Employment and Earnings Effects of Awarding Training Vouchers

Annabelle Doerr
University of Freiburg

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

Thomas Kruppe
IAB, Nuremberg

Marie Paul
University of Duisburg-Essen
RGS Econ

Anthony Strittmatter
University of St. Gallen
University of Freiburg

April 9, 2014

Abstract

In 2003, Germany moved from a system in which participants in public sponsored training programs are assigned by caseworkers to an allocation system using vouchers. Based on data for all vouchers and actual program participation, we provide IPW, OLS, and IV estimates of the employment and earnings effects of a voucher award. Our results imply that voucher recipients experience strong and lasting negative lock-in effects. On average, there are only small positive employment effects and no earnings gains 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

---

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 by the IAB. The usual caveats apply.

Corresponding Author: Bernd Fitzenberger, Dept. of Economics, University of Freiburg, 79085 Freiburg, Germany. E-mail: bernd.fitzenberger@vwl.uni-freiburg.de, Tel: +49-761-203-2332, Fax: +49-761-203-2340.

1 Introduction

Public sponsored further training is an important part of active labor market policy (ALMP) in many countries. Such programs aim at skill enhancement to improve the labor market chances of participants. In 2003, Germany moved from a system in which participants are assigned to public sponsored 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 office specifies the content of the training program, for which the voucher can be redeemed. During the years 2003 and 2004, caseworkers are urged to award a training voucher only when it can be expected that the probability to find a job after training participation lies above 70%. Allowing for more choice by 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 informed enough 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. In addition, we use regional differences in policy style to provide IV estimates.

The Adult and Dislocated Worker Program under the Workforce Investment Act (WIA) in the United States and the German Training Vouchers are two major cases using vouchers for the provision of public sponsored training.[1] In 2003, the German government spends more than 6.5 billion Euros for further training programs 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 as specified by the

[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). Traditionally, education vouchers are mostly used in the schooling system (Posner et al. 2000) and (Ladd, 2002, for a review of the literature on school vouchers).

voucher.

In the U.S., customers in the WIA program can finance participation in a training program through their government funded Individual Training Accounts (ITA). The choice is restricted to eligible training providers who offer occupational skills in demand in the local labor market. There exist several studies on the ITA's involving descriptive evidence, experimental evidence, or qualitative evaluations of the implementation (see Barnow, 2009, for an overview). In the 1970's, 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 traning the participant would enroll in, and a third group receiving counseling and a 100% subsidy. Even though the subsidy lead 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. Individuals, who were to receive an ITA under the WIA, were randomly assigned to three different treatments regarding the freedom of choice of the customer and the counseling requirements. Important findings of this experiment are that the caseworkers tend to find counseling difficult and never reject the choice of WIA customers. In fact, WIA customers rarely participate in counseling when it not mandatory and they are more likely to participate in training with less counseling requirements (McConnell et al. 2011). In the long run, employment rates do no differ across treatment groups, but those being assigned to the treatment with less freedom of choice and more counseling requirements benefit from long-run earnings gains (Perez-Johnson et al. 2011).

Heinrich et al. (2010, 2011) provide a large scale econometric evaluation of the services provided by the Adult and Dislocated Worker Program under the WIA. These studies find positive earnings effects for participation in further training programs allocated through ITA's. Individuals eligible for training may actually choose not to participate in a training program and are thus not included in the treatment group. Similarly, 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 participating in training programs after the reform in 2003 on employment and earnings 1.5 years after program start. Rinne et al. (2013) do not observe the award of vouchers itself. They define treatment as participation in a training program which is allocated using a voucher. Hence, individuals not redeeming a voucher are in the control group. When the treatment is defined as the start of a training program, this requires different assumptions to identify a causal effect than when the treatment is defined as the voucher award. In the former case, the researcher must account for the dynamic selection both into voucher award and actual participation afterwards. Moreover, the dynamic selection involved with the redemption decision may question the no-anticipation assumption typically invoked when applying a dynamic matching approach to align treated and controls in time at the start of the actual training program.

