.013)	-0.041 (0.048)	0.026 (0.047)	-0.007 (0.014)	0.042 (0.052)	-0.049 (0.051)	1,563
<i>Treated: 24 years old moving into YP within 3 months of registration</i>							
<i>Controls: ineligible (25 years old) and eligible (24 years old) non-treated</i>							
(5)	0.006 (0.014)	0.005 (0.070)	0.001 (0.071)	0.004 (0.017)	0.154* (0.075)	-0.150* (0.075)	1,049

Notes: Standard errors in parenthesis below the estimate.

* : Statistically different from zero at 5% significance level. See footnote 19 for additional explanations.

Table 4: ATT on outflows to employment and deregistration within 24 months of registration

	Outcome: employment			Outcome: deregistration			
	classical matching	correction term	adjusted matching	classical matching	correction term	adjusted matching	nr of observ.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Treated: 24 years old moving into YP within 6 months of registration</i>							
<i>Controls: ineligible (25 years old) and eligible (24 years old) non-treated</i>							
(1)	-0.008 (0.012)	0.096* (0.048)	-0.104* (0.050)	-0.032* (0.012)	-0.075 (0.049)	0.043 (0.051)	1,699
<i>Treated: 23-24 years old moving into YP within 6 months of registration</i>							
<i>Controls: ineligible (25-26 years old) and eligible (23-24 years old) non-treated</i>							
(2)	0.006 (0.008)	0.082* (0.028)	-0.076* (0.028)	-0.031* (0.008)	-0.151* (0.030)	0.120* (0.030)	4,468
<i>Treated: 24:7 to 24:12 years old moving into YP within 6 months of registration</i>							
<i>Controls: ineligible (25:1 to 25:6 years old) and eligible (24:7 to 24:12 years old) non-treated</i>							
(3)	-0.001 (0.026)	0.187 (0.152)	-0.188 (0.154)	-0.011 (0.026)	0.098 (0.147)	-0.109 (0.150)	401

Notes: Estimates for males only. Sample selection criteria varies by row as detailed in row titles. All estimates based on sample of new registrations with the employment office. “Treatment” stands for flowing into YP within 6 months of registering with employment office as first destination after registration. Row 1 compares treated aged 24 at registration with non-treated aged 24 or 25 at registration. Row 2 uses the sample of individuals at less than 2 years from 25th birthday on registration. Row 3 uses only individuals at less than 6 months from their 25th birthday at registration. The impact of treatment is estimated on the probability of moving into employment within 24 months (columns 1 to 3) and deregistering within 24 months (columns 4 to 6) of inflow. Columns 1 and 4 display standard matching estimates. Columns 2 and 5 display the correction term as specified in the right-hand side of equation (4). Columns 3 and 6 display the corrected matching estimates using the counterfactuals as specified in equation (4). Matching on the propensity score using kernel Epanechnikov weights with a bandwidth of 0.02 for a probability ranging in the unit interval. Bootstrapped standard errors based on 200 replications in brackets below the estimate.

* Statistically different from zero at 5% significance level.

Table 5: Number of observations by age group and eligibility/treatment status; instrument is calendar time; men only

		ineligibles	eligibles (1992 inflow)		
Age at		(1991 inflow)	non-participants	participants	Total
registration		(1)	(2)	(3)	(4)
registration period: Aug-Dec					
(1)	age: 20-24	44,508	33,631	9,376	87,515
(2)	age: 24	7,743	6,410	820	14,973
(3)	age: 24:7-12	3,857	3,345	189	7,391

Notes: Population of males younger than 25 when registering with the employment office during August to December 1991 or 1992. Each individual is represented at most once in each period (pre- and post-YP, 1991 and 1992 respectively). Eligibles (ineligibles) are registering in 1992 (1991) and participants are eligibles who flow into YP as the first event after registration.

Table 6: P-values for tests of Assumption 1 on older ineligible - instrument is calendar time

		equality of means		equality of distribution		nr observ.
		employment	deregistration	employment	deregistration	
		(1)	(2)	(3)	(4)	
Outflows after 12 months of registration for older ineligible men						
Comparing registration spells starting in 1992:8-12 and 1991:8-12						
(1)	25 to 29 years old	0.031	0.112	0.058	0.132	3,959
(2)	25 years old	0.040	0.711	0.071	0.479	1,928
Outflows after 24 months of registration for older ineligible men						
Comparing registration spells starting in 1992:8-12 and 1991:8-12						
(3)	25 to 29 years old	0.000	0.000	0.006	0.000	3,323
(4)	25 years old	0.000	0.000	0.003	0.000	7,913

Notes: Columns 1 and 2 display the p-values for the difference in means. Columns 3 and 4 display the p-values for the Kolmogorov-Smirnov test of equality of distributions. Rows 1 and 2 (3 and 4) describe effects on employment and registration status 12 (24) months after registration. Bootstrapped standard errors using 200 replications.

Table 7: ATT on employment and deregistration probabilities 12 months after registration

	Outcome: employment			Outcome: deregistration			
	classical matching	correction term	adjusted matching	classical matching	correction term	adjusted matching	nr of observ.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<i>Treated: 24 years old registering in 1992:8-12, YP in first 6 months</i>							
<i>Controls: ineligibles (registration 1991:8-12) and eligible (registration 1992:8-12) non-treated</i>							
(1)	-0.057*	0.069	-0.127	-0.080*	0.049	-0.129	765
	(0.020)	(0.100)	(0.101)	(0.024)	(0.103)	(0.105)	
<i>Treated: 24:7-12 years old registering in 1992:8-12, YP in first 6 months</i>							
<i>Controls: ineligibles (registration 1991:8-12) and eligible (registration 1992:8-12) non-treated</i>							
(2)	-0.021	0.100	-0.131	-0.033	-0.270	0.237	171
	(0.035)	(0.223)	(0.229)	(0.039)	(0.211)	(0.214)	
<i>Treated: 20-24 years old registering in 1992:8-12, YP in first 6 months</i>							
<i>Controls: ineligibles (registration 1991:8-12) and eligible (registration 1992:8-12) non-treated</i>							
(3)	-0.035*	0.078*	-0.112*	-0.065*	0.155*	-0.220*	8,972
	(0.010)	(0.011)	(0.011)	(0.007)	(0.024)	(0.024)	
<i>Treated: 20-24 years old registering in 1992:8-12, YP in first 6 months</i>							
<i>Controls: eligible (registration 1992:8-12) non-treated</i>							
(4)	-0.015*	0.118*	-0.133*	-0.044*	0.177*	-0.223*	8,972
	(0.007)	(0.026)	(0.026)	(0.008)	(0.025)	(0.024)	

Notes: Estimates for males only. All estimates based on sample of new registrations with the employment office during 1992:8-12 and 1991:8-12. "Treatment" stands for flowing into YP within 6 months of registering with employment office as first destination after registration. Row 1 uses observations on 24 years old at registration. Row 2 further restricts sample to men at least 6 months into their 24th year. Row 3 extends the sample to all those aged 20 to 24 and row 4 restricts the control group used in standard atching to include eligible non-treated only. The impact of treatment is estimated on the probability of employment (columns 1 to 3) or deregistering (columns 4 to 6) 12 months after registration. Details on the contents of each colmn are identical to those of table 3 and can be found in footnote 19.

* Statistically different from zero at 5% significance level.