

Assignment vs. choice: lessons from training vouchers*

Friedrich Poeschel
IAB Institute for Employment Research

August 2012

-preliminary-

Abstract

For instrumental-variable estimation using binary instruments, we offer simple methods to obtain policy-relevant insights beyond local average treatment effects. We demonstrate the methods by evaluating an element of choice introduced into active labour market policy in Germany by a reform in 2003. Instead of being assigned to training courses by caseworkers, unemployed job seekers receive vouchers allowing them to choose among approved courses. Our data record the receipt of a voucher and the participation in a course. We use exogenous variation in treatment probabilities across locally integrated labour markets as an instrument. Applying our theoretical methods, we can make statements on the distribution of treatment effects and thereby infer the priorities of caseworkers. We further examine the consequences of choice for policy effectiveness and finally explore the distributions of gains or losses from choice.

JEL Classification Numbers: H43, J68

Key words: vouchers, training, choice, local average treatment effect

*I am grateful to Joseph Altonji, Bernd Fitzenberger, Thomas Kruppe, and Gesine Stephan for helpful comments. I am indebted to Annabelle Doerr and Anthony Strittmatter who let me use their programming codes for data cleaning and variable definition. All errors are my own. Correspondence: friedrich.poeschel@iab.de

1 Introduction

Basic economic intuition suggests that the use of vouchers in public policy is beneficial: they allow individuals to exercise choice and thereby account for their specific situation. In particular, individuals who expect little from participating in some treatment can opt out. At the same time, an individual's expectation may be unfounded, or exercising choice may be associated with costs. Then also individuals opt out who would benefit a lot from treatment, and a well-informed caseworker might have made a better choice.

While vouchers in the provision of social services, especially in education and training, have been intensely debated since they were advocated by Friedman (1955), there is very little evidence on the 'pure' effect of vouchers, i.e. the effect of choice by individuals on the efficacy of policy. The lack of evidence does not reflect a lack of interest in this fundamental question for policy making. In fact, the potential benefit from individual choice features prominently in both public and academic debates despite scant evidence. Rather, it has proven very difficult to identify a pure voucher effect separately from a host of other effects.

In the case of school vouchers, the effect of school choice on student achievement typically cannot be separated from peer effects. The study by Bettinger et al. (2010) is an exception. Using data from a rare institutional setting, they find a positive voucher effect net of peer effects. In education as much as in further education, voucher effects are hard to separate from the effect of greater competition between providers of education who have to attract voucher holders. Gibbons et al. (2008) claim that of the very few studies seeking to distinguish these two effects empirically, they are the first to obtain an estimate of the pure voucher effect (which they call 'direct choice effect'). They report only weak evidence for the effect of choice on student achievement.

The aim of this paper is to examine the issue of choice vs. assignment from two angles. We firstly propose methods to detect the priorities pursued by the caseworkers who can hand out vouchers. This sheds light on the choices the caseworkers would make if they could assign individuals, and how well this assignment would be in line with the intention of the policy. We secondly propose methods to examine the consequences of choice by individuals for the effectiveness of a policy, both in terms of average and total effects. By combining the two angles, we thirdly derive insights on how these consequences of choice change with the scale of the policy.

Like other recent studies (such as Gibbons et al. (2008), Lavy (2010), Frölich and Lechner (2010)), the empirical implementation of our approach relies on exogenous variation in otherwise homogeneous geographical units. However, our approach is apparently the first to systematically exploit the difference between intention to treat and actual treatment in the context of vouchers, i.e. between the allocation of a voucher and the actual participation. Our data derive from vouchers for further education that are given to unemployed individuals as part of active labour market policy (ALMP) in Germany.

Using detailed administrative information on the receipt of vouchers and their take-up, we can thus examine a number of hypotheses and make inference on the consequences of using vouchers in policy.

There is thus far little evidence on the consequences of vouchers in ALMP (see Barnow 2009). A study by O’Neill (1977) on the GI bill constitutes a notable exception: he finds strong treatment effects but no evidence that only a favourable selection of individuals uses the voucher. In fact, vouchers in ALMP appear particularly suited for the analysis of treatment effects because peer effects are likely absent here. Instead, recent literature has relied on lotteries to examine the role of choice. Cullen et al. (2006) consider school choice lotteries and find little benefit from choice, especially for weaker students. Hastings et al. (2006) estimate the effect of obtaining one’s first choice in such a lottery, and they conclude that parents’ priorities determine whether this raises school achievement. Methodologically, this paper is related to Imbens and Rubin (1997) whose methods deliver distributions of treatment outcomes and non-treatment outcomes, but cannot deliver a distribution of treatment effects. Our methods seek to make a first step towards a distribution of treatment effects and to a distribution of consequences from choice.

The paper proceeds as follows. Section 2 provides details on the reform that introduced vouchers into German ALMP. The section also explains why the few other papers analysing this reform are unlikely to estimate the pure voucher effect. Section 3 develops our empirical methods in the context of instrumental-variable estimation with a binary instrument. After describing our data set in section 4, our estimation techniques and results are reported in section 5. Finally, section 6 explores the role of search costs for our results before section 7 offers some conclusions.

2 Background: training vouchers for the unemployed

The public provision of training for unemployed was reformed in Germany on January 1st, 2003. Before the reform, caseworkers of the Federal Employment Service had assigned training in the form of various courses to unemployed individuals, who were then expected to participate. The training was delivered by third-party (private) providers selected by the local office of the Federal Employment Service, and course fees were reimbursed.¹ After the reform, caseworkers could only assign vouchers for training to unemployed individuals, while the recipients themselves had to decide whether or not to use the voucher. Caseworkers still specified the kind of training that the voucher could be used for, but would not be allowed to suggest a particular provider. Instead, the Federal Employment Service maintained a database of courses offered by approved providers for voucher holders to choose from. Vouchers still not used after a pre-specified time (up to three months, typically depending on the kind of intended training) became invalid.²

¹See Bruttel (2005) for details of pre-reform practices and the political context of the reform.

²Schneider et al. (2006) discuss the post-reform arrangements in greater detail.

Any attempt to quantify the effect of this reform has to deal with the issue of selection: both the unemployed who receive vouchers and those who use them might exhibit particular observable and unobservable characteristics. Let us first consider caseworker's decisions. Instead of giving vouchers to unemployed who stand to benefit most from training, caseworkers might give vouchers to unemployed with good employment prospects (known as "creaming") in order to obtain placement statistics that reflect positively on ALMP and the caseworkers' decisions.³ Indeed, as part of the reform, caseworkers were instructed explicitly to limit vouchers to unemployed who would have at least a 70% chance of being employed within six months after the training. As reported extensively in Schneider et al. (2006), however, it proved difficult to apply this rule because estimates of employment chances could not be based on sufficient data or because hardly anyone in certain regions would then qualify. It appears that different local offices of the Federal Employment Service have thus reacted differently to this rule, so that the probability of receiving a voucher varies substantially across regions. The rule was accordingly abolished altogether after two years (see Kruppe (2009)).

Despite such difficulties, surveys among caseworkers, their immediate superiors, or unemployed individuals have repeatedly found that voucher recipients were strongly selected (see for example Schneider et al. (2006), Doerr and Kruppe (2012)).⁴ Such behaviour by caseworkers has been analysed in only very few papers. Bell and Orr (2002) find that caseworkers in a welfare-to-work programme select participants mainly by their education, work experience, and unobserved characteristics. While some evidence is reported that caseworkers would apparently like to engage in creaming, Bell and Orr (2002) conclude that caseworkers are very poor at predicting their clients' treatment effects. Similarly, Lechner and Smith (2007) first confirm that caseworkers in Switzerland use their discretion and systematically select the participants of ALMP. They then calculate counterfactual treatment effects for possible alternative allocations of the unemployed and conclude that the assignment chosen by caseworkers is overall no more effective than random assignment.

Turning to the decisions of voucher holders, a key concern is that voucher recipients self-select into actual training in a way that partly undermines the policy's objectives. For ALMP in general and for the German training vouchers in particular, Weber (2008) and Kruppe (2009) have respectively found that low-skilled unemployed are less likely to actually enroll in the training. Unobservable characteristics are likely to also play an important role. In particular, if one assumes that voucher holders weigh the costs and benefits of using it, low-skilled or otherwise disadvantaged unemployed might have systematically lower expected benefits or higher costs from using vouchers. For example, they might find it harder to collect and assess information on providers and available

³Heckman et al. (1997) and Bell and Orr (2002) discuss this issue and emphasise its importance.