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 public sponsored training program as an intention-to-treat.[2] From a policy perspective, it is the effect of a voucher award, which 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 for not redeeming a voucher. Our empirical analysis proceeds in two steps. First, we apply a matching strategy, which accounts for selection based on observable characteristics. This approach is implemented using both inverse probability weighting (IPW) and ordinary least squares (OLS) regressions. 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 treatment start versus no treatment start now (treatment versus waiting/continuing

[2] The exists 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 mixeds, see Card, Kluve, and Weber (2010) for a recent review.

to search) for each month of elapsed unemployment. Second, we implement an instrumental variable (IV) approach exploiting the variation in the regional conditional allocation intensity, which we take as representing differences in policy styles across regional employment offices.

Our study uses unique rich administrative data provided by the Federal Employment Agency in Germany. We have information on *all* individuals who receive 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, information which has not been previously available for evaluation studies.[3] 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. 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. OLS and IPW lead to virtually the same results. A comparison to raw differences between the treatment and control group shows a strong positive selection of voucher recipients with respect to observable characteristics. The IV estimates are somewhat imprecise and do not differ significantly from OLS. However, the point estimates for IV point towards further positive selection with respect to unobservables. 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 causal 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

[3]These data are used by Kruppe (2009) to investigate the heterogeneity in the probability of redeeming a voucher.

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 using alternative methods as well as effect heterogeneity in Section 5. The final section concludes.

2 Background

Before 2003, public sponsored training in Germany involved the direct assignment of the unemployed by caseworkers to a specific training provider and further training course. At the time, the political debate was concerned that public sponsored training was not effective and that this might have been related to the tight relationships between employment offices and training providers. The First Modern Services on the Labor Market Act (the so-called *Hartz* I Reform) introduces a voucher system for the provision of public sponsored training in January 2003. It aim is to foster market mechanisms and transparency in the training market.[4]

During an unemployment spell, individuals repeatedly meet their caseworker for counseling. The caseworker may award a training voucher to the unemployed when training is considered necessary. The caseworker denotes the objective, content, and maximal duration of the course on the voucher. The unemployed may then choose a course offered by an eligible training provider which is located within a one-day commuting zone subject to the restrictions denoted on the voucher. It is thus the task of the caseworker (potentially in discussion with the unemployed) to decide upon the training objective and the educational content of the course. The unemployed may choose the provider and the particular course. Eligible (certified) training providers are listed in an online tool provided by the employment office and providers may also advertise their courses e.g. by placing handouts in the employment office.[5] The caseworker is not supposed to

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

[5]In 2003 and 2004, the Federal employment office is in charge of the certification of the eligible

give any advice on which provider to choose, which is a response to the concern that relationships between the employment office and training provides were too tight before 2003. Training vouchers are valid for at most three months, so they can only be redeemed during this period. The German voucher system differs from the WIA system in the U.S. with regard to who makes which decision. Under the WIA, the voucher recipient faces 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 eligible provider. The choice of the content of the training is up to the voucher recipient. But in order to receive a voucher, individuals are typically required to participate in counseling which involves an assessment of skills, research on training programs and the labor market, and one-to-one discussions with the caseworker on which course to choose (McConnell et al. 2011, King and Barnow 2011). In contrast to Germany, the voucher recipient in the U.S. receives guidance on how to use the voucher but may finally decide him/herself upon the content of the training. So after a guided and mandatory decision process, the voucher recipient may decide, say, to enroll into training to become an IT specialist instead of a care nurse. In Germany, the voucher recipient may state his preference (say to become an IT specialist) before the voucher award, but finally the caseworker decides upon the content of training. And then, after the award of the voucher, the German unemployed receives no guidance by the caseworker on which training course to choose.

