DISCUSSION PAPER SERIES

IZA DP No. 8454

Employment and Earnings Effects of Awarding Training Vouchers in Germany

Annabelle Doerr
Bernd Fitzenberger
Thomas Kruppe
Marie Paul
Anthony Strittmatter

September 2014

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Employment and Earnings Effects of Awarding Training Vouchers in Germany

Annabelle Doerr
University of Freiburg

Marie Paul
University of Duisburg-Essen and RGS Econ

Bernd Fitzenberger
University of Freiburg, IFS, IZA, ROA and ZEW

Anthony Strittmatter
University of St. Gallen and University of Freiburg

Thomas Kruppe
IAB, Nuremberg

Discussion Paper No. 8454
September 2014

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

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

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

IZA Discussion Paper No. 8454
September 2014

ABSTRACT

Employment and Earnings Effects of Awarding Training Vouchers in Germany*

In 2003, Germany moved from a system in which participants in training programs for the unemployed are assigned by caseworkers to an allocation system using vouchers. Based on the rich administrative data for all vouchers and on actual program participation, we provide inverse probability weighting and ordinary least squares estimates of the employment and earnings effects of a voucher award. Our results imply that after the award, voucher recipients experience long periods of lower labor market success. On average, there are only small positive employment effects and no gains in earnings even four years after the voucher award. However, we do find significantly positive effects both for low-skilled individuals and for degree courses. The strong positive selection effects implied by our estimates are consistent with sizeable cream-skimming effects.

JEL Classification: J68, H43, C21

Keywords: active labor market policies, training vouchers, treatment effects evaluation, administrative data

Corresponding author:

Bernd Fitzenberger
Department. of Economics
University of Freiburg
79085 Freiburg
Germany
E-mail: bernd.fitzenberger@vwl.uni-freiburg.de

* We gratefully acknowledge helpful comments from Joshua Angrist, Carolyn Heinrich, Peter Mueser, and Jeffrey Smith, as well as from participants in various seminars. This study is part of the project "Regional Allocation Intensities, Effectiveness and Reform Effects of Training Vouchers in Active Labor Market Policies", IAB project number 1155. This is a joint project of the Institute for Employment Research (IAB) and the University of Freiburg. We gratefully acknowledge financial and material support from the IAB. The usual caveats apply.

1 Introduction

Vocational training for the unemployed is an important part of active labor market policy (ALMP) in many countries. Such programs aim at skill enhancement to improve chances of participants in the labor market. In 2003, Germany moved from a system in which participants are assigned to training programs by caseworkers to an allocation system using vouchers. Assigning government-funded programs using vouchers allows recipients to choose among a set of eligible training providers. At the same time the local employment agency specifies the educational objective of the training program, for which the voucher can be redeemed. During the years 2003 and 2004, caseworkers were urged to award a training voucher only when it can be expected that the probability to find a job after training participation is above 70%. Allowing more choice for the participants should result in better choices, thus increasing the effectiveness of training (Posner et al. 2000). However, there is concern that the unemployed may not be sufficiently informed to make good choices in using the training vouchers and that concerns unrelated to the effectiveness of the program may drive the redemption decision. This paper estimates the employment and earnings effects of a voucher award during the years 2003 and 2004. Using rich administrative data, our estimates control for selection with respect to a large set of observable characteristics.

The Adult and Dislocated Worker Program under the Workforce Investment Act (WIA) in the U.S. and the German Training Vouchers are two important cases that use vouchers for the provision of training.[1] In 2003, the German government spent more than 6.5 billion euros for further training programs that were allocated using vouchers. Training vouchers are awarded to the unemployed by caseworkers, if they consider training to be helpful for finding a job. A voucher recipient may choose a course offered by an eligible training provider, if the course fits the training content and the planned duration specified by the voucher.

[1]Training vouchers are not only used in the context of ALMP but also to foster training of employees (see Görlitz, 2010, for a recent evaluation of such training vouchers in Germany). Education vouchers are for the most part used in the schooling system (Posner et al. 2000) and (Ladd, 2002, for a review of the literature on school vouchers).

In the U.S., customers in the WIA program can use the fixed budget provided by the government-funded Individual Training Accounts (ITA) to pay for participation in training. The choice is restricted to eligible training providers who offer occupational skills in demand at the local labor market, but there is more choice in the content of training compared to the German case. There exist several studies on the ITA's and preceding voucher-like programs involving descriptive evidence, experimental evidence, or qualitative evaluations of the implementation (see Barnow, 2009, for an overview). In the 1970s, there was an experiment on the use of training vouchers for needy parents. Participants were randomly assigned to a group receiving counseling only, a group receiving counseling and a 50% subsidy for the costs of basically any sort of training the participant was able to enroll in, and a third group receiving counseling and a 100% subsidy. Although the subsidy led to additional enrollment in training, no positive impact on earnings was found (Barnow, 2009).

More recently, an experiment was conducted to study the relative effectiveness of different levels of counseling and control by the caseworkers. One extreme case would be to create a system in which caseworkers direct customers to a specific course through counseling, award an ITA corresponding directly to a customer's need, and have the right to reject a customer's choice. In a polar-opposite case, caseworkers can award all customers with the same fixed amount for the ITA and provide counseling upon request only. The majority of agencies use a system somewhere in between these two extremes (Perez-Johnson et al. 2011). For the experiment, individuals who were to receive an ITA under the WIA at one of seven particular sites were randomly assigned to three different treatments regarding the freedom of choice of the customer, the counseling requirements, and the award structure (fixed or customized): "structured choice model", "guided choice model", and "maximum choice model". With regard to long-term labor market outcomes, it turned out that participants of all three groups are equally likely to be employed six to eight years after the experiment, but those who were in the "structured choice" group have the highest earnings. Their earnings are significantly higher than those of the "guided choice" group, while the earnings of the "maximum choice group" lie in between (Perez-Johnson et al. 2011).

Heinrich et al. (2013) provide a large scale econometric evaluation of the services provided by the Adult and Dislocated Worker Program under the WIA. Participants receive basic job search assistance and part of them receive intensive counseling or short training courses and some are awarded an ITA for a training program of an external provider. Heinrich et al. (2013) provide separate estimation results for participants in the Adult programs (targeted to individuals with poor work histories) and participants in the Dislocated Worker programs (targeted to individuals who have been laid off). In their main analysis, they estimate the effects of participating in WIA (regardless of the services that are taken) as opposed to not entering WIA. They find large positive employment and earnings effects for the Adult program and find positive employment effects, though only small and insignificant earnings effects, for the Dislocated Worker program. Heinrich et al. (2013) also estimate the effects of receiving training through an ITA as opposed to receiving only the other services of the WIA (and possibly training not related to the WIA program), but advise the reader to interpret the results with some caution. For the Adult program, the long-run earnings effects are large, and there are also positive long-run employment effects. The authors find no positive effects for the Dislocated Worker program in their observation period of four years. Heinrich et al. (2013) estimate the effect of participating in training assigned through an ITA and do not estimate the effect of being awarded with an ITA. In the U.S., this difference may not be important, but it is important in Germany because a considerable number of those receiving a voucher do not participate in training and the timing may be important as described below.

Rinne et al. (2013) estimate the effects of actual participation in training under the voucher system in Germany. Using a dynamic matching approach, the study finds positive effects of training participation after the reform in 2003 on employment and earnings 1.5 years after the program start. Rinne et al. (2013) do not observe the award of vouchers itself but program participation spells. They do not evaluate the treatment "voucher award" but the treatment "training participation". With the latter approach, first, individuals not redeeming a voucher are in the control group and, second, the treatment start and thus also the alignment of participants and control persons occurs in the month in which the treatment starts and not in the month in which the voucher is awarded. Evaluating the treatment "training participation" requires different assumptions to identify a causal effect from those for evaluating the treatment "voucher award". In the former case, the researcher must account for the dynamic selection both for the voucher award and actual participation, while in the latter case only the selection of receiving a training voucher must be accounted for. Moreover, in the former case, the fact that potential participants have already been awarded a voucher when they sign up for training and finally start the program may call into question the assumption that individuals cannot perfectly anticipate the time of treatment (here: the start of the training spell) typically invoked when applying a dynamic matching approach.

To the best of our knowledge, our study is the first to estimate the effect of *being awarded* with a voucher for participation in a training program as an intention-to-treat effect.[2] From a policy perspective, it is the effect of the voucher award that is of prime interest, because this is the policy intervention. The caseworker decides upon the voucher award but cannot perfectly control the actual participation in training. This holds in particular because as part of the 2003 reform, caseworkers were not supposed to sanction an unemployed individual for not redeeming a voucher. We apply a matching strategy, which accounts for selection based on observable characteristics. To avoid the bias that is inevitable if a static evaluation approach is used in a dynamic setting (Frederiksson and Johansson, 2008), we follow Sianesi (2004) and estimate the effects of starting treatment now versus not starting treatment now for each month of elapsed unemployment. The alternative of not starting treatment now entails the possibility that treatment starts in the future. This evaluation approach aligns treated individuals and controls by the elapsed unemployment duration, and it only compares individuals

[2] There is a large literature estimating the effects of public sponsored training for the unemployed in Germany (see Biewen et al. 2014, Hujer, Thomsen, and Zeiss (2006), Lechner, Miquel, and Wunsch (2011, 2007), and Rinne et al. 2013). With the exception of the last study, the literature analyzes the time period before the introduction of the voucher system. The evidence on employment and earnings effects of further training is mixed; see Card, Kluve, and Weber (2010) for a recent review.

who are still unemployed at the time of the treatment start. The approach is implemented using both inverse probability weighting (IPW) and ordinary least squares (OLS) regressions. As a sensitivity analysis, we also implement an instrumental variable (IV) approach exploiting the unexplained variation in differences in policy styles across regional employment agencies.

Our study uses unique rich administrative data provided by the Federal Employment Agency in Germany. We have information on *all* individuals who received training vouchers in 2003 or 2004 and on a 3% sample of all other unemployed. Our data allow us to follow individuals for four years after the voucher award. The data include precise award dates and redemption dates for the vouchers. This information has not been previously available for evaluation studies. We merge the voucher data with individual data records from the Integrated Employment Biographies (IEB), which contains information on employment outcomes and a rich set of control variables, e.g., the complete employment and welfare history, various socioeconomic characteristics, information on health and disabilities, and regional labor market characteristics.

Our results imply that the award of a training voucher has strong and lasting negative lock-in effects. Lock-in effects of training programs can be explained by a lower job search intensity during program participation, and training programs in Germany may even last more than two years. It is four years after the voucher award that small, significantly positive employment effects are found. There are no positive effects on earnings during the observation period. OLS and IPW lead to virtually the same results. A comparison of raw differences between the treatment and control group indicates a strong positive selection of voucher recipients with respect to observable characteristics. In our sensitivity analysis, the monthly IV estimates are quite imprecise. However, at an annual frequency, the IV estimates prove more precise, and they do not differ significantly from the OLS estimates.

Allowing for effect heterogeneity identifies subgroups for which a voucher award is more effective. The employment and earnings effects are more positive for individuals without a vocational degree and for programs leading to a vocational degree. A decomposition of the effect estimates reveals that those unemployed who do not redeem the voucher do better than comparable individuals who are not awarded with a voucher in the short run, but they do much worse in the long run. This suggests that any positive effect of a voucher award actually works through participation in training.

The remainder of the paper is organized as follows: The next section gives a brief overview of the institutional background, followed by the data description. Section 4 discusses identification and estimation. We present our results on the average voucher effect and effect heterogeneity in Section 5. The final section concludes.

2 Background

Before 2003, vocational training for the unemployed in Germany involved the direct assignment by caseworkers of the unemployed to a specific training provider and training course. At the time, the political debate addressed the concern that vocational training was not effective and that this might have been related to the close relationships between local employment agencies and training providers. The First Modern Services on the Labor Market Act (the so-called *Hartz* I Reform) introduced a voucher system for the provision of training for the unemployed in January 2003. Its aim is to foster market mechanisms and transparency in the training market.[3]

During an unemployment spell, individuals repeatedly meet their caseworker for counseling. In the profiling process, the caseworker reviews their potential labor market opportunities. If there is a lack of necessary qualifications to be integrated into employment immediately, participation in a training course is considered necessary. The caseworker denotes the objective, content, and maximum duration of the course on the voucher. The unemployed individual may then choose a course offered by an eligible training provider that is located within a one-day commuting zone subject to the restrictions denoted on the voucher.[4] It is

[3]For more details on the reform, see Schneider et al. (2007).

[4]The one-day commuting zone is defined as a regional zone that can be reached by public

thus the task of the caseworker (potentially in discussion with the unemployed individual) to decide upon the training objective and the educational content of the course. The unemployed individual may choose the provider and the particular course. Eligible (certified) training providers are listed in an online tool provided by the employment agency, and providers may also advertise their courses, e.g., by placing handouts in the employment agency.[5] The caseworker is not allowed to give any advice as to the choice of provider, which is a response to the concern that the relationships between the local employment agencies and training providers were excessively close before 2003. Training vouchers are valid for at most three months, so training has to start within this period.