⁴Given the strong selection by caseworkers, training providers might have little to gain from further selecting participants, although they face in principle similar incentives to increase their success rates.

courses as a basis for their decision.⁵ Indeed, surveys conclude that voucher holders might well have trouble using the said database of approved courses (see Schneider et al. (2006), Doerr and Kruppe (2012)). On the other hand, mandatory information of voucher holders can reduce the take-up rate further (see Messer and Wolter (2009)). We will return to the issue of search costs in section 6.

Beyond selection, further issues complicate evaluating the introduction of training vouchers. While Schneider and Uhlendorff (2006) and Rinne et al. (2008) offer evaluations of this reform, they essentially compare the situations before and after the reform. There are three problems with this approach. Firstly, change-over difficulties occurred initially after the reform and affected the provision of vouchers (see Schneider et al. (2006)), so that a reform effect measured close to January 2003 would be a poor guide to effects thereafter. If one instead measures a reform effect much later than January 2003, then, secondly, other yet simultaneous changes in the labour market will likely bias the results. Even if the environment was stable over time, it would still seem hard to identify the effect of vouchers separately from the consequences of other elements of the same reform: the new selection rule (70% chance) or other changes in selection, adjustments among training providers, and changes in the programme’s objectives and scope.⁶ As a more subtle problem, thirdly, vouchers were used much less frequently after the reform (see Kruppe (2009)). Hence even if participants were selected after the reform by the same criteria as before the reform, the (observable and unobservable) characteristics of the marginal participants would likely differ substantially and thus give rise to bias. Our results below should be largely unaffected by these three problems, as we only use measurements after the reform.

3 Empirical methods

3.1 Model

We cast our model in the notation of potential outcomes. The only agents are caseworkers who implement ALMP and their clients who are currently unemployed. The outcome of interest Y is whether or not a client is (permanently) employed at a specific time after a treatment. The treatment can in our case be defined as receiving a voucher (denoted $D^V \in \{0, 1\}$) or as using the voucher to participate in training (denoted $D^T \in \{0, 1\}$). D^V and D^T respectively equal 1 if a treatment occurs. The observed outcome Y can be

⁵While the Federal Employment Service reimburses in advance such costs as child care and travel when incurred during a course, similar costs can already arise during the search for a suitable course.

⁶Schneider et al. (2006) report a greater emphasis on formal qualifications as an objective after the reform. While Rinne et al. (2008) attempt to account for changes in selection, they can only do this for selection on observables and define the voucher effect simply as the residual. Yet a voucher effect thus defined might be confounded with selection on unobservables or the effects of other reform elements.

seen as depending on either treatment:

$$Y = D^V Y_1^V + (1 - D^V) Y_0^V = D^T Y_1^T + (1 - D^T) Y_0^T$$

where Y_1^V and Y_0^V denote the observed outcomes categorised by whether or not a voucher is received, and similarly for Y_1^T and Y_0^T . One might want to know the individual treatment effect $Y_1^V - Y_0^V$, but only one of Y_1^V and Y_0^V is observed for any given client, and the same problem applies to $Y_1^T - Y_0^T$.

While the caseworker decides whether treatment D^V occurs, the client decides herself whether she uses the voucher so that treatment D^T occurs. We model the empirical behaviour of caseworkers and clients in an extended latent variable framework:

$$Y = 1[\lambda^j(X, D^j, \varepsilon) \geq 0], \quad j \in \{V, T\} \tag{1}$$

$$D^V = 1[D^{*V} \geq 0] = 1[\kappa^V(X, Z) - U^V \geq 0] \tag{2}$$

$$D^T = 1[D^{*T} \geq 0] D^V = 1[\kappa^T(X, Z) - U^T \geq 0] D^V \tag{3}$$

where X is a vector of observed random variables, Z is a scalar observed random variable, and (U^V, U^T, ε) are unobserved scalar random variables. We multiply by D^V in equation (3) to ensure that $D^T = 0$ whenever $D^V = 0$, for the simple reason that training is only available after the receipt of a voucher. Moments $E[Y]$, $E[Y_0^j]$, and $E[Y_1^j]$ for $j \in \{V, T\}$ are assumed to exist, as well as $\Pr(D^j = 1 > 0)$ so that some clients are in fact treated. For the core of our analysis, we do not make parametric assumptions on (U^V, U^T, ε) or on any of the functions $\lambda^j(\cdot)$ and $\kappa^j(\cdot)$, $j \in \{V, T\}$. In particular, as we do not assume additive separability in equation (1), treatment effects may vary across observably identical individuals. While we assume $(X, Z) \perp\!\!\!\perp (U^V, U^T, \varepsilon)$, the framework still allows for possible endogeneity of D^j : if $\text{cov}(U^j, \varepsilon) \neq 0$ (and variances are finite), estimates for the effect of D^j on Y will be biased even when controlling for X .

In the specific context of our model, D^{*V} in equation (2) may be interpreted as the client's employability that depends on X and on unobservable ability U^V : the more employable a client, the more likely the caseworkers will decide in favour of a voucher. As another possibility, D^{*V} may be the gain from training anticipated by the caseworker. Correspondingly, D^{*T} in equation (3) might be interpreted as the net gain from training anticipated by the client, which might depend on unobservable costs U^T of searching for a suitable course: the greater the gross gain relative to such costs, the more likely the client will in fact attend a course. In all these examples, $\text{cov}(U^j, \varepsilon) \neq 0$ appears well possible, so that endogeneity of D^j cannot be ruled out at all. This will be the case when the most employable clients (with high ability captured by low U^V , say) receive vouchers, as these clients will find employment more quickly than observable variables would suggest (high ε). Then a simple conditional mean over all clients $E[Y_1^j - Y_0^j | X]$ would overestimate the average treatment effect (ATE) by falsely attributing the effect of unobserved ability to

a treatment.

3.2 LATE and limitations

In response to the concerns about endogeneity of D^j , we adopt an instrumental-variable (IV) approach. For the instrument Z in the model, we use the geographic variation in the likelihood of receiving a voucher. As argued in section 2, this variation appears to be exogenous and therefore a suitable instrument. Concretely, we will consider a well-integrated local labour market, where labour market conditions are roughly comparable everywhere in the market, as the logic of arbitrage would suggest. Within this local market, Z varies because the market is (exogenously) divided into several districts for ALMP, some of whom use vouchers much more intensively than others. Given the small number of districts in a local market, Z will only take few values. Therefore, we will focus throughout on binary comparisons between a market side with high probability of receiving a voucher ($Z = z_H$) and a market side with low probability ($Z = z_L$, with $z_H > z_L$).

The correlation of an effectively binary instrument Z with D^{*j} will be too weak to identify ATE, but the correlation is perfect for those clients where the (potential) change in Z lifts $\kappa^j(X, Z)$ above the threshold U^j and thus leads to a change in D^{*j} . Thus the use of Z allows us to identify the local ATE (LATE) for this group of compliers, as shown by Imbens and Angrist (1994). To estimate LATE in the context of our non-parametric model with covariates X , we employ the methods developed by Frölich (2007). Although it is not essential to control for X , given that we assume $Z \perp\!\!\!\perp \varepsilon$, we hope to thereby obtain more robust estimates. To see this, note first that the estimated LATE might be based on few observations, because the values of X will for most clients push $\kappa^j(X, Z)$ too far below or above the threshold U^j for Z to be pivotal. After controlling for X , however, Z might be pivotal also for some of these clients.

When LATE is estimated, a client's type θ^j is categorised into one of four groups: compliers ($\theta^j = c^j$), always-participants ($\theta^j = a^j$), never-participants ($\theta^j = n^j$), and defiers ($\theta^j = d^j$). While D^{*j} for a complier switches from 0 to 1 in response to an increase in Z from z_L to z_H , D^{*j} for a defier switches from 1 to 0. Always-participants and never-participants are unaffected by the change in Z . As in Frölich (2007), the following assumptions collectively ensure that a non-empty group of compliers captures all clients whose D^{*j} changes in response to a change in Z from z_L to z_H : for $j \in \{V, T\}$,

- (i) (*Monotonicity*) defiers do not exist: $\Pr(\theta^j = d^j) = 0$
- (ii) (*Existence of compliers*) there is a strictly positive proportion of compliers:
 $\Pr(\theta^j = c^j) > 0$
- (iii) (*Unconfounded type*) the proportions of existing types are independent of Z :

for all $x \in \text{Supp}(X)$,

$$\Pr(\theta^j = \check{\theta} | X = x, Z = z_L) = \Pr(\theta^j = \check{\theta} | X = x, Z = z_H) \quad \text{for } \check{\theta} \in \{a^j, n^j, c^j\}$$