Further training programs are used to adjust the skills of the unemployed to changing requirements of the labor market and possibly to changing individual conditions of employability (due to health problems for example). They aim at improving the human capital and productivity of the participants. Participation prolongs the entitlement period for unemployment benefits. Further training mainly comprises long-term training and degree courses. Long-term training courses typically last several months to one year (on average five months in our sample) and usually involve full-time programs. Teaching takes place in class

training providers. Afterwards, the certification process is privatized.

rooms or on the job in training firms. The course curriculum may also include internships. Typical examples of further training schemes are courses on IT based accounting or on customer orientation and 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 vocational training for a care-assistance for the elderly or an office clerk. Even though the Federal employment office covers the costs typically for at most two years, these programs may last for three years and further programs exist (e.g. sponsored directly by the state government) which cover the additional costs.

Besides the opportunity to take part in an intensive further training program, training vouchers may influence future labor market opportunities through various channels (see for example Barnow, 2000, 2009, Hipp and Warner, 2008, for 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 training participants and should introduce market mechanisms into the provision of public sponsored training. The first main difference to 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 which act on 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 know better which training providers offer the best programs and which courses are most suitable for a particular unemployed. Furthermore, the choice by the unemployed may be driven by concerns unrelated to the effectiveness of the training program and some individuals may feel incapable or incompetent to find a suitable course, which may have negative effects on the motivation. The increased course choices may unfold positive effects from the providers side. One would expect that competition about potential clients will lead to a positive selection of providers remaining on the market and efficient behavior of the remaining providers. To assure that training providers offer courses that are in line with the regional labor demand, the employment offices have to plan and publish their regional and sector-specific demand once a year.[6]

A second difference to the old system is that the caseworker should not impose a sanction when a voucher is not redeemed and the unemployed provides a reasonable explanation for that. After redemption, however, training participation is mandatory. The freedom not to redeem the voucher may change the attitude of the unemployed towards this service perceiving it more like an offer and less like an assignment. This could exert a positive attitude effect such that the unemployed may value that a costly service is offered to him or her and reciprocates by increasing 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. The employment offices are supposed to award vouchers such that, according to their prediction, at least 70% of the voucher recipients will find a job within six months after training.[7]

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 receive a training voucher in 2003 or 2004. The data are generated from internal administrative data and includes precise award and redemption dates for each voucher - information which has not been previously 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). The IEB is a rich administrative data base which is the source of the subsamples of data used in all recent years studies evaluating German ALMP. It is a merged

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

[7] This 70% rule was abolished after the time period considered here.

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*). 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.[8] 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.

Our control persons are from the same data base: A three percent random sample (based on twelve days of birth of 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.[9] We only consider unemployed individuals who are eligible for unemployment benefits.[10] This sample choice reflects the main target group for the training

[8]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.

[9]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.

[10]Note that, in particular, this condition excludes training programs for mothers returning to the

vouchers. In order 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 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 a 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 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, whereof 50,796 individuals are awarded with a voucher during their first twelve months of unemployment and 82,397 observations are in the control group. 42,331 individuals in our sample redeem their vouchers. This results into a redemption rate of 83%. We observe 8,465 vouchers that are awarded but not redeemed.[11]

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 show the mean value of the respective control variable in the treatment and in the control subsample. In column six and seven we distinguish between those who redeem the voucher and those who do

labor market after longer employment interruptions.

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

not. Voucher recipients are on average more often middled-aged, single or single-parent and females than individuals in the control group. They show less health problems. Individuals who redeem the training voucher and thus participate in a training course are on average slightly older and more healthy 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 past 7 years, 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 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 show problems of this kind. The fraction of people with motivation deficits or past incapacities is even lower for individuals who redeem the voucher.

4 Identification and Estimation

We consider voucher awards during the first twelve months of unemployment in the first unemployment spell between January 2003 and December 2004. Each unemployed is observed for at least 48 months. Denote the indicator for a voucher award as an intention to treat 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 months since the voucher is awarded). 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. In the results section, we report a weighted average of the twelve monthly dynamic treatment effects (see Appendix A for details).

4.1 Approaches Conditioning on Observables

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

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

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

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}.[12] This is formalized by the following dynamic version of the conditional mean independence assumption.

Assumption 1 *(Strong Ignorability).*

i) Dynamic mean independence assumption:
$E[Y^0_{imt}|D_{im} = 1, X_{im} = x] = E[Y^0_{imt}|D_{im} = 0, X_{im} = x]$ 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 $t = 1, ..., 48$.