The German voucher system differs from the WIA system in the U.S. with regard to who makes which decision. WIA customers face two main restrictions: The content of the course must relate to an occupation in demand on the local labor market (which is defined by the local agency), and similar to the German case, the training provider must be listed as an eligible provider. The choice of the content of the training is left to the customer. However, the customer typically has to undergo counseling, which involves an assessment of skills, research on the training programs and the labor market, and face-to-face discussions with the caseworker about the course to choose (McConnell et al. 2011, King and Barnow 2011). In contrast to the German case, WIA customers in the U.S. receive guidance on how to use the voucher but may finally make the decision regarding the content of the training. Thus, after a guided and mandatory decision process, the voucher recipient may decide, for example, to enroll in training to become an IT specialist instead of a care nurse. In Germany, the voucher recipient may state his preference (for example, to become an IT specialist) before the voucher award, but ultimately, the caseworker decides upon the content of the training. Then, after the award of the voucher, the German unemployed individual receives no guidance by the caseworker regarding the choice of training course. Thus, com-

transport in a reasonable amount of time. For a training course with six or more hours a day, commuting times of up to 2.5 hours are reasonable. For a training course with less than six hours a day, the reasonable commuting time is reduced to two hours.

[5] In 2003 and 2004, the Federal Employment Agency was in charge of the certification of the eligible training providers. Afterwards, the certification process was privatized.

pared to the old system, the German voucher does not introduce more freedom of choice with regard to the contents to be studied. However, it nevertheless represents an important change because it allows the unemployed to choose a provider and also to decide not to redeem the voucher. Previously, the unemployed basically received a letter notifying them that they had to present themselves for a training program at a certain date and a certain place. The new system allows for some choice, and for the first time, it treats the unemployed as clients who are eligible for a costly service that may also make a difference for them.

Vocational training programs are used to adjust the skills of the unemployed to the changing requirements of the labor market and possibly to change the conditions of the employability of the individual (due to health problems, for example). Their goal is to improve the human capital and productivity of the participants. Participation prolongs the entitlement period for unemployment benefits.[6] Further training mainly comprises long-term training and degree courses. Long-term training courses typically last several months to one year (in our sample, an average of five months) and usually involve full-time programs. Teaching takes place in class rooms or on the job in training firms. The course curriculum may also include internships. Typical examples of training schemes are courses on IT-based accounting or on customer orientation and the sales approach. With a typical duration of two to three years, degree courses (similar to the former retraining programs) last much longer and lead to a full new vocational degree within the German apprenticeship system. Thus, they cover, for example, the full curriculum of the vocational training for care-assistance for the elderly or for an office clerk. Although the Federal Employment Agency typically covers the costs for at most two years, these programs may last for three years and other programs exist (e.g., those sponsored directly by the state government) that cover the additional costs.

In addition to the opportunity to take part in an intensive training program, training vouchers may influence future labor market opportunities through various channels (see, for example, Barnow, 2000, 2009, Hipp and Warner, 2008, for

[6]The duration of unemployment benefits varies between 12 and 36 months depending on previous employment and age.

a discussion of the potential advantages and disadvantages of using vouchers for the allocation of further training programs). Training vouchers are expected to improve the self-responsibility of the training participants and should introduce market mechanisms into the provision of training. The first main difference with the old system is that the voucher recipients have a choice with regard to the course and the provider. This is expected to change the behavior of the training providers and the selection of those providers that participate in the market. Voucher recipients have the freedom to choose the training provider and the particular program, which should lead to efficient outcomes if they know their needs best. However, it may be the case that experienced caseworkers have a better understanding of the training providers that offer the best programs and the courses that are the most suitable for a particular unemployed individual. Furthermore, the choice on the part of the unemployed individual may be driven by concerns unrelated to the effectiveness of the training program, and some individuals may feel incapable of finding a suitable course, which may have negative effects on motivation. The increased course choices may have a positive effect on the provider side. One would expect that competition for potential clients will have a positive effect on the selection of providers remaining on the market in addition to strengthening the efficiency on their part. To ensure that training providers offer courses that are in line with the regional labor demand, the local employment agencies have to plan and publish their regional and sector-specific demand once a year.[7]

A second difference with the old system is that the caseworker does not impose a sanction when a voucher is not redeemed and the unemployed individual provides a reasonable explanation. After redemption, however, training participation is mandatory. The freedom not to redeem the voucher may change the attitude of the unemployed individual toward this service; the voucher may be perceived as being more like an offer and less like an assignment. This could exert a positive attitude effect such that the unemployed individual may value the fact that a costly service is being offered to him or her and may reciprocate

[7] This is similar to the WIA, stipulating that the local agency provides a list of occupations in demand at the local level.

by increasing the search effort or by participating wholeheartedly in the training program.

Together with the voucher system, the labor market reform in 2003 introduced a new assignment criterion for the award of a voucher. According to predictions, the caseworkers in local employment agencies are supposed to award vouchers such that at least 70% of the voucher recipients find a job within six months after training ends.[8]

3 Data Description

This study is based on unique data provided by the Federal Employment Agency of Germany. These data contain information on *all* individuals in Germany who received a training voucher in 2003 or 2004. The data are generated from internal administrative data and include precise award and redemption dates for each voucher - information that previously has not been available for evaluation purposes.

For each voucher recipient, we merge the information on training vouchers to the individual's data record in the Integrated Employment Biographies (IEB).[9] The data contain detailed daily information on employment subject to social security contributions, receipt of transfer payments during unemployment, job search, and participation in different active labor market programs as well as rich individual information.[10] Thus, we are able to enrich the information from the voucher data with a large set of personal characteristics and a long labor market history for all voucher recipients.

[8]Because this prediction was always made intuitively by the caseworker, the real integration rate never reached this level. The 70% rule was abolished after the time period considered here.

[9]The IEB is a rich administrative data base that is the source of the subsamples of data used in all recent-year studies evaluating German ALMP. It is a merged data file containing individual data records collected in four different administrative processes: the IAB Employment History (*Beschäftigten-Historik*), the IAB Benefit Recipient History (*Leistungsempfänger-Historik*), the Data on Job Search originating from the Applicants Pool Database (*Bewerberangebot*), and the Participants-in-Measures Data (*Maßnahme-Teilnehmer-Gesamtdatenbank*).

[10]A more detailed description of the IEB in English can be found on the website of the Research Data Center of the Federal Employment Agency (http://fdz.iab.de/en.aspx). The version of the IEB we use in this project has been supplemented with some personal and regional information not available in the standard version.

Our control persons are from the same data base: A 3% random sample (based on twelve days of birth during the year) of those individuals in Germany who experience at least one switch from employment to non-employment (of at least one month) between 1999 and 2005 has been drawn. When constructing our sample of analysis, we apply the same selection rules for voucher recipients and control persons. We account for the fact that we use a 100% sample of voucher recipients and a 3% sample of non-recipients by using weights in all tables and estimations.

We consider an inflow sample into unemployment consisting of individuals who became unemployed in 2003, after having been continuously employed for at least three months. Entering unemployment is defined as the transition from (non-subsidized, non-marginal) employment to non-employment of at least one month plus a subsequent (not necessarily immediate) contact with the employment agency, either through benefit receipt, program participation, or a job search spell.[11] We only consider unemployed individuals who are eligible for unemployment benefits.[12] This sample choice reflects the main target group for the training vouchers. To exclude individuals eligible for specific labor market programs targeted to youths and individuals eligible for early retirement schemes, we only consider persons aged between 25 and 54 years at the beginning of their unemployment spell.

We aggregate the spell information in the original data into calendar months. We follow a person in the sample from the month of his or her first inflow into unemployment until the end of 2004 with regard to the voucher award and until the end of 2008 with regard to the employment outcome. We do not consider individuals who receive a training voucher after December 2004 because the next step of the labor market reforms also affecting training was implemented in January 2005. Information from prior periods is exploited when constructing the covariates referring to the labor market history. The focus is on the first voucher

[11]Subsidized employment refers to employment in the context of an ALMP. Marginal employment refers to employment of a few hours per week only; this is due to specific social security regulations in Germany.

[12]Note that, in particular, this condition excludes training programs for mothers returning to the labor market after longer employment interruptions.

awarded. We distinguish the two outcome states non-subsidized, non-marginal employment (henceforth denoted as employment) and non-employment as alternative states. As an alternative outcome variable, we use monthly earnings. The panel data set for the analysis is completed by adding personal, occupational, and regional information. Covariates on individual characteristics refer to the time of inflow into unemployment, whereas covariates on regional characteristics are updated each month.

The final sample includes 133,193 unweighted observations, of which 50,796 individuals are awarded with a voucher during their first twelve months of unemployment and 82,397 observations are in the control group. There are 42,331 individuals in our sample who redeem their vouchers. This amounts to a redemption rate of 83%. We observe 8,465 vouchers that are awarded but not redeemed.[13]

Tables 1 to 4 report the mean values for the most important socioeconomic and labor market characteristics of the individuals in the evaluation sample. In the first two columns of each table, we display the mean value of the respective control variable in the treatment and in the control subsample. In columns six and seven, we distinguish between those who redeem the voucher and those who do not. Voucher recipients are on average more often middle-aged, single or single-parent and females than the individuals in the control group. They exhibit fewer health problems. Individuals who redeem the training voucher and thus participate in a training course are on average slightly older and healthier than individuals who do not redeem their voucher. In addition, the fraction of individuals with children living in the same household is somewhat higher, and the children are on average older than the children of individuals not redeeming a voucher.

Voucher recipients hold a higher schooling degree on average. Furthermore, they tend to have more successful employment histories in the previous 7 years, and in particular, they had higher earnings. The share of individuals with stable employment and no participation in an active labor market program in the past is remarkably higher in the treatment group, already suggesting a strong positive

[13] These individuals would be in the control group if we used the sample design of Rinne et al. (2013).

selection of the treated. We have also information about potential placement handicaps of the unemployed, e.g., indirect information about past psycho-social or drug problems, lack of motivation, received sanction from the caseworker or past incapacities due to illness, pregnancy or child care. Those receiving a training voucher are less likely to exhibit problems of this type. The fraction of people with motivation deficits or past incapacities is even lower for individuals who redeem the voucher.

4 Identification and Estimation

Our analysis will rely on a dynamic selection-on-observables identification strategy, which is motivated by the richness of our administrative data. As a sensitivity analysis, we investigate the robustness of the main results by providing instrumental variable (IV) estimates, which exploit the unexplained variation in policy styles across regional employment agencies.

We consider voucher awards during the first twelve months of unemployment in the first unemployment spell between January 2003 and December 2004. Each unemployed individual is observed for at least 48 months. The indicator for a voucher award as an intention to treat is denoted by $D_{im} \in \{0,1\}$ (with individuals $i = 1, ..., N$ and $m = 1, ..., 12$ indicating the elapsed unemployment duration at the time when the voucher is awarded in months). The outcome variable is denoted by Y_{imt} (where $t = 1, ..., 48$ indicates the number of months since the award of the voucher). We consider employment and monthly earnings as outcome variables, and we estimate the effect of the voucher award (not the actual training participation). To avoid the bias that is inevitable if a static evaluation approach is used in a dynamic setting (Frederiksson and Johansson, 2008), we follow Sianesi (2004) and estimate the effect of treatment start versus no treatment start (treatment versus waiting) for each month of elapsed unemployment duration. The treatment is the award of a voucher, i.e., the intention to assign further training. In the results section, we report a weighted average of the twelve monthly dynamic treatment effects (see Appendix A for details).

The potential outcomes are indicated by Y_{imt}^d, where $d = 1$ under treatment and 0 otherwise. For each individual unemployed until month m, only the realized outcome $Y_{imt} = Y_{imt}^1 \cdot D_{im} + Y_{imt}^0 \cdot (1 - D_{im})$. Our goal is to estimate the expected difference between the outcomes Y_{imt}^0 and Y_{imt}^1 for treated individuals

$$\gamma_{mt} = E[Y_{imt}^1 | D_{im} = 1] - E[Y_{imt}^0 | D_{im} = 1].$$

Hence, $E[Y_{imt}^1 | D_{im} = 1]$ is identified from observed data. In contrast, $E[Y_{imt}^0 | D_{im} = 1]$ involves the expected counterfactual non-treatment outcome for treated individuals. To identify this parameter, we need to make further assumptions.

4.1 Identification Strategy

Assuming that there is only selection on observables, it is possible to control for all confounding variables that jointly influence the treatment probability and the potential non-treatment outcome, summarized by the vector of pre-treatment variables X_{im}. This is formalized by the following dynamic version of the conditional mean independence assumption.

Assumption 1 *(Strong Ignorability).*

i) Dynamic mean independence assumption (DMIA):
$E[Y_{imt}^0 | D_{im} = 1, X_{im} = x] = E[Y_{imt}^0 | D_{im} = 0, X_{im} = x]$ and

ii) Common support: $p(x) < 1$, where $p(x) = Pr(D_{im} = 1 | X_{im} = x)$

hold jointly for all $m = 1, ..., 12$ and $t = 1, ..., 48$.