(iv) (*Exclusion restriction*) the instrument Z is exogenous: for all $x \in \text{Supp}(X)$,

$$E[Y_0^j | X = x, Z = z_L, \theta^j = \check{\theta}] = E[Y_0^j | X = x, Z = z_H, \theta^j = \check{\theta}] \quad \text{for } \check{\theta} \in \{n^j, c^j\}$$

$$E[Y_1^j | X = x, Z = z_L, \theta^j = \check{\theta}] = E[Y_1^j | X = x, Z = z_H, \theta^j = \check{\theta}] \quad \text{for } \check{\theta} \in \{a^j, c^j\}$$

(v) (*Common support*) X has the same support on both market sides:

$$\text{Supp}(X | Z = z_L) = \text{Supp}(X | Z = z_H)$$

A discussion of these assumptions in some detail is offered by Frölich and Lechner (2010). Based on these assumptions, Frölich (2007) shows that LATE for our non-parametric model is given by

$$\Delta^j(z_H, z_L) = E(Y_1^j - Y_0^j | \theta^j = c^j) = \frac{\int (E(Y | X, Z = z_H) - E(Y | X, Z = z_L)) dF_X}{\int (E(D^j | X, Z = z_H) - E(D^j | X, Z = z_L)) dF_X} \quad (4)$$

where $\Delta^j(z_H, z_L)$ denotes the LATE of treatment $j \in \{V, T\}$ estimated with a change from $Z = z_L$ to $Z = z_H$, without making the conditionality on X explicit. Frölich and Lechner (2010) apply this result and also point out that four potential outcomes are similarly identified from the results in Frölich (2007). In particular,

$$E(Y_1^j | \theta^j = c^j) = \frac{\int [E(YD^j | X, Z = z_H) - E(YD^j | X, Z = z_L)] dF_X}{\int [E(D^j | X, Z = z_H) - E(D^j | X, Z = z_L)] dF_X} \quad (5)$$

and the expression for $E(Y_0^j | \theta^j = c^j)$ is obtained by substituting $(1 - D^j)$ for D^j in the numerator of equation (5). Next,

$$E(Y_1^j | \theta^j = a^j) = \frac{\int E(YD^j | X, Z = z_L) dF_X}{\int E(D^j | X, Z = z_L) dF_X} \quad (6)$$

and, by substituting $(1 - D^j)$ for D^j and z_H for z_L everywhere in equation (6), an expression for $E(Y_0^j | \theta^j = n^j)$ is obtained. Equations (4) through (6) use only observable information and can thus be estimated.

However, the conclusions that may be drawn from LATE more generally and from the results above in particular are limited and leave a number of important questions open. For example, in their application of these results to Swiss data on ALMP, Frölich and Lechner (2010) find the following average results over 18 local labour markets they consider (see their Table 5):

$$E(Y_1 | \theta = c) > E(Y_0 | \theta = n) > E(Y_1 | \theta = a) > E(Y_0 | \theta = c) \quad (7)$$

In words, they find that the never-participants have better employment chances ex ante than the compliers ($E(Y_0|\theta = n) > E(Y_0|\theta = c)$) while the compliers have better employment chances ex post than the always-participants ($E(Y_1|\theta = c) > E(Y_1|\theta = a)$). As an interpretation of this ordering, Frölich and Lechner (2010) suggest that caseworkers prioritise those unemployed with the worst employment chances ex ante, who thus typically become always-participants, while the never-participants do not need any assistance to find employment and the compliers make up an intermediate group.

While such insights would, in our view, be highly relevant and valuable for ALMP and for programme evaluation more generally, let us examine whether the results in equation (7) put us in a position to draw such a conclusion. Thus suppose that caseworkers prioritise the unemployed with the worst employment chances ex ante, so that in particular $E(Y_0|\theta = a) < E(Y_0|\theta = c)$, even though these unemployed overall benefit less from training than others, i.e.

$$E(Y_1|\theta = c) - E(Y_0|\theta = c) > E(Y_1|\theta = a) - E(Y_0|\theta = a)$$

However, let us now consider a very different possibility: the caseworkers prioritise those unemployed who overall benefit most from training, so that

$$E(Y_1|\theta = c) - E(Y_0|\theta = c) < E(Y_1|\theta = a) - E(Y_0|\theta = a) \tag{8}$$

Yet with $E(Y_1|\theta = a) < E(Y_1|\theta = c)$ from equation (7), equation (8) again implies $E(Y_0|\theta = a) < E(Y_0|\theta = c)$. Hence these two contrasting interpretations both fit the results in equation (7). To distinguish among them, one would have to know by how much $E(Y_0|\theta = c)$ exceeds $E(Y_0|\theta = a)$, yet the latter is not observed. Therefore, reliable conclusions about caseworkers' behaviour cannot be reached this way.⁷

3.3 Comparing local effects: extensive margin

Based on simple comparisons between measurements of LATE, this section proposes a method for distinguishing between alternative hypothesis about caseworkers' behaviour (and we thus focus on treatment D^V). Let us describe their behaviour in terms of an approximate rule for every remaining voucher they can assign (assuming that they cannot give vouchers to everyone). One imagines that caseworkers rank their clients by some criterion and then hand out the first voucher to the top-ranked client, the second voucher to the second-ranked client and so on until vouchers have run out. To a null hypothesis and rankings by treatment effect and by need, as suggested by Lechner and Smith (2007), we add the ranking implicit in creaming as a third possibility. The effects and outcomes

⁷That apart, the interpretation by Frölich and Lechner (2010) surprisingly suggests that ALMP increases ex-ante mediocre employment chances of compliers ($E(Y_0|\theta = c) < E(Y_0|\theta = n)$) to the best ex-post employment chances of all groups ($E(Y_1|\theta = c) > E(Y_0|\theta = n)$), i.e. better chances than those who did not seem in need of treatment.

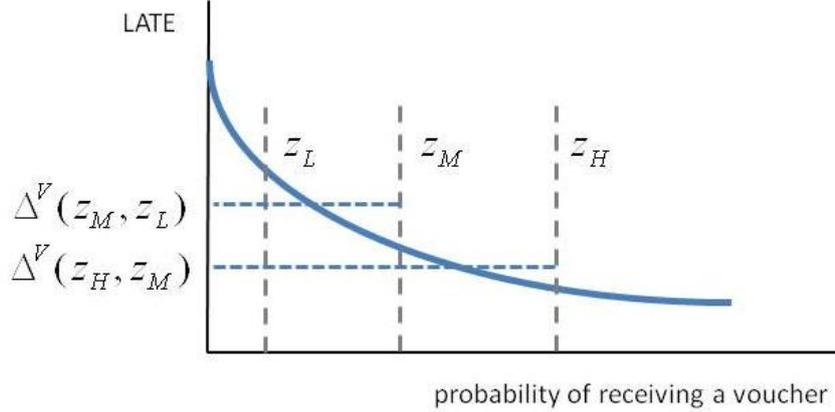


Figure 1: Measurements of LATE under effect-based ordering

considered here are all unconditional because it seems hard to imagine that caseworkers calculate conditional expectations before they make their decisions.

- (R) (*Random ordering*) Caseworkers assign vouchers at random.
- (E) (*Effect-based ordering*) Caseworkers tend to assign vouchers in descending order of the expected treatment effect $E[Y_1^V - Y_0^V]$.
- (N) (*Need-based ordering*) Caseworkers tend to assign vouchers in ascending order of the non-treatment outcome $E[Y_0^V]$.
- (S) (*Cream-skimming*) Caseworkers tend to assign vouchers in descending order of the treatment outcome $E[Y_1^V]$.

Let us reconsider a locally integrated labour market where the likelihood of receiving a voucher differs exogenously. However, suppose we can now delimit *three* geographically separate areas according to this likelihood: unemployed in area H have the highest probability of receiving a voucher ($Z = z_H$), while those in area M and L have medium ($Z = z_M$) and low probability ($Z = z_L$), respectively. We assume that caseworkers in all areas assign training according to roughly the same criteria, yet at exogenously varying intensity. This assumption is motivated by idiosyncratic reactions of local offices of the Federal Employment Service to the new selection rule (70% chance): in the absence of easily implementable guidelines, different offices chose different internal rules of thumb. Then effect-based ordering implies that the marginal recipient of a voucher in area H will likely benefit less from training than the marginal recipient in area M, who in turn will likely benefit less than the marginal recipient in area L. This gives us a monotonically downward-sloping curve for the LATE as depicted in figure 1.

We can now calculate one LATE using the exogenous variation in Z between $Z = z_H$ and $Z = z_M$ and another using the variation between $Z = z_M$ and $Z = z_L$. We denote these two treatment effects by $\Delta^V(z_H, z_M)$ and $\Delta^V(z_M, z_L)$, respectively. Each can be

identified in exact analogy to equation (4). The downward-sloping curve we expect under hypothesis (E) then implies

$$\underline{\Delta}^V \equiv \Delta^V(z_M, z_L) - \Delta^V(z_H, z_M) > 0 \quad (9)$$

as indicated in figure 1. This implication provides a very first test of hypothesis (E): if estimation results do not conform with equation (9), this will constitute evidence against hypothesis (E).