Assumption i) implies that after controlling for X_{im} there are no other variables that jointly influence the expected value of Y^0_{imt} and D_{im}. There are many

[12] This assumption is well known as strong ignorability, conditional unconfoundedness, selection on observables or conditional independence assumption (see e.g. Imbens, 2004).

factors that affect the probability of a voucher award and future labor market outcomes, e.g. age, health, family status and employment history. Because of our rich administrative data set that includes detailed information on the personal and socioeconomic characteristics, information about past employment and welfare histories as well as regional information, it may well be the case that we observe all relevant confounding factors. There are many matching studies which use the same or very similar administrative data. In particular, Biewen et al. (2014) and Lechner and Wunsch (2011) investigate the plausibility of the conditional independence assumption. These studies suggest that our data set is rich enough to satisfy this assumption for actual training participation in the time period before the introduction of training vouchers.

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. Given Assumption 1,

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

is identified from 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. Taking IPW as a benchmark, we specify our parametric OLS regressions to allow for sufficient flexibility.

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 $m = 1, ..., 12$ 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 treated and nontreated are well matched with respect to observables (see Appendix B for details).

Although IPW has some optimality properties, some critical issues may arise. First, 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 a estimation of an appropriate specification for the treatment probability (we rely on probit estimates). To show 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. Even though the implicit parametric assumptions may not hold, OLS might provide a good estimate of average treatment effects.[13] Since 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 as IPW.

4.2 Instrumental Variable Approach

In the case of selection into treatment based on factors unobserved by the researcher, an instrumental variable (IV) approach may be appropriate to obtain consistent estimates of the treatment effects (for the subset of compliers in the

[13] 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 OLS finds exactly the conditional expectation function in fully saturated models, thus providing the fully nonparametric estimates for such a case.

random coefficients case). We use an IV approach to assess the impact of selection on unobservables. 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. 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. In our case, the variation in the conditional employment district specific allocation intensity, that 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 just 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 1 shows the differences in unconditional award intensities across employment office districts in Germany. In some areas of Germany, there exists 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. Furthermore, they decide upon the targetting of training vouchers. The differences in voucher award intensities can partly be explained by differences in attitudes of caseworkers in different employment offices. This may hold in particular for the first two years after the reform, when views about the introduction of vouchers varied strongly across employment offices.[14] To understand this, one has to recall that the implementation of training vouch-

[14]Brenke et al. (2006) provide information about the reform assessment by employees of employment offices soon after the implementation of the reform and Doerr and Kruppe (2012) for an assessment at a later date.

ers resulted in a major change of the assignment into further training both for training participants and caseworkers in the employment offices. Prior to the reform, the caseworkers had been responsible for the assignment of the unemployed to training courses. After the reform, the unemployed have the choice of training provider and a voucher award should be restricted to unemployed for whom the employment rate is predicted to be at least 70% after training participation. Thus, the reform involved a loss of discretionary power, possibly leading to negative attitudes regarding the reform. A positive impact can be seen in the reduced work burden for the caseworkers regarding actual course choice. Furthermore, some caseworkers supported the idea of freedom of choice for the unemployed while others have been skeptical. Finally, the caseworkers may differ in their assessment of which groups are helped best with the award of a training voucher (e.g. less educated or more educated unemployed). We assume that such variation in the award of training vouchers partly reflect differences in the exogenous policy style.

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) for the dummy variable voucher award to individual i in month m

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

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 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 5 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 mostly 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, 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 typically rule-of-thumb threshold in the literature) in 10 out of 12 months.[15] Nevertheless, our subsequent results show that our IV estimates of the treatment effects by month after treatment involve a fairly large estimation error. For this reason, we also report IV estimates averaged by year since treatment, and we do not provide IV estimates accounting

[15] Because our IV approach builds on full interactions between regions and the other characteristics considered, we are in a case with many instruments. However, our data set is very large as well. In the second stage, we provide estimates which are pooled at the national level. The standard rules-of-thumb applied in the IV literature (F-Statistics above 10) do not apply to this case.

for effect heterogeneity.