The DMIA states that conditional on a given unemployment experience and a vector of observed covariates, the sequence of potential outcomes associated with not receiving the treatment in a particular month is mean independent of the treatment status in this month. In a dynamic context, not receiving the treatment in the current month entails the possibility of participation in later months. Our matching approach will produce valid estimates if we consider all the determinants that jointly influence treatment status (i.e., voucher award) and potential outcomes. Conditional on these determinants, individuals are randomly allocated to receiving a voucher or not in a given month, and the treated and non-treated have the same predictions of future treatment or employment chances. We argue in the following that these assumptions are plausible in light of a voucher assignment in Germany and the rich information in our data.

The literature (e.g., Heckman et al. (1999) and Mueser et al. (2007) with regard to US programs and Biewen et al. (2014) and Lechner and Wunsch (2013) with regard to German training programs) stresses the importance of conditioning flexibly on lagged employment and wages, benefit receipt history, basic personal characteristics and local labor market characteristics. These pieces of information are all available in our data, and we use them in a flexible way. The literature addresses the plausibility of the conditional independence assumption (CIA, which is the static counterpart of DMIA) with regard to directly assigning a training program, but we believe that the award of a voucher to be used for a training program involves a similar selection process, which is perhaps less demanding with regard to the CIA because the actual start of the program is not part of this selection. Although training participation was mandatory under the old system in Germany, there may have been individuals who have talked the caseworker into not assigning a program or who have not started it even though they had to. Such cases are demanding for the CIA and do not have to be accounted for in our case. Our data allows us to control for the full labor market history of the previous seven years and on important local labor market characteristics. In their sensitivity analysis, Biewen et al. (2014) find that it is very important to exactly match on the elapsed unemployment duration in months. This is implemented in the present paper by the dynamic approach. Note that the award of a voucher is left to the discretion of the caseworker; thus, from the perspective of the unemployed, the receipt of a voucher cannot be perfectly anticipated. Moreover, the data involves pieces of information that are collected by the caseworker as a basis for his counseling activities and assignment decisions (see also Biewen, Fitzenberger, Osikominu, and Paul (2014)). To be specific, we consider the following variables that reflect part of the caseworker's information on the motivation, plans and labor market prospects of a particular unemployed individual: the caseworker's assessment of the job-seeker's current health status, information on his/her previous health status (during the previous 6 years before the start of the current unemployment spell), a dummy variable indicating whether the unemployed person appeared to lack motivation (e.g., failed to attend regular meetings), dummies indicating whether the job-seeker dropped out of a program, whether benefits were withdrawn, and whether the person participated in a program providing psychosocial support, where all variables refer to the previous 3 years unless stated otherwise. In addition, we include the employment and welfare history as sequences of the previous 7 years before the start of the current unemployment spell and variables indicating whether the job-seeker is looking for a part-time job.

The common support assumption ii) requires that it is possible in large samples to identify for each treated observation some comparable non-treated comparison observations. We apply some simple support tests but are not concerned about the failure of this assumption (see discussion in Lechner and Strittmatter, 2014). Given Assumption 1,

$$E[Y^0_{imt}|D_{im}=1] = E\left[\frac{(1-D_{im})\cdot p(X_{im})}{Pr(D_{im}=1)\cdot(1-p(X_{im}))}\cdot Y_{imt}\right],$$

is identified from the observed data on $\{Y_{imt}, D_{im}, X_{im}\}$ (Hirano, Imbens, and Ridder, 2003). For estimation, we use inverse probability weighting (IPW) and ordinary least squares (OLS). For both approaches, we perform exact matching on the elapsed unemployment duration and the duration since the award of the voucher. Thus, we align treated individuals and controls by the elapsed unemployment duration, and we only compare individuals who are still unemployed at the time of the treatment start. Taking IPW as a benchmark, we specify our parametric OLS regressions to allow for sufficient flexibility.

4.2 Estimation Strategy

Asymptotic theory suggests that IPW has some efficiency advantage in comparison to classical matching estimators in large samples (Heckman, Ichimura, and Todd, 1997, Hirano, Imbens, and Ridder, 2003). Moreover, recent simulation studies support this result (Busso, DiNardo, and McCrary, 2009). Concerning the reweighting technique, we follow the suggestions of Busso, DiNardo, and McCrary (2009) and use weights that sum up to one as a small sample correction. The average effect for the treated is estimated by

$$\hat{\gamma}_{mt} = \sum_{i=1}^{N} \frac{D_{im}}{\sum_{i=1}^{N} D_{im}} \cdot Y_{imt} - \sum_{i=1}^{N} \frac{(1-D_{im}) \cdot \frac{\hat{p}(X_{imt})}{1-\hat{p}(X_{imt})}}{\sum_{i=1}^{N} \frac{(1-D_{im}) \cdot \hat{p}(X_{imt})}{1-\hat{p}(X_{imt})}} \cdot Y_{imt},$$

where $t = 1, ..., 48$ indicates the time after treatment and $m = 1, ..., 12$ indicates the elapsed unemployment duration until treatment. The propensity score $p(X_{im})$ is specified as a probit model. We perform different balancing tests to ensure that the treated and non-treated are well matched with respect to observable characteristics (see Appendix B for details).

Although IPW has some optimality properties, some critical issues may arise. First, the IPW estimators for the average treatment effect for the treated may exhibit fat tails when the treatment probability is close to one. However, the treatment probability in our application is far below one. Second, the implementation of the IPW estimator relies on the estimation of an appropriate specification for the treatment probability (we rely on probit estimates). To demonstrate that our results are robust and not driven by specific issues with one estimator, we contrast the IPW estimates with the estimates obtained by a very flexible OLS regression. Although the implicit parametric assumptions may not hold, OLS might provide a good estimate of the average treatment effects.[14] Because nearly all of the control variables in this study are binary (excluding the earnings history and regional characteristics), our model is very flexible. We find that OLS leads to qualitatively and quantitatively very similar results to those of IPW. Using the same specification as the OLS outcome regressions, we implement an IV approach as a sensitivity analysis (see Appendix C for details). In addition, the IV esti-

[14] Angrist and Pischke (2009) suggest that OLS results often do not differ substantially from results obtained by more demanding non-parametric or semi-parametric estimators in many cases. In particular, they emphasize that the OLS finds exactly the conditional expectation function in fully saturated models, thus providing the non-parametric estimates for such a case.

mates do not differ significantly from the OLS estimates. Therefore, our detailed analysis of heterogeneous treatment effects will rely on the OLS estimates.

5 Results

We first discuss the OLS and IPW estimates of the average treatment effects for the treated. Then, we investigate the heterogeneity of the treatment effects across skill groups and across the type of training programs based on the OLS estimates. Finally, we decompose the effect estimates by whether the treated actually redeem the training voucher. Inference is based on a bootstrap clustering at the individual level, thus resampling all observations over time for an individual. Calculating all estimates based on the same resample allows us to test for differences between different estimators.

5.1 Average Treatment Effects for the Treated

5.1.1 Baseline Results

This section discusses the estimated average effects of a voucher award on employment and earnings based on OLS and IPW. We provide graphical evidence on the descriptive average differences between the treated and the non-treated and on the estimated average treatment effects for the treated. As explained above, we estimate separately the effect of treatment versus waiting for each of the first twelve months of elapsed unemployment durations. We only report the average over these twelve months (further month-specific results are available upon request). On the time axis, we depict the months since the voucher receipt, and on the vertical axis, the outcome variable is depicted. Diamonds indicate a significant effect for the corresponding month. In each figure, the results for the employment (earnings) outcome are placed to the left (right).

Figure 1 depicts the descriptive (unconditional) differences between the treated and nontreated (top line) together with the average treatment effects based on different estimators (OLS and IPW). The OLS and IPW results imply a very long and pronounced lock-in effect. It takes approximately 40 months until the negative effect reaches zero for the employment and even longer for earnings; the lock-in effect is much longer than what is typically found in studies for Germany (see, e.g., Biewen et al. 2014 or Rinne et al. 2013). However, these studies restrict their sample to participants in long-term training and do not consider the much longer degree courses, and the treatment start is defined by the actual start of the training program. Only at the end of our observation period of four years after the award of the voucher, the OLS results imply a very small positive and significant treatment effect (approximately 1-2 percentage points - henceforth, ppoints) for employment. The effect for earnings remains negative even 48 months after the treatment. The results obtained from using IPW are basically the same as those obtained using OLS. This finding suggests that we use sufficient flexibility in our specification of the OLS regression.

Figure 1 indicates that there are strong changes in the slopes of the treatment effect at approximately 12 to 14, 24 to 26, and 36 to 38 months. This finding can be explained by the fact that many programs have a duration of 12, 24 or 36 months and that the majority of treated individuals enter training within the first two months after receiving the voucher (see, Figure 2). Figure 3 displays the average employment and average earnings for treated individuals under treatment and under non-treatment (using the weights of the IPW estimation). Employment under non-treatment is higher than under treatment for the first 3 years after treatment. It takes 40 months after treatment until the employment effect becomes positive.

The descriptive effect in Figure 1 involves a shorter and less pronounced lock-in effect than that of the OLS estimates. This suggests positive selection based on observables both for employment and earnings. As discussed in Section 3, the treated are clearly a positive selection of the unemployed with regard to their labor market chances. Their labor market history is better, with less unemployment experience and higher earnings in the past; they hold higher schooling degrees, suffer less from health problems and less sanctions and are less likely to have dropped out of programs. This positive selection corresponds to the requirement of awarding vouchers only to those unemployed individuals who are expected to have at least a 70% chance of entering employment soon after the program. The control group for the descriptive effect has average characteristics and will thus have a lower employment rate than the matched control group (see column 4 in Tables 1 to 4 for the average characteristics of the matched control group). Because the treated are unemployed individuals with relatively good labor market chances, many of them would have found a job in the short or medium run, if they had not been treated.

As a robustness check, we investigate the sensitivity of our OLS results with respect to selection on unobservables using an IV approach (Appendix C describes the details of the sensitivity analysis). To construct an instrument for the voucher award, we use the remaining variation after having controlled for a large set of individual and regional characteristics. These controls account for individual and regional differences in labor market conditions, which are likely to affect the outcome variables directly.[15] We interpret the remaining regional variation as differences in regional policy style, which can be explained by preferences and sentiments regarding the use of training vouchers. Although the instrument used is highly significant (see Appendix C, Table 6), the IV effect estimates at the monthly frequency are quite imprecisely estimated, and often not significant (these results are not reported in the paper and they are available upon request). To gain precision, we consider average effects by the year since the voucher award (Table 7 in Appendix C). The yearly IV employment (earnings) effects are much more precisely estimated, and they prove to be significantly negative during the first three (two) years. The difference between the yearly IV estimates and yearly OLS estimates is negative for all four years, although never significantly so. In addition, the joint test of equality between OLS and IV (reported at the bottom of Table 7 in Appendix C) during years 1 to 4 and during years 2 to 4 never exhibits significant differences. Thus, although the IV point estimates suggest positive selection on unobservables (i.e., OLS would be upward biased), there are

[15] Regional policy variation in the treatment intensity has been used by a number of studies evaluating labor market policies. For example, Frölich and Lechner (2010) exploit regional variation for the evaluation of Swiss ALMP; Markussen and Roed (2014) use regional variation to construct an instrument for participation in vocational rehabilitation programs in Denmark; and Rehwald, Rosholm, and Rouland (2013) instrument participation in activation measures for sick-listed workers in Norway.

no significant differences between the OLS and the IV estimates. Therefore, our interpretation of the empirical results and our detailed analysis of heterogeneous treatment effects will rely on the OLS estimates.

In sum, the results so far imply that a voucher award leads to a strong and very long negative lock-in effect. It takes four years after the voucher award to find small, significantly positive employment effects. There are no positive effects on earnings within the observation period. Different estimators (OLS and IPW) based on a selection on observables assumptions basically provide the same results, and the OLS estimates do not differ significantly from our IV estimates. Raw employment differences indicate that with regard to observables, voucher recipients represent a strong positive selection with respect to both outcomes (for example, voucher recipients are less likely to be older than 50, and they have earned higher wages in their previous jobs). Altogether, our findings are consistent with cream-skimming by the caseworkers. This seems undesirable because many of the voucher recipients would have found a job much sooner anyway, if they had not received a voucher, and there are no sufficient average positive long-term effects over the course of four years to compensate for the lock-in period.

5.2 Heterogeneous Effects by Skill Level

The mostly negative average treatment effects reported so far may hide heterogeneous treatment effects, which for some subgroups may even be significantly positive. Now, we investigate the differences in effect estimates by skill level. We focus on the OLS results, and additionally, we refer to the descriptive differences. We first investigate effect heterogeneity by vocational degree.[16] One may be concerned that low-skilled individuals may not cope well with a voucher award. They may not find the best training provider, they may not redeem the voucher, or they may be more easily discouraged during participation. However, they may gain significantly by a major investment in their human capital and by obtaining a course certificate or even a vocational degree. Of the treated in our sample, 22%

[16] We have also looked into effect heterogeneity by gender. The effects of the voucher are quite similar for men and women. If at all, women face a little less deep lock-in effect, and the effect estimates are slightly more positive at the end of the observation period.

do not hold a vocational degree (low-skilled individuals). Of the treated, 11% are high-skilled, holding an academic degree. The majority of the treated hold a vocational degree (medium-skilled). The top line in Figure 4 depicts the effect of a voucher award for the group of those without a vocational degree. The lock-in effects last for approximately three years (this is one year shorter than for the whole sample), and four years after the award of the voucher, we find a significant positive employment effect of nearly 6 ppoints and a significant positive earnings effect of approximately 160 euro. In contrast, the effect for the high-skilled is strongly negative over the whole observation period, and there is also no positive effect for the medium-skilled.