For reliable inference, however, one would also want to know whether a given estimate $\hat{\underline{\Delta}}^V$ is statistically significant. It seems that a standard two-sample t-test between population means (using the Welch approximation to allow for different population variances) can be applied here, as we will explain below that the prerequisites of this test appear to be met approximately by the situation at hand. The test statistic for testing the hypothesis $H_0 : \underline{\Delta}^V \leq 0$ against $H_1 : \underline{\Delta}^V > 0$ is then

$$t_0 = \frac{\hat{\underline{\Delta}}^V}{\sqrt{s_{\Delta^V(z_M, z_L)}^2 + s_{\Delta^V(z_H, z_M)}^2}} \approx t_{(df)} \quad (10)$$

where $s_{\Delta^V(z_M, z_L)}^2$ denotes the sample variance of $\Delta^V(z_M, z_L)$ and df is a placeholder for the degrees of freedom that will reflect the Welch approximation. If H_0 can be rejected at a high significance level, we will conclude that a positive difference between $\Delta^V(z_M, z_L)$ and $\Delta^V(z_H, z_M)$ appears statistically significant.

The standard two-sample t-test between some means \bar{x}_1 and \bar{x}_2 would employ a test statistic $(\bar{x}_1 - \bar{x}_2)/\sqrt{s_{\bar{x}_1}^2 + s_{\bar{x}_2}^2}$ for testing H_0 and H_1 . It requires that \bar{x}_1 and \bar{x}_2 be sample means from two independent random samples that are sufficiently large and are drawn from the two populations whose means are to be compared. The test (using the Welch approximation) seems roughly applicable to our framework for the following reasons. Equation (4) defines both $\Delta^V(z_M, z_L)$ and $\Delta^V(z_H, z_M)$ as conditional means ($s_{\Delta^V(z_M, z_L)}^2$ and $s_{\Delta^V(z_H, z_M)}^2$ thus being the variances of means). The relation between these means is characterised by Heckman and Vytlačil (1999, 2000). Under mild additional assumptions,⁸ the probability transform \tilde{U}^V of U^V is uniformly distributed and one can write

$$(\Pr(z_H) - \Pr(z_L))\Delta^V(z_H, z_L) = \int_{\Pr(z_L)}^{\Pr(z_H)} E[Y_1^V - Y_0^V | X = x, \tilde{U}^V = u^V] du^V \quad (11)$$

where $\Pr(z)$ denotes $\Pr(D^V = 1 | Z = z)$. Let us next split up the intergral: for some z_M with $\Pr(z_L) < \Pr(z_M) < \Pr(z_H)$, equation (11) may be written down for the interval $z_M - z_L$ and for $z_H - z_M$. Since these equations' right-hand sides add up to that of

⁸Conditional on X , $\kappa^V(X, Z)$ is assumed a non-degenerate random variable and, with respect to the Lebesgue measure on R^2 , (U^V, ε) is assumed absolutely continuous.

equation (11), the same must hold for the left-hand sides, so that

$$\Delta^V(z_H, z_L) = \frac{\Pr(z_M) - \Pr(z_L)}{\Pr(z_H) - \Pr(z_L)} \Delta^V(z_M, z_L) + \frac{\Pr(z_H) - \Pr(z_M)}{\Pr(z_H) - \Pr(z_L)} \Delta^V(z_H, z_M) \quad (12)$$

which is a weighted average, as noted in Heckman and Vytlacil (2000). Equation (13) clarifies that $\Delta^V(z_M, z_L)$ and $\Delta^V(z_H, z_M)$ are, while directly comparable, calculated for two different groups of compliers. Recall that the compliers used for $\Delta^V(z_M, z_L)$ are clients who do not receive (use) a voucher when $Z = z_L$ but do when $Z = z_M$; by contrast, the compliers for $\Delta^V(z_H, z_M)$ do not receive (use) a voucher when $Z = z_M$ but do when $Z = z_H$. Hence these two groups of compliers do not overlap, so that we can define them as two populations from which all random samples will be independent.

Further, as LATE is calculated for the entire group of compliers, $\hat{\Delta}^V(z_M, z_L)$ and $\hat{\Delta}^V(z_H, z_M)$ can be regarded as sample means based on the entire population, which is a random sample. That the groups of compliers are sufficiently large can be ensured by an appropriate definition of market sides. Apparently the only remaining discrepancy between the test's prerequisites and our framework is the source of variance: $\hat{\Delta}^V(z_M, z_L)$ does not vary around the true LATE because it is only based on a small random sample, but rather because it is estimated through IV methods. However, given the perfect correlation between Z and D^{*V} for the group of compliers, we would conjecture that the variance of $\hat{\Delta}^V(z_M, z_L)$ is similar to the variance of a mean calculated directly from the sample, provided the group of compliers is sufficiently large.

Finally, by essentially the same logic as above, hypothesis (N) about caseworkers' priorities implies

$$E[Y_0^V | \theta^V = c_{HM}^V] - E[Y_0^V | \theta^V = c_{ML}^V] > 0 \quad (13)$$

where c_{HM}^V (c_{ML}^V) refers to the compliers on market sides H and M (M and L). Similarly, hypothesis (S) implies

$$E[Y_1^V | \theta^V = c_{HM}^V] - E[Y_1^V | \theta^V = c_{ML}^V] < 0 \quad (14)$$

Note that these two equations again use means defined over the group of compliers. Hence the same arguments that motivated the test in equation (10) suggest that exactly analogous tests can be used to assess whether the estimated mean differences in treatment and non-treatment outcomes are statistically significant.

In conclusion, this section proposes simple methods that derive policy-relevant insights from comparing two measurements of LATE. Under the monotonicity assumption implicit in hypotheses (E), (N), and (S), such comparisons tell us whether treatment effects and outcomes for always-participants or never-participants are higher or lower than those for compliers. For example, if a downward-sloping curve of treatment effects cannot be rejected, this indicates that the effects for always-participants are higher, while those for never-participants are lower. This kind of insight on the behaviour of caseworkers

and the distribution of treatment effects cannot be derived from just one measurement of LATE: a single point can belong to a line with arbitrary slope, so that there is no way of distinguishing between a downward-sloping or upward-sloping curve. We have demonstrated this shortcoming above for the case of Frölich and Lechner (2010) where the distinction of two geographical areas only permits one measurement of LATE.⁹

3.4 Comparing local effects: intensive margin

Let us next draw comparisons between the LATE of receiving a voucher ($D^j = D^V$) and the LATE of attending the course ($D^j = D^T$). Rather than considering different treatment probabilities as in the previous section, we now examine the LATE for a given probability but for different treatments. Their effects differ because the recipients of vouchers decide themselves whether to attend a course, and some thus do not undergo any training.

There might be many reasons why some clients do not use their voucher: high costs of finding a provider, low expected benefit from training, or high opportunity costs that arise because clients have less time for job search when ‘locked’ in a course. Equation (3) is general enough to capture all these possibilities: clients choose not to use their voucher when the expected benefit $\kappa^T(X, Z)$ is sufficiently low (possibly negative) and/or U^T is sufficiently high. Similarly, one can think of U^T as composite error including noise that is always observed together with $\kappa^T(X, Z)$, so that clients’ perception of their benefit differs from the statistical expectation of their benefit. However defined, U^T can thus outweigh even relatively large values of $\kappa^T(X, Z)$.

By consequence, it is not clear a priori which individual treatment effects would have been realised by the clients who do not use the voucher, compared to those clients who do. If there is any systematic difference between unrealised and realised treatment effects, ATE for voucher users will differ from ATE for voucher recipients. In addition to ATE, let us also define the total treatment effect (TTE) as the sum of all treatment effects, i.e. ATE multiplied by the number of clients. This measure will change with the number of voucher users even if ATE is unaffected. For both ATE and TTE, we consider here three hypotheses: compared to all voucher recipients,

- (G) (*Gains dominate*) the average (total) treatment effect for voucher users is greater
- (L) (*Losses dominate*) the average (total) treatment effect for voucher users is smaller
- (B) (*Gains and losses balance*) the average (total) treatment effect for voucher users is roughly equal.

Which of these hypotheses applies likely determines the effectiveness of a voucher policy, relative to assignment of courses by caseworkers (i.e. as if all vouchers had to be used).

⁹It is worth noting that, with a fully parametric model, one can also employ the two measurements of LATE to obtain an estimate of ATE, as explained in Oreopoulos (2006).

Let us first turn to average effects. In case (G), for example, the voucher policy is more cost-effective than assignment of courses would be: given that the costs of a course are incurred, the associated treatment effect is on average higher using vouchers than using assignment. Under hypothesis (L), in turn, assignment would be more cost-effective.

However, LATE instead of ATE is identified in our non-parametric set-up, so that all comparisons will be made between LATE of voucher users and LATE of voucher recipients. In other words, we will compare the average treatment effects between two groups of compliers defined for different treatments: between $\theta^T = c^T$ and $\theta^V = c^V$. For these comparisons to be informative, the clients with $\theta^T = c^T$ have to be a weak subset of the clients with $\theta^V = c^V$. More precisely, these comparisons require

$$\Pr(\theta^V = a^V \cap \theta^T = c^T) = 0 \quad (15)$$

as LATE for the voucher users could change in any way relative to voucher recipients if it also counts some clients with $\theta^V = a^V$. (Other types cannot possibly be counted here because they do not receive a voucher.) To see that the standard assumptions above in fact imply equation (15), focus on the type $\theta^T = c^T$. These compliers, whose behaviour is governed by equation (3), switch to $D^T = 1$ where Z switches to $Z = z_H$ (provided $D^V = 1$). Yet this switch cannot be motivated by the benefit from training being higher when $Z = z_H$, as this would violate the exclusion restrictions in assumption (iv). Nor can the switch be motivated by U^T being lower when $Z = z_H$, as this would violate $Z \perp\!\!\!\perp U^T$. Rather, the reason why these compliers use the voucher only when $Z = z_H$ (and the reason why $\kappa^T(X, Z)$ depends on Z) is that they only receive a voucher when $Z = z_H$. This means that these compliers ($\theta^T = c^T$) also have $\theta^V = c^V$ and cannot have $\theta^V = a^V$. Since equation (15) therefore holds, hypothesis (G) would be supported by

$$\underline{\Delta}^{T-V}(z_H, z_L) \equiv \Delta^T(z_H, z_L) - \Delta^V(z_H, z_L) > 0 \quad (16)$$

while $\underline{\Delta}^{T-V}(z_H, z_L) < 0$ and $\underline{\Delta}^{T-V}(z_H, z_L) \approx 0$ would support hypotheses (L) and (B), respectively. Figure 2 depicts an example for equation (16). A policy of assigning courses would realise individual treatment effects from any part of the distribution, even if they are negative due to lock-in effects, for example. Under a voucher policy, negative treatment effects would not normally be realised, alongside positive treatment effects that are now outweighed by the costs of finding a provider, for example. In the case depicted, the unrealised effects below the distribution's mean $\Delta^V(z_H, z_L)$ dominate the unrealised effects above it, so that the (local) average effect for the voucher users $\Delta^T(z_H, z_L)$ lies above $\Delta^V(z_H, z_L)$. In this case, giving clients the choice whether to use their voucher has increased the policy's cost-effectiveness.

Comparisons between averages remain silent on changes in the number of treated clients and therefore on the policy's overall effectiveness. Hence let us now turn to comparisons between total treatment effects. We consider a social planner who chooses the

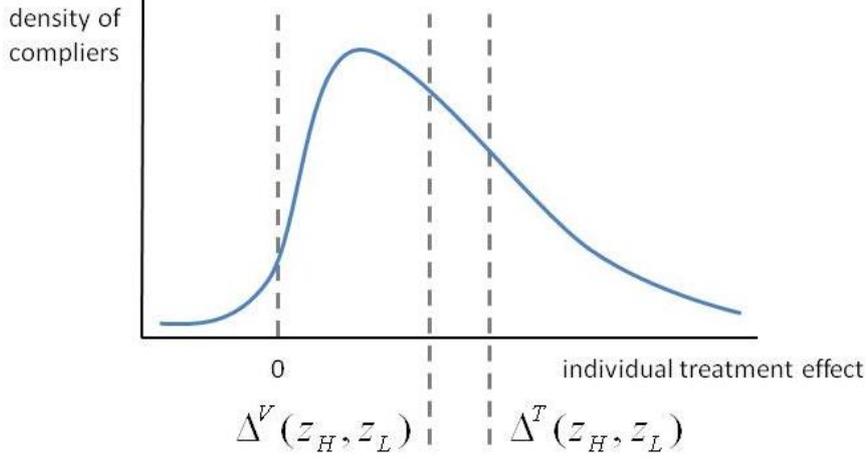


Figure 2: LATE for voucher receipt and for training attendance

policy with the greater total treatment effect:

$$\max \left[\sum_i D_i^V (Y_{1i}^V - Y_{0i}^V), \sum_i D_i^T (Y_{1i}^T - Y_{0i}^T) \right]$$

where i indexes individual clients, the discount rate is assumed negligible for simplicity, and the relevant effects are unconditional as before. The policy thus selected does not necessarily maximise welfare, as welfare comparisons would also have to include the costs incurred by clients and caseworkers under different policies. Since these costs are often very hard to measure, choosing the overall most effective policy may in practice be the best available path of action.

Since our set-up identifies the treatment effect $Y_1^j - Y_0^j$ only for the group of compliers, we focus first on the local TTE, i.e. the TTE for the compliers. To obtain this measure in analogy to TTE, we use LATE instead of ATE and the number of compliers instead of all clients. While the number of compliers is unknown, it can also be estimated: following Frölich (2007), the proportion of compliers is identified as

$$\Pr(\theta^j = c^j) = \int [E(D^j|X, Z = z_H) - E(D^j|X, Z = z_L)] dF_X, \quad j \in \{V, T\}$$

so that this proportion of all clients estimates the number of compliers. By equation (15), comparisons between the compliers for $D^j = D^V$ and the compliers for $D^j = D^T$ are informative. Hence, with regards to the local TTE, hypothesis (G) is equivalent to

$$\begin{aligned} \underline{\Delta}^{TT-TV}(z_H, z_L) &\equiv \Pr(\theta^T = c^T) I^T \Delta^T(z_H, z_L) \\ &\quad - \Pr(\theta^V = c^V) I^V \Delta^V(z_H, z_L) > 0 \end{aligned} \quad (17)$$

where I^j is the number of observations employed to calculate $\Delta^j(z_H, z_L)$, and accordingly for the other hypotheses. Under hypothesis (G), the social planner thus prefers vouchers

to assignment (at least for compliers) because the local TTE for voucher users exceeds the local TTE for voucher recipients. Hence the choice given to clients has led to an increase of the policy’s total effectiveness in this example, as clients with negative individual treatment effects choose not to use their voucher.

In conclusion, with data on both the receipt and the take-up of vouchers, we can make empirically-based statements on the role of vouchers for policy effectiveness. Concretely, this section has proposed methods to examine whether, for a subset of individuals, vouchers increase or decrease the policy’s cost-effectiveness and its overall effect.

3.5 Comparisons along both margins

The comparisons in the previous section suffer from the problem discussed earlier: they only make statements about one particular group of compliers and remain silent on always-participants, for example. To go a step further, the comparisons in this section involve both the intensive and the extensive margin. Recall from section 3.3 that one can just as well define three sides in a local labour market that differ along the extensive margin. This does not only permit two measurements of LATE, but likewise also two of the comparisons along the intensive margin we discussed in section 3.4. We thereby obtain some insights on the distribution of gains and losses from leaving the choice to clients. Let us consider the following hypotheses: the difference between the effects of training and the effects of vouchers, as a proportion of the latter,

(I) (*Increasing returns from choice*) increases along the extensive margin, i.e. with the number of treated individuals,

(D) (*Decreasing returns from choice*) decreases along the extensive margin

(C) (*Constant returns from choice*) does not change along the extensive margin

For example, hypothesis (D) implies for average effects that a proportional difference along the intensive margin is greater when fewer individuals receive vouchers:

$$\frac{\underline{\Delta}^{T-V}(z_M, z_L)}{\Delta^V(z_M, z_L)} > \frac{\underline{\Delta}^{T-V}(z_H, z_M)}{\Delta^V(z_H, z_M)} \quad (18)$$

provided $\underline{\Delta}^{T-V}(z_M, z_L) > 0$. For total effects, hypothesis (D) similarly implies

$$\frac{\underline{\Delta}^{TT-TV}(z_M, z_L)}{\Pr(\theta^V = c_{ML}^V)I^V \Delta^V(z_M, z_L)} > \frac{\underline{\Delta}^{TT-TV}(z_H, z_M)}{\Pr(\theta^V = c_{HM}^V)I^V \Delta^V(z_H, z_M)} \quad (19)$$

provided $\underline{\Delta}^{TT-TV}(z_M, z_L) > 0$. As all elements in equations (18) and (19) can be estimated, such comparisons would be a first step towards a forecast whether the gains (or losses) from choice will grow or decline as a policy is scaled up. In combination with insights from section 3.3, one also obtains an indication for which kind of client the gains

(or losses) from choice are greatest. Consider for example the case that equations (14) and (18) are both found to hold alongside $\underline{\Delta}^{T-V}(z_M, z_L) > 0$. This would suggest that clients with higher $E[Y_1^V]$ gain more from choice, possibly because choice allows them to avoid large adverse lock-in effects.