Can we say more on why only low-skilled individuals benefit on average? A potential explanation would be that the low-skilled have a shorter lock-in effect because they had a lower probability to redeem the voucher. In our sample, this is not the case: 21.8% of those individuals who redeem the voucher hold no vocational degree, and the share is approximately the same (22.1%) among those who do not redeem the voucher. Furthermore, the average time spent in a training program (conditional on redeeming the voucher) is 14 months for the low-skilled and 10 for the high-skilled. Thus, shorter courses or early dropout do not explain a shorter lock-in period. Furthermore, from month 8 to month 24, the employment effects for the low-skilled are almost parallel to those of the medium-skilled, with a stronger lock-in effect in the levels for the medium-skilled. After month 25, the line for the low-skilled increases more rapidly. This is the time at which the participants in the longer courses complete their courses and search intensively for jobs. Note that low-skilled individuals participate more often in degree courses (44% as opposed to 22% among the medium-skilled), and participants in a degree course spend on average two years in their course. Hence, participants in degree courses (after a quick redemption of the voucher) re-enter the labor market with their new degree approximately 25 to 36 months after the voucher award, and Figure 4 indicates the strongest increase for the low-skilled during that time. These results suggest that the low-skilled voucher recipients eventually do better in finding a job compared to the medium-skilled.

Substantiating this finding, Figure 5 displays the employment rates of the treated and matched controls by skill level. After 36 months, the treated low-skilled exhibit nearly the same employment rate as the treated with a higher skill level. In contrast, the matched low-skilled controls exhibit a much lower employment rate than the matched controls for the two other skill levels.

The effect heterogeneity by skill level seems to be stronger under the voucher system than under the old system in Germany, and the voucher award is more effective for the low-skilled. This may be surprising, as one could fear that in particular, the low-skilled may be overstrained by finding a suitable program. Rinne et al. (2011) and Biewen et al. (2014) find little evidence for effect heterogeneity by skill level for long-term training in the pre-reform period.[17] With regard to degree programs, there exists relatively little prior evidence, because to look beyond the lock-in effect of these very long programs, one needs an observation period of at least three or four years. A series of studies using data from the 1990s are an exception, as they have an extraordinarily long period to observe the labor market outcomes of up to eight years. These studies find positive employment effects for the long retraining program, which is closest to the degree courses investigated in this paper (see Fitzenberger and Völter, 2007, Fitzenberger, Osikominu, and Völter, 2008, Lechner, Miquel, and Wunsch, 2007). In line with our findings, Lechner, Miquel, and Wunsch (2011) estimate the largest positive effects for low-skilled women without a vocational degree. For the U.S., Heinrich et al. (2013) find more positive results for the WIA program for all services as well as for training in particular under the Adult program than for the Dislocated Worker program. Participants in the Adult program are more negatively selected than in the Dislocated Worker program.

5.3 Heterogeneous Effects by Type of Training

In light of the above results, we now distinguish between the two types of training programs: long-term training and degree courses (for the most part retraining).

[17] As one exception, Biewen et al. (2014) report a slightly more positive effect of long-term training for low-skilled males who start their program in months 4 to 6 of the unemployment spell (see the online appendix of Biewen et al. 2014).

Because the type of program (length of the course and the objective of the course) is specified by the voucher, we can treat the two alternatives as multiple exclusive treatments. Here, we do not consider some very special programs or unredeemed vouchers (for the latter, see the next section).

Tables 1 to 3 indicate that participants in degree courses are younger, more likely to be female and unemployed, and earn lower wages in the recent past than participants in long-term training. Degree courses have typically a very long duration. It is thus not surprising that we find long and very deep lock-in effects of more than 3 years, reducing the employment probability by nearly 36 ppoints and earnings by over 600 euro per month. However, after 48 months, the employment effect is 8 ppoints, and earnings gains are relatively large with over 100 euro per month (Figure 9). Thus, degree courses involve high costs due to a very long and deep lock-in period, but after three to four years, they considerably increase the labor market chances. Considering long-term training programs, we find a pronounced lock-in period of approximately 12 months. This lock-in period is comparable to Rinne et al. (2013). However, after this pronounced lock-in period, the estimated effects remain negative for the whole observation period although the effect size is reduced over time. In contrast to our results, Rinne et al. (2013) find a positive employment effect of approximately 7 ppoints at the end of their observation period of 1.5 years after the program start. In Rinne et al. (2013), those who do not redeem a voucher are members of the control group and are likely to form good matches to control for selection. Furthermore, the alignment between the treated and controls in Rinne et al. refers to the start of participation in the training program, when a number of individuals who were comparable at the time of the voucher award (among them, some of those who did not redeem a voucher) may have found a job in the meantime and are thus excluded from the control group. This may induce an upward bias in the effect estimates.

Figures 10 and 11 compare the effect estimates for long-term training and degree courses. Interestingly, the difference with the descriptive effect is a little stronger for long-term courses than for degree courses (Figure 12), suggesting that the effect of cream-skimming is stronger for long-term training. Correspondingly, a comparison of the characteristics of the control group to the treatment group of the degree courses and to the treatment group of long-term training (the last two columns in Table 1 to 3) also suggests that the positive selection on observables is somewhat stronger for long-term training.

When discussing the results on effect heterogeneity by skill group, we have suggested that the positive employment effects for the low-skilled may result from those low-skilled who participate in degree courses. Table 2 confirms that a higher share of participants in degree courses is low skilled (36.3%) than in long-term training (15.6%). Furthermore, degree courses generally exhibit more positive long-term effects than long-term training. Shedding further light on these findings, Figure 13 distinguishes results by skill level and by type of training. In degree courses, we find at least small positive employment effects for all skill levels. We also find positive effects for the low-skilled in long-term training, and the highest positive effect materializes for the low-skilled in degree courses. Positive earnings effects can be found for the low-skilled participating in both types of training and for the medium-skilled taking degree courses. Thus, degree courses seem in general more effective than long-term training and the low-skilled benefit in general from the award of a voucher. In contrast, awarding a voucher for long-term training on average seems ineffective for the medium- and high-skilled.

5.4 Unredeemed Vouchers

The award of a voucher may have an effect by allowing the individual to participate in a training program, but it may also have an effect on the labor market outcomes themselves. Figures 14 to 16 display the effect estimates by the redemption decision. These OLS estimates do not allow for a causal interpretation because the redemption decision itself is endogenous (see discussion above). Nevertheless, these descriptive findings provide a statistical decomposition of the average effect estimates.

Individuals who redeem their vouchers (at 83%, this is the majority among the treated) exhibit the same pattern as for the effect for all treated. However, both the positive and the negative effect estimates are slightly more pronounced. Individuals who do not redeem their voucher are first better off than the corresponding control group of unemployed not being awarded with a voucher. This positive effect may represent a threat effect because individuals may fear being assigned to a mandatory active labor market program three months after the voucher award, such as, for example, a job creation scheme. Note, however, that individuals are not supposed to be sanctioned by a cut in unemployment benefits, if they do not redeem a voucher. However, the positive effect may also be due to those individuals who receive a job offer quickly and who therefore do not redeem the voucher. This positive effect may be the result of higher motivation because the award of a voucher may boost their attachment to the labor market and thus increase their search effort. However, because not redeeming a voucher is not sanctioned, some unemployed with a training voucher may just enjoy their unemployment benefits for three months without being pushed to find a job (note that these are not the ones who find a job quickly). For these individuals, employment chances may have deteriorated over time.

After five months, the effect turns negative. Three potential reasons for this are the following: First, those who do not redeem the voucher may participate in other programs; second, the threat effect may lead to negative consequences in the medium to long run (individuals may have taken unstable or unsuitable jobs); and third, those who do not succeed in finding a training course may suffer from a loss in motivation. Although we do not estimate the causal effects of actual voucher redemption, the findings suggest that the average long run effects of actual training participation are slightly better than the effects of a voucher award.

6 Conclusions

This paper estimates the effect of the award of a training voucher on employment and earnings for the unemployed in Germany. We use rich administrative data on all training vouchers awarded in 2003 and 2004 and on participation in training programs after the redemption of the voucher. We estimate the average effect of a voucher award in a flexible way by OLS and by inverse probability weighting (IPW) as alternatives to control for selection on observables.

Our results imply that the award of a training voucher on average has strong and lasting negative lock-in effects. It takes four years after the voucher award to find small, significantly positive employment effects. There are no positive effects on earnings during the observation period. The two methods based on selection on observables assumptions (IPW and OLS) lead to nearly the same results. The OLS estimates do not differ significantly from our IV estimates, which we obtained in a sensitivity analysis exploiting the unexplained variation in differences in policy styles across regional employment offices. A comparison to raw employment differences indicates that with regard to observables, voucher recipients represent a strong positive selection both regarding employment and earnings. The strong positive selection effects implied by our estimates are consistent with sizeable cream-skimming effects.

An investigation of effect heterogeneity by skill group and by type of training indicates a more positive picture for some subgroups and a more negative one for others: Individuals without a vocational degree are more successful in finding a job after training than higher skilled individuals and the voucher leads to considerable positive long-run effects. Despite strong and lasting lock-in effects, programs leading to a vocational degree work better than those that do not. The strongest positive effects are found for individuals without a vocational degree participating in degree courses. Our study lacks a comprehensive cost-benefit analysis for these subgroups because the observation period is too short to assess whether the positive effects found are sustained after our observation period. Finally, a statistical decomposition by the redemption decision suggests that those treated, who do not redeem the voucher, do better in the short run but worse in the long run than comparable individuals who do not receive a voucher.

Overall, the award of a voucher on average does not improve the labor market perspectives of the voucher recipients. The disappointing result is that, even though most recipients use the voucher to participate in training, they often are not better in the long run, as if they had not been awarded with a voucher. At the same time, they suffer from a lock-in effect that seems to be particularly pronounced due to the strong positive selection of voucher recipients. There are two exceptions to these overall negative findings: Voucher recipients who do not hold a vocational degree and participants in degree courses benefit significantly in the long run.

References

Angrist, J., and J. Pischke (2009): *Mostly Harmless Econometrics*. Princeton University Press.

Barnow, B. (2000): "Vouchers for Federal Training Programs," in *Vouchers and the Provision of Public Services*, ed. by C. Steuerle, V. Doorn Ooms, G. Peterson, and R. Reischauer. Brookings Institution Press, Washington, D.C.

Barnow, B. (2009): "Vouchers in US Vocational Training Programs: An Overview of What We have Learned," *Journal of Labor Market Research (Zeitschrift für ArbeitsmarktForschung)*, 42, 71–84.

Biewen, M., B. Fitzenberger, A. Osikominu, and M. Paul (2014): "The Effectiveness of Public Sponsored Training Revisited: The Importance of Data and Methodological Choices," *Journal of Labor Economics*, forthcoming.

Busso, M., J. DiNardo, and J. McCrary (2009): "New Evidence on the Finite Sample Properties of Propensity Score Matching and Reweighting Estimators," *IZA Discussion Paper*, 3998.

Card, D., J. Kluve, and A. Weber (2010): " Active Labour Market Policy Evaluations: A Meta-Analysis," *The Economics Journal*, 120, F452–F477.

Fitzenberger, B., A. Osikominu, and R. Völter (2008): "Get Training or Wait? Long Run Employment Effects of Training Programs for the Unemployed in West Germany," *Annales d'Economie et de Statistiquer*, 91-92, 321–355.

FITZENBERGER, B., AND R. VÖLTER (2007): "Long-run Effects of Training Programs for the Unemployed in East Germany," *Labour Economics*, 14, 730–755.

FREDERIKSSON, P., AND P. JOHANSSON (2008): "Dynamic Treatment Assignment – The Consequences for Evaluations Using Observational Studies," *Journal of Business Economics and Statistics*, 26, 435–445.

FRÖLICH, M., AND M. LECHNER (2010): "Exploiting Regional Treatment Intensity for the Evaluation of Labour Market Policies," *Journal of the American Statistical Association*, 105, 1014–1029.

GÖRLITZ, K. (2010): "The Effect of Subsidizing Continuous Training Investments – Evidence from German Establishment Data," *Labour Economics*, 17 (5), 789–798.

HECKMAN, J., H. ICHIMURA, AND P. TODD (1997): "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *Review of Economic Studies*, 64(4), 605–654.

HECKMAN, J., R. LALONDE, AND J. SMITH (1999): "The Economics and Econometrics of Active Labor Market Programs," in *Handbook of Labor Economics, Vol. 3*, ed. by O. Ashenfelter, and D. Card, pp. 1865–2086. North Holland.

HEINRICH, C., P. MUESER, K. TROSKE, K. JEON, AND D. KAHVECIOGLU (2013): "Do Public Employment and Training Programs Work?," *IZA Journal of Labor Economics*, 2:6, 1–23.

HIPP, L., AND M. E. WARNER (2008): "Market Forces for the Unemployed? Training Vouchers in Germany and the USA," *Social Policy and Administration*, 42, 77–101.