4 Data

Our empirical analysis draws on three large sets of individual level data. All data was supplied by the public Federal Employment Service (*Bundesagentur für Arbeit*) in Germany. The first set consists of data on the training vouchers that are awarded by caseworkers of the Service. These data cover the entire population of vouchers awarded from 2003 to 2005; in particular, they record when and where the voucher was awarded and whether it was eventually used. We can link these data to information on the voucher recipients available from integrated employment histories (data called IEB).¹⁰ These data on the clients of the Federal Employment Service record employment, unemployment benefit receipt, and participation in ALMP with daily frequency, along with demographic information. While the receipt of benefits requires registration with the Federal Employment Service, unemployed who do not register are not observed. Finally, we use a third data set to complement the demographic information in the IEB by additional information such as civil status, number of children, and health status.

Geographical information on a client's place of residence (from the IEB) allows us to select individuals belonging to a particular local labour market. For identification strategy, we seek a locally integrated labour market characterised by exogenous differences in the probability of receiving a voucher. The local labour market we choose is the Ruhr area, a densely populated area made up by a number of medium-sized cities. Due to coal mining, these cities have grown next to each other and have evolved into an almost unified urban conglomerate. Today the Ruhr area counts several million inhabitants and is well integrated by public transport and other infrastructure. While this area thus appears as an integrated labour market in the sense that its inhabitants face roughly the same labour market conditions, it is also split up into a dozen policy districts for ALMP, as shown in Figure 3 below.

Counts of the vouchers awarded in the Ruhr area's policy districts (available on request) reveal a similar pattern for 2004 and 2005: where the probability of receiving a voucher was high in 2004, it was also high in 2005. The overall number of vouchers awarded, however, is high only in 2003 and 2004. It thus appears that policy differences immediately after the reform in January 2003 did not persist and settled only in 2004. We therefore choose to further focus on the year 2004 when the overall number of vouchers is still high. Hence we reduce our data set to those individuals who live in the Ruhr area

¹⁰A negligible proportion of the vouchers cannot be linked up because of some data cleaning carried out in the generation of the IEB. We disregard these vouchers.



Figure 3: Labour market policy districts in North-Rhine Westfalia.
 Districts used in estimation: Bochum, Dortmund, Düsseldorf, Duisburg, Essen, Gelsenkirchen, Hagen, Oberhausen, Recklinghausen, Solingen, Wuppertal.
 As of 25/07/2012. Source: Bundesagentur für Arbeit

at some point in 2004. By dropping those individuals for whom the place of residence in 2004 is missing and cannot be easily inferred, we implicitly assume that those individuals are missing at random.

Table 1 below provides descriptive statistics on the vouchers awarded in the policy districts of the Ruhr area. The differences across districts in the probability of receiving a voucher (see voucher rate) are marked. Neither we nor internal sources of the Federal Employment Service are aware of any reason that would make these differences appear endogenous; to the contrary, it can be shown that observable differences between districts do not explain the differences in treatment probability. Instead, it appears to us that such differences have arisen in response to the vacuum left by an assignment rule for vouchers that could hardly be satisfied (see section 2). Such idiosyncratic policy choices would imply that the differences in treatment probability are exogenous, as is required for our identification strategy.

Table 1: Treatment probabilities across ALMP districts in 2004

District	id.	unempl. inflow	vouchers	voucher rate in%	unused vouchers	training	training rate in%
Bochum	321	6849	1190	17.6	150	1053	15.4
Dortmund	333	12372	1932	16.6	171	1750	14.1
Düsseldorf	337	7290	1173	16.0	96	1069	14.7
Duisburg	341	6199	1148	18.7	250	902	14.6
Essen	343	7699	1800	23.6	138	1674	21.7
Gelsenkirchen	345	7400	791	11.1	143	664	9.0
Hagen	347	5228	827	16.2	108	732	14.0
Oberhausen	371	4575	384	9.2	61	349	7.6
Recklinghausen	375	6611	1026	16.2	223	847	12.8
Solingen	385	2586	645	25.3	64	588	22.7
Wuppertal	391	6463	1446	21.2	226	1148	17.8

All numbers are based on the selected sample. District id. is the internal code used by the Bundesagentur für Arbeit. Treatment rates are defined as the share of treated clients (receiving a voucher or attending training) among all inflowing unemployed in 2004 (with unemployment duration longer than 4 weeks). Unused vouchers do not exactly equal vouchers minus training due to slight inconsistencies in our data definitions.

As we want to look at ALMP, we further limit our data set to those who are unemployed at some point in 2004. For the treatment D^V , those unemployed who receive a voucher make up the treatment group, all others being the control group. For the treatment D^T , the treatment group is limited to those who both receive a voucher and use it, all others again being the control group. A suitable instrument to address the likely endogeneity of these treatments is the probability of treatment: we have argued that it offers considerable exogenous variation, and it still correlates with the occurrence of treatment.

The outcome variable is binary and indicates whether or not a client is in employment after one year has elapsed.¹¹ By employment we mean dependent work subject to social security contributions, of any duration. The time elapsed is counted from the day the voucher is awarded (treatment groups) or from the first day of unemployment (control groups).¹² For the relevant covariates such as age, education, and employment history, we also use the values they take on these dates. The award of the voucher is taken as the starting date also for the training because the following training spells themselves can arise from other ALMP programmes, so that the relevant starting date cannot be unambiguously identified.

5 Estimation and inference

5.1 Estimator

The approach we take to the estimation of the model in section 3.1 is nonparametric. This avoids imposing any particular structure on equation (1) and allows notably for heterogeneous treatment effects. The nonparametric estimator we use is analysed in Frölich (2007), where it is found to be \sqrt{n} -consistent and efficient. Its implementation for a binary instrument is available from Frölich and Blaise Melly.

The innovation of the estimator is to allow for the case that the exclusion restriction for the instrument Z does not hold unconditionally, but only conditionally on covariates X . While we will consider both cases, we will rely on propensity scores whenever covariates are involved, in order to avoid the curse of dimensionality in our context of limited sample sizes. Based on a consistently estimated propensity score $\pi_i = \Pr(Z = z_H | X = x_i)$, the estimator of LATE can then be written as

$$\hat{\Delta}^j(z_H, z_M) = \frac{\sum_{i:z_i=z_H} y_i - \hat{m}_{z_M}(\hat{\pi}_i) - \sum_{i:z_i=z_M} y_i - \hat{m}_{z_H}(\hat{\pi}_i)}{\sum_{i:z_i=z_H} d_i - \hat{\mu}_{z_M}(\hat{\pi}_i) - \sum_{i:z_i=z_M} d_i - \hat{\mu}_{z_H}(\hat{\pi}_i)} \quad (20)$$

with $\mu_z(\pi_i) = E[D|\pi(X) = \pi_i, Z = z]$ and $m_z(\pi_i) = E[Y|\pi(X) = \pi_i, Z = z]$. The former is estimated by local logit regression and the latter by local linear regression. Hence estimation results will be given as deviations from means.

Recall that we apply this estimator to what appears to be an integrated local labour market, in the sense that the initial situation is very similar across market sides. Therefore, if the estimation results with and without covariates are roughly the same, this will indicate that the labour market defined is indeed an integrated local labour market. In other words, if accounting for covariates does not substantially alter the results, it ap-

¹¹The time horizon of six months that underlies the 70% rule is too short for our context where all clients in 2004 are considered together.

¹²The fact that those in the treatment group have already spent some time in unemployment before they receive a voucher might bias our results. This may be avoided by matching on unemployment duration, which is beyond the scope of the present version of this paper.

pears that market sides do not substantially differ with respect to covariates. And while only observable covariates can be taken into account here, we would argue that unobservable covariates are unlikely to substantially differ across market sides when observable covariates do not.

Finally, the same estimator can be used to estimate the treatment outcome \hat{Y}_1^j . To see this, note that equation (4) is again obtained when YD^j in equation (5) is redefined as a dependent variable Y . Redefining $Y(1 - D^j)$ analogously, we can also estimate the non-treatment outcome \hat{Y}_0^j .

5.2 Results

We delimit market sides such that policy districts with similar probabilities of treatment, as shown in table 1, are grouped together. Concretely, the side labelled H consists of Essen, Solingen, and Wuppertal; the side labelled M consists of Bochum, Dortmund, Düsseldorf, Duisburg, Hagen, and Recklinghausen; and the side labelled L consists of Gelsenkirchen and Oberhausen. As each market side has a common border with every other market side, comparisons between any two market sides appear equally valid.

The first four lines in table 2 present estimation results for $D^j = D^V$, while the last four lines consider $D^j = D^T$. In each set of results, the first two lines give the unconditional results respectively for comparisons between H and M and between M and L . The next two lines repeat these comparisons conditional on covariates. The covariates we employ are intended to capture a range of individual characteristics: demographic characteristics (education, age group, sex, nationality, civil status, children, health status, disability/incapacity status), economic characteristics (sector, employment history, previous unemployment incidence, employment as trainee, reason for end of last job) and characteristics decided by caseworkers (motivation, history of sanctions, assignment to problem groups). More macro-level characteristics are not included: even if they differ across market sides, it is not clear that they matter for a given client who can look for a job on any market side.

The estimates of LATE in table 2 that are obtained without covariates suggest large effects: treated compliers on market sides H and M are around 30% more likely to be employed one year after treatment than non-treated compliers. This effect is still around 20% for treated compliers on market sides M and L . These estimates are also all significant at the 1%-significance level. The estimates of LATE with covariates are lower in all cases, suggesting effects of 20% and less. Moreover, they are often insignificant, possibly because these lower point estimates coincide with fewer observations (due to missing values in some covariates). As explained above, the difference between conditional and unconditional estimates points to differences between the market sides. We therefore conclude that LATE is probably more accurately estimated with covariates, so that the large effects obtained without covariates should not be taken at face value. However, it

Table 2: Non-parametric estimates for the Ruhr area

	treatment	market sides	$\hat{\Delta}^j$ (s.e.)	\hat{Y}_1^j (s.e.)	\hat{Y}_0^j (s.e.)	prop. of compliers	covariates	obs.
(1)	$j = V$	H,M	0.341* (0.079)	0.373* (0.060)	-0.032 (0.075)	0.062	no	81833
(2)	$j = V$	M,L	0.257* (0.064)	0.418* (0.042)	-0.161* (0.061)	0.081	no	74070
(3)	$j = V$	H,M	0.204 (0.118)	0.315* (0.091)	-0.111 (0.112)	0.043	yes	74098
(4)	$j = V$	M,L	0.178** (0.081)	0.411* (0.056)	-0.233* (0.077)	0.069	yes	66766
(5)	$j = T$	H,M	0.310* (0.081)	0.327* (0.060)	-0.017 (0.077)	0.062	no	79574
(6)	$j = T$	M,L	0.201* (0.069)	0.345* (0.045)	-0.144** (0.066)	0.076	no	71547
(7)	$j = T$	H,M	0.196 (0.119)	0.304* (0.089)	-0.108 (0.112)	0.043	yes	71982
(8)	$j = T$	M,L	0.118 (0.087)	0.352* (0.058)	-0.233* (0.082)	0.065	yes	64410

With Y denoting the dependent variable, $E[Y|X, Z]$ is estimated by local linear regression with $h = \infty$ and $\lambda = 1$. $E[D|X, Z]$ is estimated by local logit regression with $h = \infty$ and $\lambda = 1$. Propensity score matching is applied, and 5% of the propensity scores are trimmed for the variance estimation. For $j = V$, the treatment group without (with) covariates has 12537 (11780) obs., the control group 70540 (62586). For $j = T$, the treatment group without (with) covariates has 10850 (10214) obs., the control group 71611 (63574).

* Significant at 1% level.

** Significant at 5% level.

turns out that all our further conclusions are the same whether or not they are based on conditional or unconditional estimates.

When we use the estimates of LATE to calculate $\underline{\Delta}^V$ as in equation (9), we consistently find $\underline{\Delta}^V < 0$ and thus evidence against effect-based ordering. Yet $\underline{\Delta}^V$ is not statistically different from 0, according to the test we propose in section 3.3: for both the unconditional and the conditional estimates of LATE, the test statistic is below 1 and therefore not significant at any significance level. Hence there is neither conclusive evidence for nor against effect-based ordering (i.e. hypothesis (E)).

The estimates for the post-treatment outcome Y_1^j reported in table 2 are all significant at the 1%-significance level. They suggest that the probability of being employed a year later is between 30% and 42% higher for treated compliers than the average probability over all compliers. In turn, the estimates for the non-treatment outcome Y_0^j (while often insignificant) suggest that this probability is up to 23% lower for non-treated compliers than the average probability over all compliers. These differences between post-treatment and non-treatment outcomes would be in line with large treatment effects, but also with cream-skimming (i.e. hypothesis (S)). Now note that equation (14) consistently does not hold for estimated treatment outcomes, which is evidence against cream-skimming. We therefore conclude that treatment effects appear to be large, even if clients are not ordered by treatment effect. Instead, it appears that clients are ordered by need (i.e. hypothesis (N)), as equation (13) consistently holds for the estimated non-treatment outcomes.

Let us next turn to the intensive margin discussed in section 3.4. Both $\underline{\Delta}^{T-V}(z_H, z_M)$ and $\underline{\Delta}^{T-V}(z_M, z_L)$, defined in exact analogy to equation (16), are consistently negative. This is evidence for hypothesis (L): among the compliers, the average treatment effect for voucher users is smaller than for all voucher recipients. In other words, those who do not use their voucher would have had above-average treatment effects. The evidence for hypothesis (L) is confirmed by the estimated treatment effects in total: both $\underline{\Delta}^{TT-TV}(z_H, z_M)$ and $\underline{\Delta}^{TT-TV}(z_M, z_L)$, defined in exact analogy to equation (17), are also consistently negative.

Combining intensive and extensive margins, we find that equation (18) holds consistently with an inverted sign: for average treatment effects, there is evidence of increasing returns from choice (i.e. hypothesis (I)). More exactly, the losses from choice found for the intensive margin decrease along the extensive margin. This finding again extends to total treatment effects: equation (19) also holds consistently with an inverted sign, so that the loss in the policy's total treatment effect found for the intensive margin also decreases along the extensive margin.

In conclusion, the application of our methods to estimation results for the Ruhr area allows us to make the following statements: while caseworkers here apparently prioritise clients most in need of better employment chances, treatment effects seem still sizeable. The clients who do not use their voucher tend to have above-average treatment effects, so that both the cost-effectiveness and the overall impact of the policy are lower than they

would be without leaving clients this choice. The difference between the policies with and without choice appears to shrink, however, as the scale of the policy increases.

6 The role of search costs for voucher take-up

A primary concern associated with the use of vouchers are the search costs involved in using it. In the context of training for unemployed, voucher recipients incur such search costs when they look for a suitable training provider. As a complicating factor, these search costs are not observed, although they may be systematically higher for some groups of unemployed than for others. It has therefore often been speculated that voucher recipients with little education, lack of access to information and networks, or high family obligations find it hard to use vouchers, and thus do not use it in disproportionately many cases. This would constitute a problem for virtually any ALMP, and a massive problem for the numerous policies meant to assist in particular the most disadvantaged in society.

Some evidence that unemployed with little education, for example, are indeed less likely to use their voucher is presented by Kruppe (2009). On the background of our finding above that those who do not use their voucher would have had above-average treatment effects, we inquire in this section whether high search costs keep clients from using their voucher. In the absence of data on search costs, one has to rely on indirect evidence. Weber (2008) estimates a double-hurdle model, where being offered training corresponds to the first hurdle, and actually enrolling in the course corresponds to the second. However, this approach cannot link the enrollment decision to any unobserved costs. Here we instead employ a method developed by Altonji et al. (2008) based on the sensitivity analysis in Rosenbaum (2002). This method can be applied to a fully parametric variant of our model:

$$Y = 1[Y^* \geq 0] = 1[X'\beta + \alpha D^T + \varepsilon \geq 0] \quad (21)$$

$$D^T = 1[D^{*T} \geq 0]D^V = 1[X'\gamma - U^T \geq 0]D^V \quad (22)$$

The third model equation is not needed because we will focus on voucher recipients only, so that $D^V = 1$ in all cases and only non-trained voucher recipients form the control group. Z is left out in equation (22) because it only plays a role for an instrumental-variable approach (while we include all market sides here).