HIRANO, K., G. IMBENS, AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71(4), 1161–1189.

Hujer, R., S. Thomsen, and C. Zeiss (2006): "The Effects of Vocational Training Programmes on the Duration of Unemployment in Eastern Germany," *Allgemeines Statistisches Archiv*, 90, 299–322.

King, C. T., and B. S. Barnow (2011): "The Use of Market Mechanisms," in *The Workforce Investment Act: Implementation Experiences and Evaluation Findings*, ed. by D. J. Besharov, and P. H., pp. 81–111. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

Ladd, H. F. (2002): "School Vouchers: A Critical View," *Journal of Economic Perspectives*, 16(4), 3–24.

Lechner, M., R. Miquel, and C. Wunsch (2007): "The Curse and the Blessing of Training the Unemployed in a Changing Economy: The Case of East Germany after Unification," *German Economic Review*, 8, 468–509.

——— (2011): "Long- run Effects of Public Sector Sponsored Training," *The Journal of the European Economic Association*, 9, 742–784.

Lechner, M., and A. Strittmatter (2014): "Practical Procedures to Deal with Common Support Problems in Matching Estimation," *SEPS Discussion Paper*, 2014-10.

Lechner, M., C. Wunsch, and P. Scioch (2013): "Do Firms Benefit from Active Labour Market Policies?," *WWZ Discussion Paper 2013/11.*

Markussen, S., and K. Roed (2014): "The Impacts of Vocational Rehabilitation," *IZA Discussion Paper Series*, No. 7892.

Mueser, P., T. K., and A. Gorislavsky (2007): "Using state administrative data to measure program performance," *Review of Economics and Statistics*, 89, 761–783.

Perez-Johnson, I., Q. Moore, and R. Santillano (2011): "Improving the Effectiveness of Individual Training Accounts: Long-Term Findings from an Experimental Evaluation of Three Service Delivery Models," *Final Report, Mathematica Policy Research, Princeton, NJ.*

Posner, P., R. Yetvin, M. Schneiderman, C. Spiro, and A. Barnett (2000): "A survey of voucher use: variations and common elements," *Vouchers and the Provision of Public Services*, pp. 503–539.

Rehwald, K., M. Rosholm, and B. Rouland (2013): "Activating Sick-Listed Workers: Evidence from a Randomized Experiment," *Unpublished Manuscript, University of Aarhus.*

Rinne, U., M. Schneider, and A. Uhlendorff (2011): "Do the skilled and prime-aged Unemployed benefit more from Training? Effect Heterogeneity of Public Training Programmes in Germany," *Applied Economics*, 43, 3465–3494.

Rinne, U., A. Uhlendorff, and Z. Zhao (2013): "Vouchers and Caseworkers in Public Training Programs: Evidence from the Hartz Reform in Germany," *Empirical Economics*, 45(3), 1089–1127.

Schneider, H., K. Brenke, B. Jesske, L. Kaiser, U. Rinne, M. Schneider, J. Steinwede, and A. Uhlendorff (2007): "Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission, Bericht 2006," *IZA Research Report*, 10.

Sianesi, B. (2004): "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s," *The Review of Economics and Statistics*, 86, 133–155.

Smith, J., and P. Todd (2005): "Does Matching Overcome LaLonde's Critique of Nonexperimental Methods?," *Journal of Econometrics*, 125(1-2), 305–353.

Wooldridge, J. (2010): *Econometric Analysis of Cross-Section and Panel Data*. 2nd edition, MIT Press, Cambridge, Massachusetts.

A Averaging across Starting Dates

Following a dynamic treatment evaluation approach (Sianesi, 2004, Frederiksson and Johansson, 2008), we estimate the effect of a voucher award versus waiting for each of the first twelve months of the unemployment period m separately. In the first month, the treatment group includes only individuals who are awarded with a training voucher during the first month. Individuals who either receive a voucher later or never are in the control group. In the second month, we drop all individuals who have left the risk set in the first month, i.e., received a voucher or found employment in the first month. The treatment group in the second month consists of voucher recipients that are awarded with a voucher in their second month of the unemployment period. Everybody in the risk set who does not receive a voucher in the second month belongs to the control group. This procedure continues until month twelve. By using this dynamic approach, we end up with twelve different treatment effects for each of the twelve different times of elapsed unemployment duration. To communicate our results, we reduce the dimension of the results by reporting a weighted average of the twelve dynamic treatment effects in the following. The weights are calculated as the fraction of treated in the respective month of the total number of treated individuals

$$\hat{\gamma}_t = \frac{\sum_{m=1}^{M} \sum_{i=1}^{N} D_{im} \cdot \hat{\gamma}_{mt}}{\sum_{m=1}^{M} \sum_{i=1}^{N} D_{im}}.$$

Given that we observe the labor market outcomes of each individual for 48 months after treatment ($t = 1, ..., 48$), we specify a separate model for each month after treatment. This induces flexibility in all parameters with respect to the duration since treatment.

B Matching Quality

We assess the matching quality by displaying the means of the matched control group for different control variables in Tables 1-3. Further, we report the standardized differences before and after matching. The standardized differences are defined as

$$SD = \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{0.5(\sigma^2_{X_1} + \sigma^2_{X_2})}} \cdot 100,$$

where $\bar{X}_d$ is the mean and $\sigma^2_{X_d}$ the variance in the respective treatment group $d \in \{0,1\}$. Before matching, we observe standardized differences larger than 40. After matching, the standardized differences are always below one, suggesting a very good matching quality.

We also apply a second balancing test following an approach of Smith and Todd (2005). Therefore, we run the regression

$$x_k = \hat{\beta}_0 + \hat{\beta}_1 D_{im} + \hat{\beta}_2 \hat{p}(X_{im}) + \hat{\beta}_3 D_{im} \hat{p}(X_{im}) + \hat{\varepsilon}_{im},$$

where x_k indicates the specific control variable. We perform a joint F-test for the null hypothesis that $\hat{\beta}_1$ and $\hat{\beta}_3$ equal zero. In Table 5, we report the summarized results of the test for each of the twelve treatment times. Overall, we run 1,272 regressions, of which the test indicates a rejection of the null hypothesis in only 74 cases. We take the results of the assessment as an indication that the propensity score is well balanced and acceptable for the performance of the IPW estimations. Because we control directly for X_{im} in the OLS and IV regressions, it is not necessary to assume that the propensity score is balanced for these estimators.

C Sensitivity Analysis: Instrumental Variable Approach

As a robustness check, we apply an instrumental variable (IV) approach. In the case of selection into treatment based on factors unobserved by the researcher, an IV approach may provide consistent estimates of the treatment effects (for the subset of compliers in the random coefficients case). We use an IV approach to assess the impact of selection on unobservables. If the results that we obtain from the IV, OLS, and IPW approach do not differ significantly, we argue that our OLS and IPW approaches control sufficiently for all confounding variables.

To construct an instrument for the voucher award, we exploit the variation in the conditional regional-specific allocation intensity of training vouchers. Regional policy variation in the treatment intensity has been used by a number of studies evaluating labor market policies (see references in footnote 15). In our case, the variation in the conditional employment district-specific allocation intensity, which we name conditional regional policy style, can be explained by preferences and sentiments regarding the use of training vouchers that differ across employment offices. This preference is assumed to be independent of the regional labor market characteristics after controlling for a large set of individual and regional characteristics. The implicit assumption is that solely living in a region with a high or low allocation intensity, without receiving a voucher, has no influence on the potential outcomes.

The number of vouchers awarded per unemployed varies across and within employment offices. As an indication of the between variation, Figure 17 displays the differences in unconditional award intensities across employment office districts in Germany. In some areas of Germany, there exist large differences even between neighboring districts. The employment offices themselves decide upon how much of their budget is used for training vouchers and how much for alternative instruments of ALMP. Lechner, Wunsch, and Scioch (2013) argue that local employment offices have a high degree of autonomy in defining the mix of ALMP they are implementing, which partly depends on preferences that are unrelated to the labor market. Furthermore, they decide upon the targeting of the training vouchers. The differences in voucher award intensities can partly be explained by differences in attitudes of the caseworkers in different employment offices.

Apart from the policy style, the allocation intensity is likely to depend upon regional labor market characteristics reflecting differences in labor demand and supply. To identify the policy style, we use the residual variation after controlling both for individual characteristics of the unemployed and the aforementioned regional covariates. Specifically, the latter comprise the characteristics of the stock of unemployed in a region, the number of vacancies for full time jobs, the share of foreigners among the unemployed, and the industry structure of employment in the region.

We implement our IV approach in two steps analogous to Procedure 21.1 in Wooldridge (2010, p. 939). In the first step, we allow for a full interaction of the regional policy style with all covariates considered. For each region, we estimate a separate linear probability model (the point estimates are robust to estimating a probit model) for the dummy variable voucher award to individual i in month m

$$D_{im} = \alpha_{0,r} + X'_{irm} \cdot \alpha_{mr} + v_{im}, \tag{1}$$

where X_{irm} involves regional and individual covariates and r (with $r = 1, ..., 181$) refers to the region of individual i. Based on these estimates, we calculate the predicted probabilities $\hat{p}_{im} = \hat{\alpha}_{0,r} + X'_{irm} \cdot \hat{\alpha}_{mr}$ for a voucher award. These probabilities reflect differences across regions in the labor market conditions and across individuals with different labor market outcomes, both of which we do not want to use as exogenous variation in voucher awards. As instruments, we only use the residual differences, which we allow to differ by individual characteristics and which we attribute to exogenous differences in the policy style.

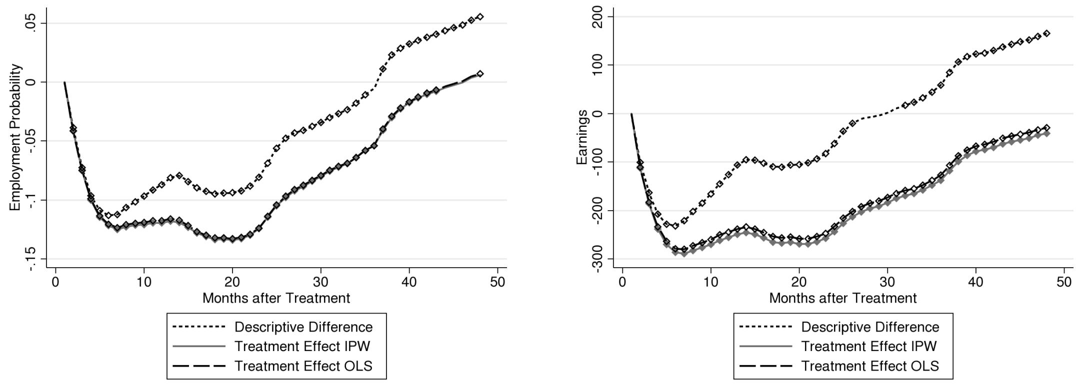
In the second stage, we run IV regressions, which are pooled across regions, using $\hat{p}_{im}$ as the conditionally exogenous instrument while controlling in the outcome equation (the second stage of IV for employment or earnings outcomes) for differences across regions in the labor market conditions as in the first stage of the Wooldridge Procedure. Thus, we do not exclude regional supply and demand effects and individual characteristics of the unemployed from the outcome regressions. Correspondingly, the conditional variation in $\hat{p}_{im}$ given all other regressors used in the outcome regressions presumably reflects the aforementioned heterogeneous differences in the policy style across regions.

Table 6 provides the F-statistics for the significance of the single instrument $\hat{p}_{im}$ in the first stage of the IV regressions for month m based on clustered bootstrap standard errors. These F-statistics lie above 1000 and for the most part above 2000; thus, in a formal sense, the instruments are very strong for the second stage. However, our instruments are based on region-specific estimates of the variations in voucher awards, and we also report adjusted F-statistics, for which we divide the aforementioned F-statistics by the number of regions minus one. We think these adjusted F-statistics provide a better assessment of the bite of the instrument. The adjusted F-statistics are larger than 10 (the typical rule-of-thumb threshold in the literature) in 10 out of 12 months. Nevertheless, our IV estimates of the treatment effects at a monthly frequency (that is, the frequency at which we report the OLS and IPW results in the main part of the paper) involve a fairly large estimation error and are often not significant (these IV results at the monthly frequency are not reported in the paper, and they are available upon request). For these reasons, our sensitivity analysis only reports the IV and OLS estimates averaged by the year since treatment; see Table 7.

The yearly IV employment (earnings) effects are significantly negative during the first three (two) years. The treatment effects estimated by OLS and IV (the second and third column) remain negative and insignificant in the case of IV. The second-to-last column displays the difference between the descriptive estimates and the OLS estimates. This difference is always significantly positive, which is consistent with positive selection based on observables in all four years as discussed in the main part of the paper. This is also the case for earnings. The last column displays the difference between the IV estimates and OLS estimates. The difference is consistently negative, though never significantly so. In addition, the joint test of equality between OLS and IV (reported at the bottom of Table 7) during years 1 to 4 and during years 2 to 4 never exhibit significant differences. Thus, for yearly treatment effects, there are no significant differences between the OLS and the IV estimates.

Figures and Tables

Figure 1: Effect of a voucher award on employment and earnings averaged over elapsed unemployment durations until treatment.