Now consider a client with relatively high search costs, represented by a relatively high U^T . We would expect that the same search costs also complicate the client's job search following or during the training, and therefore adversely affect the client's probability Y of being employed one year later. This adverse effect is likely to manifest in a relatively low ε , especially if covariates X do not include variables that capture the client's search costs. In other words: the employment chances of a client with high unobserved search costs are likely to be lower than observed variables would suggest. Formally, this would

mean $\text{cov}(U^T, \varepsilon) \neq 0$.

The method offered by Altonji et al. (2008) tests whether such a correlation is likely to exist. It uses seemingly unrelated bivariate probit to estimate the parametric model under the assumption that

$$\begin{bmatrix} U^T \\ \varepsilon \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \quad (23)$$

The range of (absolute) values considered for ρ is bounded from above by the value that equates the influence of unobservable variables on Y with the influence of observable variables, so that

$$\frac{\text{cov}(\varepsilon, U^T)}{\text{var}(\varepsilon)} = \frac{\text{Cov}(X'\gamma, X'\beta)}{\text{Var}(X'\beta)} \quad (24)$$

Given the reduced sample size, we can only fit a number of covariates X , while ε likely reflects the influence of omitted and unobserved variables. When we include education, sex, employment as trainee, lack of motivation, sanction history, and assignment to problem group, the null hypothesis $\rho = 0$ in a likelihood-ratio test cannot be rejected at any significance level. This result appears robust to marginal changes of the covariates included. Importantly, the covariates do not include several variables that would reflect high search costs, such as disability/incapacity status, nationality (reflecting language problems), family status, children, and health status. While these omitted variables very likely affect Y , they do apparently not give rise to a correlation between U^T and ε , although U^T should capture these variables in so far as they create search costs.

The test result offers indirect evidence that search costs do not drive voucher take-up: if they did, one would expect a significant correlation between U^T and ε , as search costs likely also register in ε . Hence clients who do not use their voucher, even though the effects from training would tend to be above average, probably do so for other reasons than search costs. Concretely, they might underestimate their personal benefit from training, or they might anticipate high opportunity costs of training (e.g. lock-in effects) that lead them to rather focus on job search. These issues would call for a very different response from policy makers than search costs would.

7 Conclusions

This paper has developed simple methods to obtain insights beyond LATE in the context of instrumental-variable estimation of treatment effects. The methods are applied to evaluate the use of vouchers in active labour market policy. Insights are thus offered on the priorities of caseworkers, the role that choice exercised by clients plays for the effectiveness of the policy, and the distribution of gains or losses from choice. For the example of the Ruhr area's local labour market, we have thus found evidence that caseworkers implement

a need-based ordering of clients, that policy effectiveness decreases as a result of choice, and that such losses from choice are concentrated among the first treated clients. As these clients were also found to be most in need of improvement to their employment chances, it emerges overall that vouchers might be inadequate to serve the most disadvantaged clients. The reasons might be false expectations or different priorities on the side of these clients, as indirect evidence suggests that search costs do not prevent them from using a voucher.

References

- [1] ALTONJI, JOSEPH G., ELDER, TODD E. UND TABER, CHRISTOPHER R. (2008): "Using Selection on Observed Variables to Assess Bias from Unobservables when Evaluating Swan-Ganz Catheterization," *American Economic Review (Papers & Proceedings)* 98, 345-350
- [2] BARNOW, BURT S. (2009): "Vouchers in U.S. Vocational Training Programs: An Overview of What We Have Learned," *Journal for Labour Market Research* 42, 71-84
- [3] BELL, STEPHEN H. AND ORR, LARRY L. (2002): "Screening (and Creaming?) Applicants to Job Training Programs: the AFDC Homemaker-Home Health Aide Demonstrations," *Labour Economics* 9, 279-301
- [4] BETTINGER, ERIC, KREMER, MICHAEL, AND SAAVEDRA, JUAN E. (2010): "Are Educational Vouchers Only Redistributive?," *Economic Journal* 120, F204-F228
- [5] BRUTTEL, OLIVER (2005): "Delivering Active Labour Market Policy Through Vouchers: Experiences with Training Vouchers in Germany," *International Review of Administrative Sciences* 71, 391-404
- [6] CULLEN, JULIE BERRY, JACOB, BRIAN A., AND LEVITT, STEVEN (2006): "The Effect of School Choice on Participants: Evidence from Randomized Lotteries," *Econometrica* 74, 1191-1230
- [7] DOERR, ANNABELLE AND KRUPPE, THOMAS (2012): "Bildungsgutscheine und Zertifizierung aus Sicht der Arbeitsverwaltung," IAB mimeo
- [8] FRIEDMAN, MILTON (1955): "The Role of Government in Education," in SOLO, ROBERT A. (ed.): *Economics and the Public Interest*, 123-144, New Brunswick, NJ: Rutgers University Press
- [9] FRÖLICH, MARKUS (2007): "Nonparametric IV Estimation of Local Average Treatment Effects with Covariates," *Journal of Econometrics* 139, 35-75
- [10] FRÖLICH, MARKUS AND LECHNER, MICHAEL (2010): "Exploiting Regional Treatment Intensity for the Evaluation of Labor Market Policies," *Journal of the American Statistical Association* 105, 1014-1029
- [11] GIBBONS, STEPHEN, MACHIN, STEPHEN, AND SILVA, OLMO (2008): "Choice, Competition, and Pupil Achievement," *Journal of the European Economic Association* 6, 912-947
- [12] HASTINGS, JUSTINE S., KANE, THOMAS J., AND STAIGER, DOUGLAS O. (2006): "Preferences and Heterogeneous Treatment Effects in a Public School Choice Lottery," NBER mimeo
- [13] HECKMAN, JAMES, HEINRICH, CAROLYN, AND SMITH, JEFFREY (1997): "Assessing the Performance of Performance Standards in Public Bureaucracies," *American Economic Review (Papers & Proceedings)* 87, 389-395
- [14] HECKMAN, JAMES AND VYTLACIL, EDWARD (1999): "Local Instrumental Vari-

- ables and Latent Variable Models For Identifying and Bounding Treatment Effects,” *Proceedings of the National Academy of Sciences* 96, 4730-4734
- [15] HECKMAN, JAMES AND VYTLACIL, EDWARD (2000): “The Relationship Between Treatment Parameters Within A Latent Variable Framework,” *Economics Letters* 66, 33-39
- [16] IMBENS, GUIDO W. AND ANGRIST, JOSHUA D. (1994): “Identification and Estimation of Local Average Treatment Effects,” *Econometrica* 62, 467-475
- [17] IMBENS, GUIDO W. AND RUBIN, DONALD B. (1997): “Estimating Outcome Distributions for Compliers in Instrumental Variables Models,” *Review of Economic Studies* 64, 555-574
- [18] LAVY, VICTOR (2010): “Effects of Free Choice Among Public Schools,” *Review of Economic Studies* 77, 1164-1191
- [19] LECHNER, MICHAEL AND SMITH, JEFFREY (2007): “What Is the Value Added by Caseworkers?,” *Labour Economics* 14, 135-151
- [20] KRUPPE, THOMAS (2009): “Bildungsgutscheine in der aktiven Arbeitsmarktpolitik,” *Sozialer Fortschritt* 58, 9-19
- [21] MESSER, DOLORES AND WOLTER, STEFAN C. (2009): “Money Matters: Evidence from a Large-Scale Randomized Field Experiment with Vouchers for Adult Training,” CESifo mimeo
- [22] O’NEILL, DAVE M. (1977): “Voucher Funding of Training Programs: Evidence from the GI Bill,” *Journal of Human Resources* 12, 425-445
- [23] OREOPOULOS, PHILIP (2006): “Estimating Average and Local Average Treatment Effects of Education When Compulsory Schooling Laws Really Matter,” *American Economic Review* 96, 152-175
- [24] RINNE, ULF, UHLENDORFF, ARNE, AND ZHAO, ZHONG (2008): “Vouchers and Caseworkers in Public Training Programs: Evidence from the Hartz Reform in Germany,” IZA mimeo
- [25] ROSENBAUM, PAUL R. (2002): *Observational Studies*, Second edition, New York: Springer
- [26] SCHNEIDER, HILMAR, BRENKE, KARL, JESKE, BIRGIT, KAISER, LUTZ, RINNE, ULF, SCHNEIDER, MARC, STEINWEDE, JACOB, AND UHLENDORFF, ARNE (2007): “Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission - Bericht 2006,” IZA mimeo
- [27] SCHNEIDER, HILMAR AND UHLENDORFF, ARNE (2006): “Die Wirkung der Hartz-Reform im Bereich der beruflichen Weiterbildung,” *Journal for Labour Market Research* 39, 477-490
- [28] WEBER, ANDREA (2008): “Individual Incentives in Program Participation: Splitting up the Process in Assignment and Enrollment,” IZA mimeo