Diamonds indicate significant effects.

Figure 2: Fraction of individuals in training after the award of a voucher.

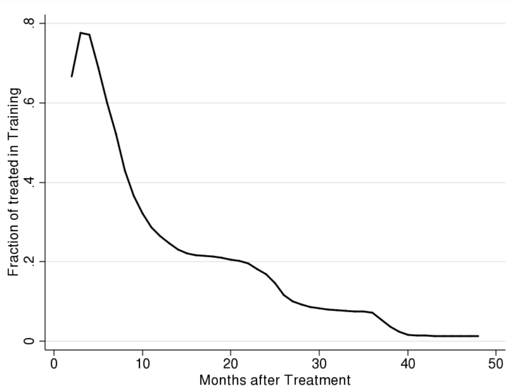


Figure 3: Comparison of average employment and average earnings between treatment and matched control group averaged over elapsed unemployment durations until treatment.

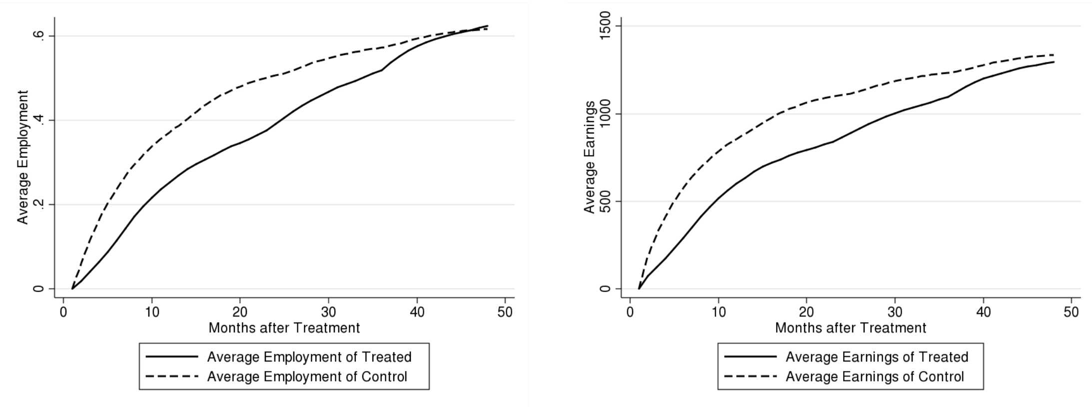


Figure 4: Heterogeneous effects on employment and earnings by skill group (OLS) averaged over elapsed unemployment durations until treatment.

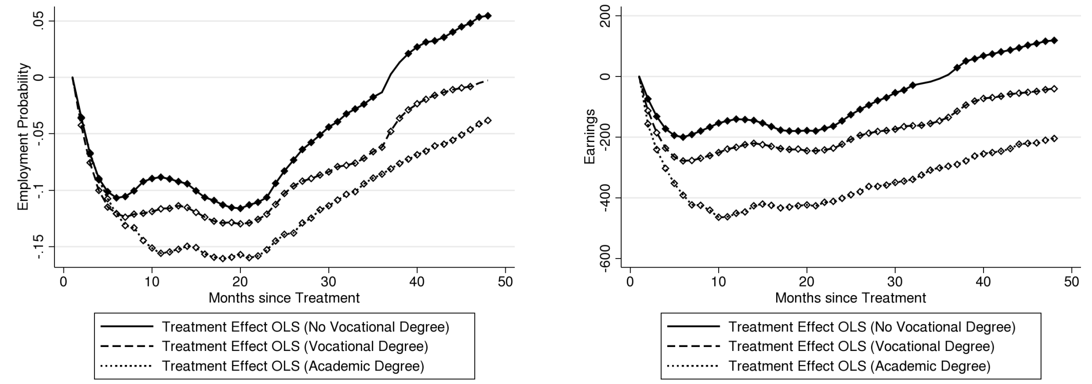


Figure 5: Comparison of average employment of treated and matched control group by skill group averaged over elapsed unemployment durations until treatment.

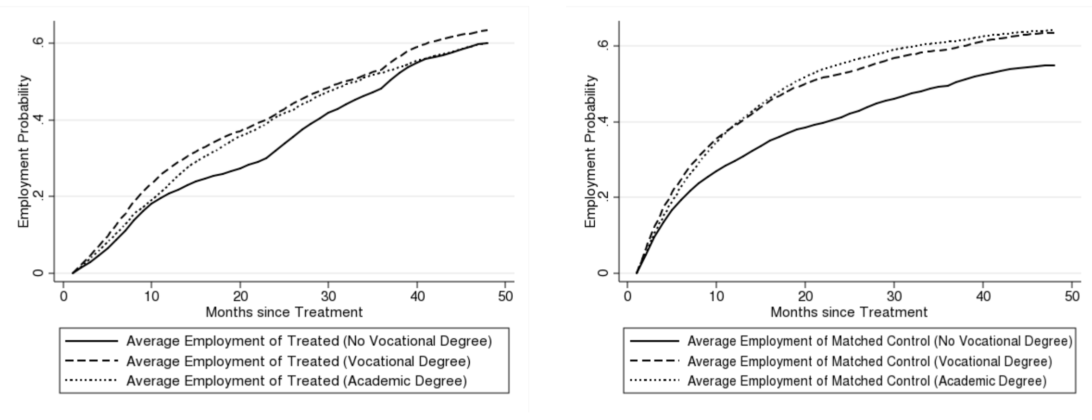


Figure 6: Effect of a voucher award on employment and earnings for individuals without vocational degree averaged over elapsed unemployment durations until treatment.

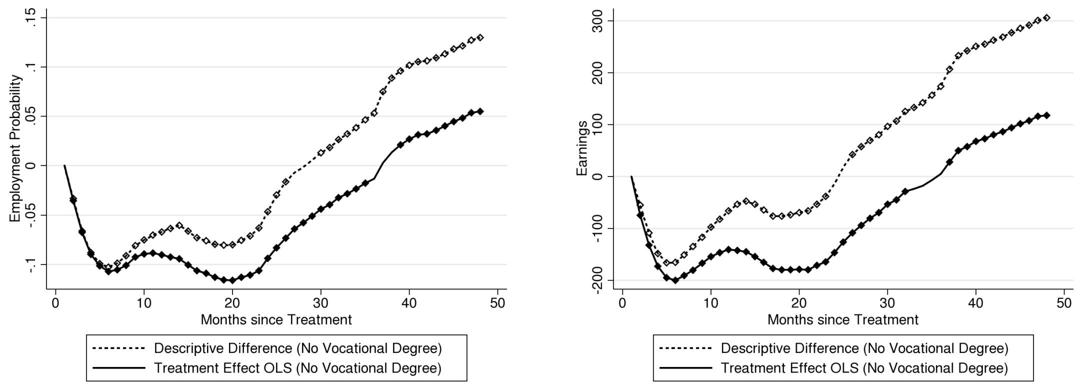


Figure 7: Effect of a voucher award on employment and earnings for individuals with vocational degree averaged over elapsed unemployment durations until treatment.

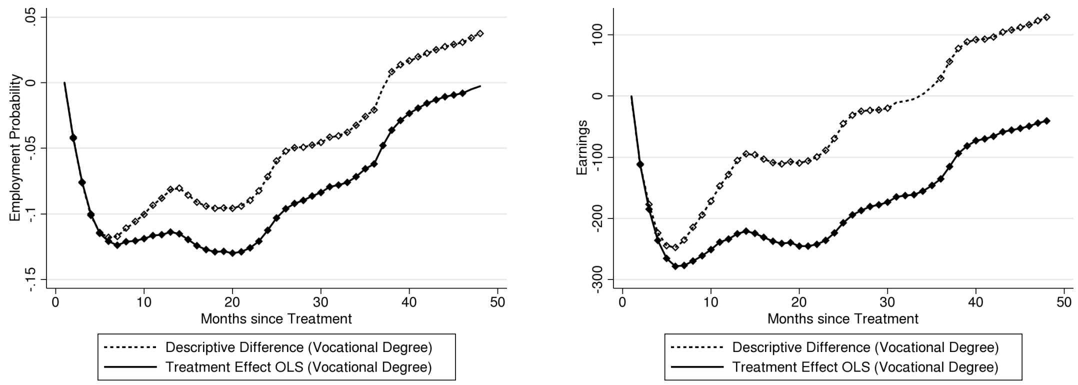


Figure 8: Effect of a voucher award on employment and earnings for individuals with academic degree averaged over elapsed unemployment durations until treatment.

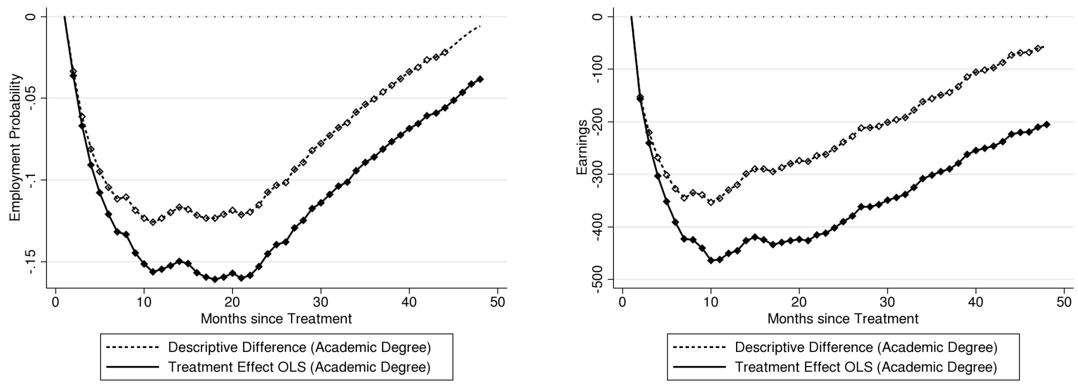


Figure 9: Heterogeneous effects on employment and earnings with regard to the type of training (OLS) averaged over elapsed unemployment durations until treatment

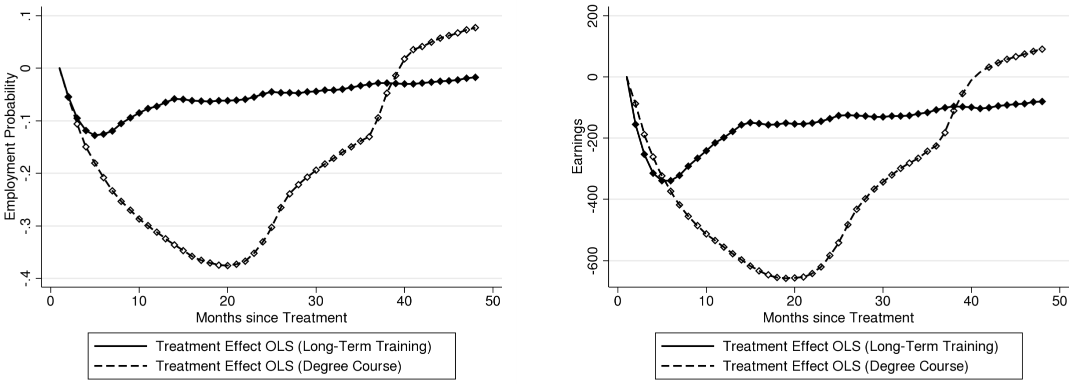


Figure 10: Effect of a voucher award on employment and earnings for individuals participating in long-term courses averaged over elapsed unemployment durations until treatment.

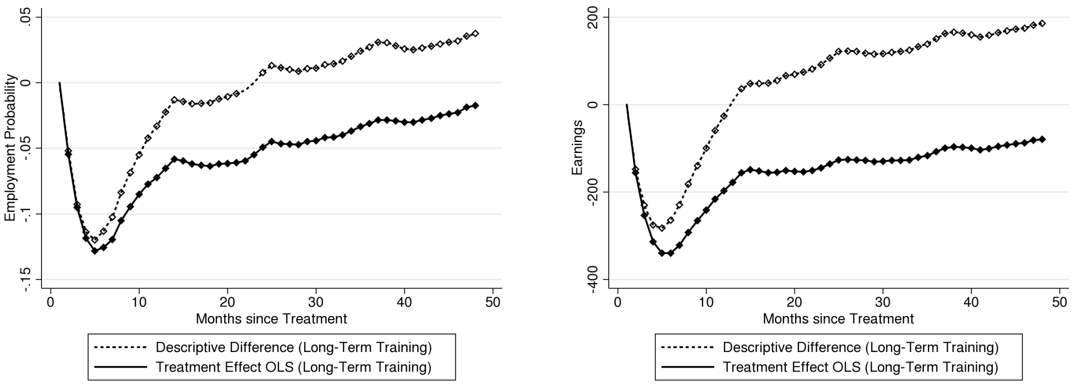


Figure 11: Effect of a voucher award on employment and earnings for individuals participating in degree courses averaged over elapsed unemployment durations until treatment.

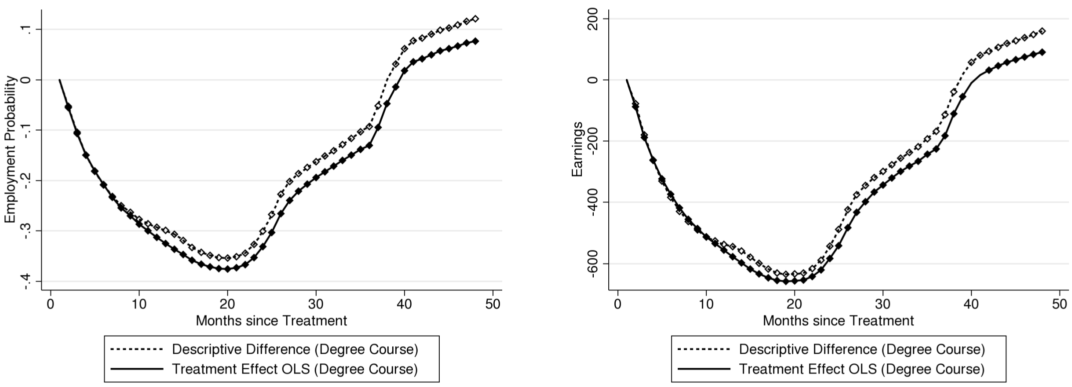


Figure 12: Comparison of average employment of treated and matched control group by course type averaged over elapsed unemployment durations until treatment.

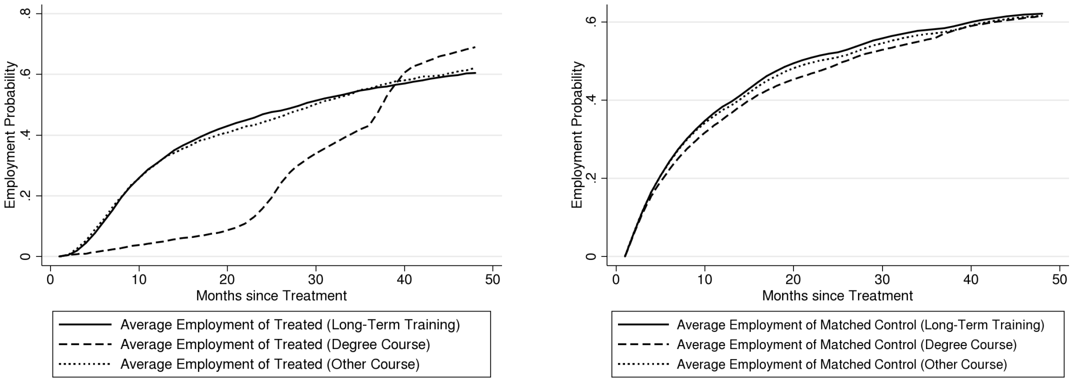


Figure 13: Heterogeneous effects on employment and earnings with regard to the type of training and vocational degree (OLS) averaged over elapsed unemployment durations until treatment

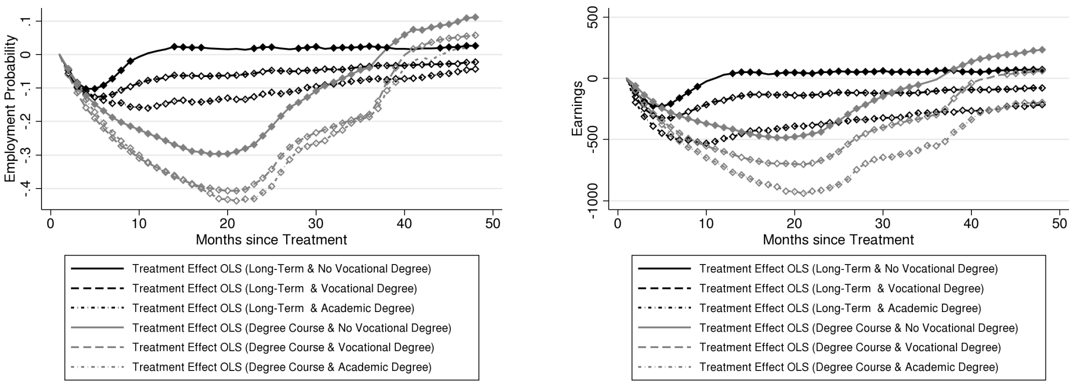


Figure 14: Heterogeneous effects on employment and earnings with regard to the redemption decision (OLS) averaged over elapsed unemployment durations until treatment.

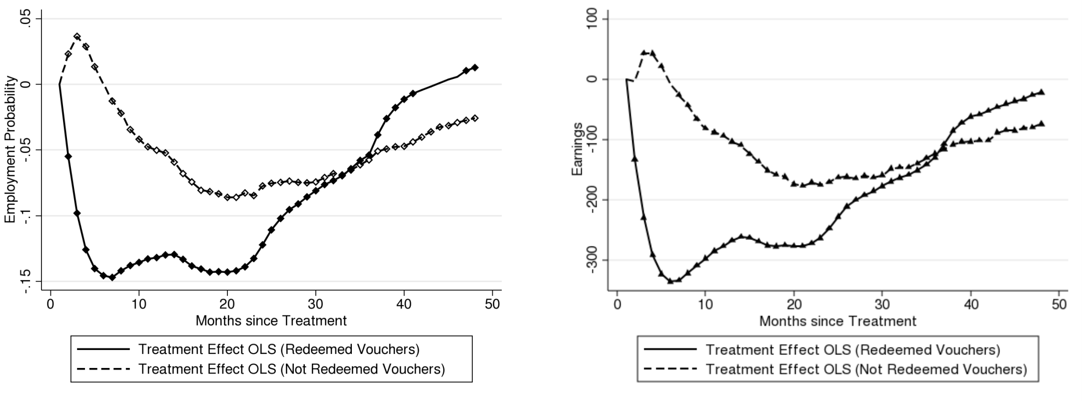


Figure 15: Effect of a voucher award on employment and earnings for individuals who redeem the voucher averaged over elapsed unemployment durations until treatment.

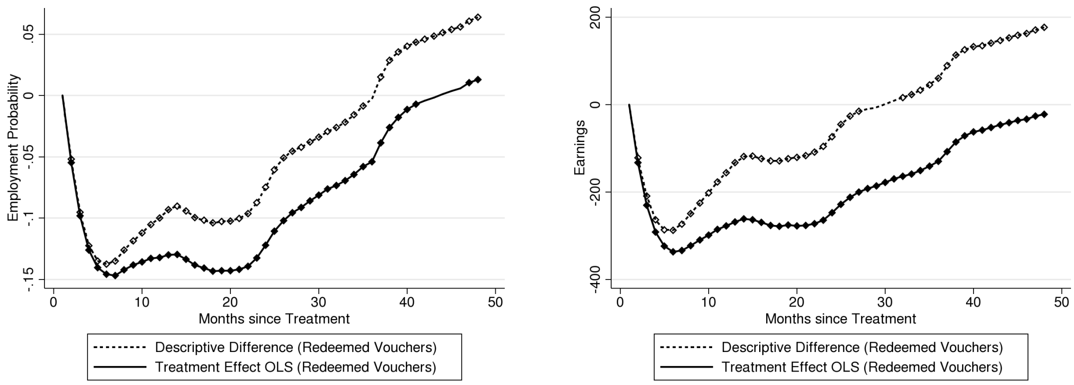


Figure 16: Effect of a voucher award on employment and earnings for individuals who do not redeem the voucher averaged over elapsed unemployment durations until treatment.

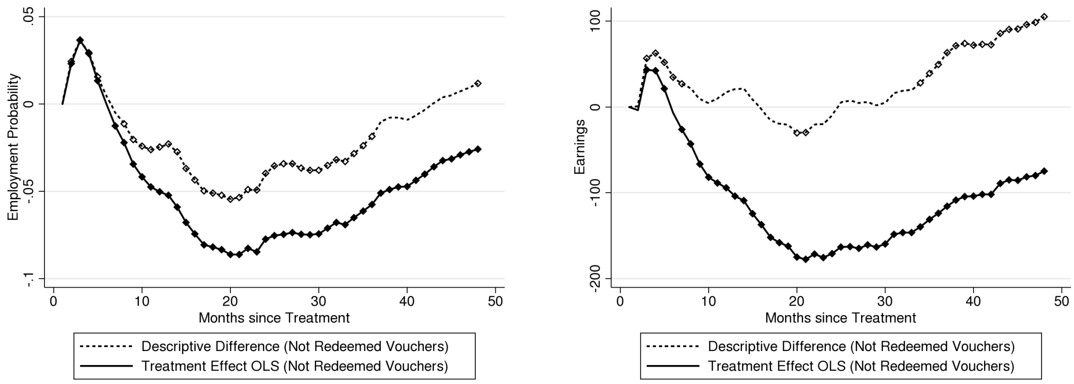
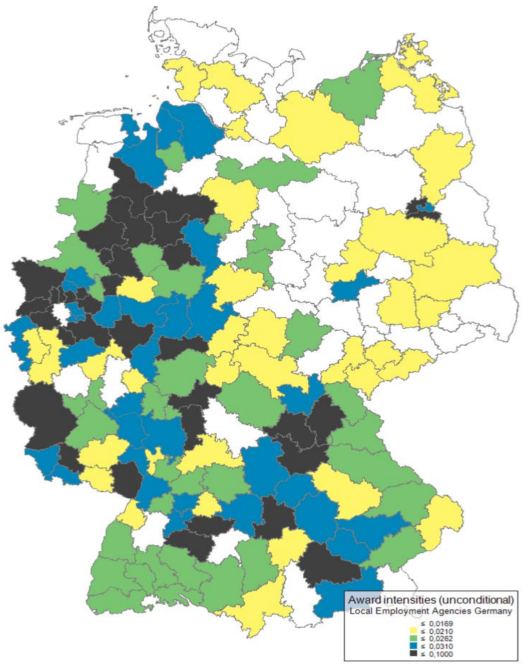


Figure 17: Regional Differences in Voucher Awards per Unemployed



Notes: Differences in unconditional award intensities across employment office districts. Min= 0.08%, Max= 5.59%, Mean= 2.43%, Award Intensity = #Voucher Recipients/#Unemployed by District.

Table 1: Means and Standardized Differences (SD) for Personal Characteristics

	Treatment-group	Control-group	SD before Matching	Matched Controlgroup	SD after Matching	Voucher redeemed	Voucher expired	Degree Courses	Long-term Courses
Female	0.446	0.431	6.630	0.445	0.180	0.446	0.445	0.490	0.416
Age									
25-29 years	0.156	0.155	1.530	0.158	0.430	0.154	0.166	0.234	0.126
30-34 years	0.189	0.176	3.540	0.1900	0.170	0.189	0.193	0.250	0.171
35-39 years	0.233	0.205	6.700	0.233	0.190	0.234	0.226	0.245	0.229
45-49 years	0.142	0.155	3.860	0.141	0.220	0.143	0.139	0.074	0.165
50-54 years	0.071	0.115	15.340	0.070	0.180	0.070	0.074	0.015	0.088
Nationality									
Germany	0.928	0.906	8.000	0.929	0.200	0.930	0.923	0.910	0.938
Outside EU	0.031	0.060	14.210	0.031	0.120	0.030	0.031	0.040	0.027
Missing	0.017	0.007	8.580	0.016	0.160	0.016	0.020	0.019	0.015
Marital Status									
Single	0.322	0.310	3.810	0.323	0.260	.318	0.344	0.287	0.337
Single parent	0.071	0.058	5.150	0.071	0.150	.076	0.069	0.098	0.061
Married	0.462	0.484	4.590	0.462	0.120	.467	0.437	0.441	0.477
Missing	0.102	0.100	3.660	0.101	0.280	.100	0.107	0.125	0.082
Child	0.363	0.355	2.850	0.363	0.160	0.369	0.335	0.420	0.351
Age of youngest child									
One year	0.012	0.011	1.980	0.012	0.090	0.012	0.011	0.014	0.011
Between 1 and 3 years	0.035	0.031	2.510	0.035	0.100	0.036	0.033	0.042	0.034
Between 3 and 6 years	0.065	0.061	2.160	0.065	0.130	0.066	0.059	0.085	0.061
Between 6 and 10 years	0.082	0.075	2.860	0.087	0.110	0.082	0.080	0.103	0.074
Older than 14 years	0.086	0.098	4.100	0.086	0.150	0.088	0.078	0.081	0.091
Missing	0.638	0.647	2.860	0.639	0.160	0.633	0.666	0.581	0.650
Disabled	0.020	0.026	3.980	0.020	0.150	0.019	0.026	0.007	0.024
Health									
Health problems	0.094	0.120	8.330	0.094	0.220	0.092	0.107	0.081	0.096
Health problems before unemployment	0.040	0.050	4.910	0.040	0.070	0.039	0.046	0.033	0.040
N	50,796	82,397				42,331	8,465	10,976	26,721

Omitted Categories:
Age: 40-44 years
Nationality: Member EU
Marital Status: Common law marriage
Age of youngest child: Between 10 and 14 years

Table 2: Means and Standardized Differences (SD) for Education, Occupation, and Sector

	Treatment-group	Control-group	SD before Matching	Matched Controlgroup	SD after Matching	Voucher redeemed	Voucher expired	Degree Courses	Long-term Courses
Education									
No schooling degree	0.041	0.068	11.980	0.041	0.070	0.041	0.042	0.046	0.038
University entry degree	0.225	0.173	13.030	0.226	0.360	0.227	0.214	0.163	0.267
Missing	0.012	0.014	2.480	0.012	0.110	0.012	0.014	0.016	0.010
Vocational Training									
No vocational degree	0.218	0.230	7.400	0.217	0.350	0.218	0.221	0.363	0.156
Academic degree	0.108	0.089	6.450	0.109	0.450	0.110	0.099	0.050	0.146
Missing	0.012	0.014	2.400	0.012	0.130	0.012	0.014	0.016	0.010
Classification of Occupation									
Farmer, Fisher	0.013	0.024	8.310	0.013	0.190	0.013	0.011	0.019	0.012
Technical	0.077	0.054	9.370	0.078	0.170	0.078	0.074	0.024	0.105
Service	0.621	0.580	8.350	0.621	0.130	0.612	0.627	0.629	0.616
Other	0.004	0.005	3.420	0.004	0.190	0.004	0.004	0.006	0.003
Part-time work									
Full-time	0.804	0.789	8.140	0.805	0.270	0.805	0.801	0.773	0.832
Missing	0.071	0.081	3.930	0.071	0.290	0.070	0.076	0.082	0.061
Part-time work desired									
Desired	0.830	0.823	4.480	0.830	0.230	0.831	0.825	0.821	0.850
Missing	0.085	0.085	4.270	0.085	0.310	0.084	0.088	0.108	0.065
Type of work									
White-collar	0.475	0.381	19.030	0.476	0.210	0.474	0.479	0.335	0.536
Missing	0.106	0.109	6.660	0.106	0.140	0.108	0.096	0.133	0.091
Azubi	0.029	0.018	11.880	0.029	0.310	0.031	0.021	0.049	0.012
Sector									
Agriculture	0.009	0.015	5.890	0.009	0.110	0.009	0.008	.011	.008
Mining	0.002	0.002	1.210	0.002	0.090	0.002	0.001	.002	.002
Utilities	0.002	0.002	1.140	0.002	0.110	0.002	0.002	.001	.002
Construction	0.068	0.100	11.450	0.068	0.150	0.068	0.067	.056	.074
Trade	0.150	0.132	5.170	0.150	0.140	0.149	0.155	.140	.153
Hotels and Restaurants	0.028	0.038	5.120	0.028	0.120	0.028	0.033	.038	.024
Traffic, Transportation	0.054	0.056	1.470	0.053	0.160	0.054	0.054	.065	.051
Financial Services	0.020	0.013	5.180	0.019	0.140	0.020	0.018	.015	.022
Renting	0.010	0.010	1.290	0.010	0.070	0.010	0.010	.006	.012
Data processing	0.144	0.118	7.770	0.143	0.240	0.143	0.147	.093	.170
Public Sector, Education	0.056	0.062	4.680	0.056	0.240	0.055	0.057	.059	.057
Health and social services	0.074	0.072	14.600	0.074	0.280	0.075	0.067	.137	.042
Other Services	0.040	0.042	2.240	0.040	0.130	0.041	0.038	.049	.038
Temporary Employment	0.133	0.171	12.690	0.134	0.360	0.132	0.136	.142	.129
N	50,796	82,397				42,331	8,465	10,976	26,721

Omitted Categories:
Education: Schooling degree without Abitur
Vocational Training: Vocational Degree
Classification of Occupation: Miner and Manufacturing
Part-time work: Part-time
Part-time work desired: Not desired
Type of work: Blue-collar
Sector: Production

Table 3: Means and Standardized Differences (SD) for Employment/Unemployment/ALMP History

	Treatment-group	Control-group	SD before Matching	Matched Controlgroup	SD after Matching	Voucher redeemed	Voucher expired	Degree Courses	Long-term Courses
Noticeable problems									
Problem group	0.018	0.025	4.790	0.018	0.180	0.018	0.017	0.015	0.020
Sanction	0.011	0.031	14.010	0.011	0.110	0.011	0.014	0.014	0.008
Lack of Motivation	0.108	0.134	9.160	0.108	0.110	0.106	0.116	0.133	0.095
Incapacity	0.136	0.213	21.000	0.136	0.250	0.128	0.180	0.124	0.129
Dropout	0.012	0.054	23.650	0.012	0.210	0.012	0.013	0.015	0.010
Employment History (last 7 years), Sequences (1 for employed, 0 for unemployed)									
Mostly employed in last period (i.e., 1111000, 1101000, 1000101)									
Mostly unemployed (i.e., 1000010)	0.170	0.223	13.180	0.171	0.290	0.170	0.173	0.228	0.150
3 years employed, close (i.e., 1111010)	0.131	0.095	11.280	0.131	0.100	0.131	0.132	0.135	0.127
3 years employed, far (i.e., 1100111)	0.026	0.055	14.690	0.026	0.190	0.026	0.027	0.023	0.027
3 years unemployed, close (i.e., 1000011)	0.012	0.025	9.969	0.012	0.120	0.012	0.011	0.010	0.012
3 years unemployed, far (i.e., 1101000)	0.099	0.088	3.640	0.099	0.210	0.099	0.095	0.112	0.095
Mixed employment (i.e., 1101101)	0.049	0.061	5.430	0.049	0.170	0.049	0.049	0.053	0.047
Mostly unemployed in last period (i.e., 0111000, 0101000, 0000101)									
Mostly employed (i.e., 0101101)	0.014	0.030	10.650	0.014	0.090	0.014	0.014	0.013	0.015
3 years employed, close (i.e., 0111001)	0.004	0.006	2.640	0.004	0.080	0.004	0.005	0.006	0.004
3 years employed, far (i.e., 0100111)	0.001	0.004	5.570	0.001	0.110	0.001	0.001	0.001	0.001
Program History (last 3 years), Sequences									
Often in programs	0.012	0.034	14.970	0.012	0.260	0.012	0.012	0.014	0.012
No programs	0.911	0.774	38.420	0.910	0.380	0.911	0.910	0.907	0.911
History of Wages While Employed (measured as average daily wages)									
Real wage (t-1)	67.435	58.960	27.860	67.501	0.200	67.354	67.889	58.196	71.637
Real wage (t-2)	61.086	48.079	36.580	61.169	0.220	60.979	61.665	50.649	65.550
Real wage (t-3)	54.875	44.204	27.780	54.815	0.200	54.835	55.120	44.087	59.399
Real wage (t-4)	49.820	43.230	16.930	49.679	0.350	49.700	50.493	39.210	54.133
Real wage (t-5)	45.191	40.172	12.790	45.090	0.250	45.137	45.514	34.742	49.441
Real wage (t-6)	41.583	37.529	11.290	41.503	0.210	41.497	42.045	31.417	45.675
Real wage (t-7)	39.530	36.242	10.120	39.453	0.200	39.378	40.346	29.289	43.470
N	50,796	82,397				42,331	8,465	10,976	26,721

Omitted Categories:
Mostly employed in last Period: Mostly Employed
Mostly unemployed in last period: 3 years unemployed (far) and Mixed Employment
History of programs (last 3 years): Seldom in programs

Table 4: Means and Standardized Differences (SD) for Regional Characteristics

	Treatment-group	Control-group	SMD before Matching	Matched Controlgroup	SMD after Matching	Voucher redeemed	Voucher expired	Degree Courses	Long-term Courses
Unemployment and Population									
Unemployment rate ´	12.195	12.842	12.31	12.221	0.504	12.255	11.907	12.745	12.430
Share of male unemployed	0.565	0.561	10.332	0.565	0.292	0.564	0.568	0.563	0.565
Share of German unemployed	0.858	0.871	14.674	0.858	0.437	0.859	0.851	0.868	0.857
Share of vacant fulltime jobs	0.794	0.789	6.586	0.794	0.196	0.794	0.795	0.790	0.793
Population per km^2	590.595	560.973	3.850	591.575	0.179	566.358	714.376	532.299	632.596
Industries									
Management of forests and agriculture	0.012	0.013	16.829	0.012	0.515	0.012	0.011	0.013	0.012
Fishing	0.005	0.005	4.070	0.005	0.161	0.005	0.005	0.005	0.005
Mining	0.010	0.010	3.477	0.010	0.240	0.010	0.010	0.010	0.010
Energy and water supply	0.064	0.067	14.450	0.064	0.428	0.064	0.062	0.066	0.064
Construction	0.150	0.150	2.693	0.150	0.127	0.150	0.149	0.149	0.150
Trade	0.028	0.028	3.265	0.028	0.224	0.028	0.028	0.029	0.028
Hotels and Restaurants	0.056	0.057	9.124	0.056	0.403	0.056	0.055	0.057	0.056
Transport and Communications	0.038	0.037	7.663	0.038	0.249	0.038	0.039	0.037	0.038
Bank and insurance business	0.118	0.116	5.452	0.118	0.215	0.117	0.120	0.116	0.120
Real estate activities	0.065	0.067	12.416	0.065	0.265	0.065	0.065	0.067	0.065
Public administration and defense	0.040	0.043	12.124	0.041	0.518	0.041	0.040	0.041	0.041
Education	0.118	0.117	3.118	0.118	0.125	0.117	0.118	0.118	0.118
Healthcare and social sector	0.047	0.047	3.795	0.047	0.207	0.047	0.048	0.047	0.048
Services	0.001	0.001	13.367	0.001	0.507	0.001	0.001	0.001	0.001
Production at the household level	0.001	0.001	2.630	0.001	0.324	0.001	0.001	0.001	0.001
Extraterritorial organizations and bodies	0.000	0.000	5.766	0.000	0.207	0.000	0.000	0.000	0.000
Other	0.000	0.000	8.644	0.000	0.310	0.000	0.000	0.000	0.000
N	50,796	82,397				42,331	8,465	10,976	26,721

Omitted Categories:
Industries: Manufacturing industry

Table 5: Balancing Test (Smith and Todd, 2005)

Elapsed Unempl. Duration (in months)	**Weighted Obs**	**Treated**	**Number of Parameters**	**# sign.**
1	2,151,575	8,419	106	9
2	2,037,131	4,497	106	4
3	1,861,567	4,721	106	7
4	1,707,959	4,664	106	6
5	1,586,653	4,554	106	7
6	1,491,415	4,355	106	5
7	1,403,392	4,131	106	9
8	1,332,685	3,873	106	6
9	1,266,373	3,509	106	10
10	1,204,959	3,241	106	4
11	1,151,255	2,718	106	5
12	1,097,295	2,114	106	2
			1,272	74

Table 6: F-statistics for Instrument in First Stage

	Elapsed unemployment duration (in months)					
	1	2	3	4	5	6
F-statistic	2762.82	1077.72	2053.54	2088.80	2486.04	2442.94
Adj. F-statistic	15.35	5.99	11.41	11.60	13.81	13.57
No. Treated	8,419	4,497	4,721	4,664	4,554	4,355
No. Wght. Obs	2,151,575	2,037,131	1,861,567	1,707,959	1,586,653	1,491,415

	Elapsed unemployment duration (in months)					
	7	8	9	10	11	12
F-statistic	2134.11	2891.15	3178.19	3163.80	3242.71	2657.31
Adj. F-statistic	11.86	16.06	17.66	17.58	18.02	14.76
No. Treated	4,131	3,873	3,509	3,241	2,718	2,114
No. Wght. Obs	1,403,392	1,332,685	1,266,373	1,204,959	1,151,255	1,097,295

The F-statistic refers to the test of the significance of the fitted treatment probability in the first stage of the IV estimates. The adjusted F-statistic is the F-statistics divided by 180 (number of employment offices minus one).

Table 7: Yearly Treatment Effects

	Desc. Difference	**OLS**	**IV**	**Desc. Diff - OLS**	**Diff. IV-OLS**
	Effects on Employment Probability				
year 1	**-0.085** (0.001)	**-0.097** (0.002)	**-0.145** (0.037)	**0.012** (0.001)	-0.048 (0.037)
year 2	**-0.087** (0.003)	**-0.126** (0.003)	**-0.180** (0.057)	**0.039** (0.001)	-0.055 (0.057)
year 3	**-0.031** (0.003)	**-0.078** (0.003)	**-0.147** (0.058)	**0.047** (0.002)	-0.069 (0.058)
year 4	**0.038** (0.003)	**-0.011** (0.003)	-0.087 (0.060)	**0.049** (0.002)	-0.075 (0.060)
	Effects on Monthly Earnings				
year 1	**-164.72** (3.55)	**-220.20** (3.93)	**-389.59** (128.11)	**55.48** (2.39)	-169.38 (127.33)
year 2	**-97.72** (5.76)	**-247.55** (5.76)	**-280.84** (122.95)	**149.83** (3.98)	-33.29 (122.63)
year 3	8.82 (6.07)	**-169.92** (6.08)	-202.20 (133.58)	**178.75** (4.27)	-32.28 (133.54)
year 4	**132.26** (6.22)	**-58.48** (6.22)	-89.48 (138.21)	**190.75** (4.38)	-31.00 (138.15)

Bold font indicates significance at 5% level. Wald test statistics for the joint significance of the difference between IV and OLS over several years imply for employment a p-value = 0.558 over years 1 to 4 and a p-value = 0.562 over years 2 to 4 and for earnings a p-value = 0.661 over years 1 to 4 and a p-value = 0.989 over years 2 to 4.