5 Results

We first discuss OLS, IPW, and IV estimates of the average treatment effects for the treated. Then, we investigate the heterogeneity of the treatment effects across skill groups and across type of training programs based on 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 OLS and IPW 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 voucher receipt and on the vertical axis the outcome variable. Triangles and 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 2 depicts the descriptive (unconditional) differences between the treated and nontreated (top line) together with the average treatment effects based on different estimators (OLS, middle line; IV, bottom line). The OLS results imply a very long and pronounced lock-in effect. It takes about 40 months until the negative effect turns 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 (about 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 not depicted in Figure 2, because they are basically the same as those obtained using OLS. This can be seen in Figure 3 which shows the differences between the results of the different estimators. This suggests that we use sufficient flexibility in our specification of the OLS regression.

Figure 2 evidently shows that there are strong changes in the slopes of the treatment effect at about 12 to 14, 24 to 26, and 36 to 38 months. This can be explained by the fact that many programs have a duration of 12, 24 or 36 months and most treated individuals enter training within the first two months after receiving the voucher (Figure 4). Figure 5 shows 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 three years after treatment. It takes 40 months after treatment until the employment effect turns positive.

The descriptive effect in Figure 2 involves a shorter and less pronounced lock-in effect than 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 or prior dropouts 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 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.

5.1.2 IV Results

Figure 2 also includes the IV effect estimates both for employment and earnings. Being quite noisy, the monthly IV estimates are often not significant pointwise. The IV estimates are typically more negative than the OLS estimates, and the comparison of the point estimates would imply strong positive selection based on unobservables. This would be consistent with the hypothesis that, in addition to cream skimming behavior with regard to observables as discussed above, there is also such behavior with regard to characteristics that are unobservable to the researcher, but observable to the caseworker. However, the difference between monthly OLS and IV estimates (Figure 3) is significant only for a few months for employment and never for earnings. To gain precision, we additionally consider average effects by year since voucher award (Table 7). The yearly IV employment (earnings) effects are significantly negative during the first three (two) years. While the descriptive difference is positive in the fourth year (first column of Table 7), the treatment effects estimated by OLS and IV (second and third column) remain negative and insignificant in the case of IV. The second last column shows 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 above. This is also the case for earnings. The last column shows the difference between IV estimates and OLS estimates. The difference is consistently negative, but never significantly so. Also 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 show significant differences. Thus, even for yearly treatment effects, there are no significant differences between the OLS and the IV estimates.

Summarizing, 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 methods (OLS and IPW) based on a selection on observables assumption basically provide the same results. Raw employment differences show 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 years and they have earned higher wages in their previous jobs). Our IV strategy does not provide precise estimates of the treatment effects, and the IV estimates are not significantly different from the OLS estimates. However, the IV point estimates are consistent with a further positive selection based on unobservables. 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 IPW results and the descriptive differences. We do not report IV results (these results are available upon request), because they are even more imprecise than the IV results reported above and because the IV results reported above did not differ significantly from the OLS estimates. As a caveat, based on the estimates reported above, it is not unlikely that the OLS estimates are upward biased, but the evidence for such a bias is not significant.

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 strongly by a major investment into their human capital and by obtaining a course certificate or even a vocational degree. 22% of the treated in our sample do not hold a vocational degree (low skilled individuals). 11% of the treated are high skilled, holding an academic degree. The majority of the treated holds a vocational degree (medium skilled). The top line in Figure 6 depicts the effect of a voucher award for the group of those without a vocational degree. The lock-in effects lasts for about 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 almost 6 ppoints and a significant positive earnings effect of about 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 medium skilled. To substantiate the findings, Figures 8 to 10 show that the IPW estimates are very close to the OLS estimates.

Can we say more on as to why only low skilled individuals benefit on average? A potential explanation would be that 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 about 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 levels for the medium skilled. After month 25, the line for the low skilled increases more strongly. This is the time at which participants

[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 a slightly more positive at the end of the observation period.

of 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 about 25 to 36 months after the voucher award, and Figure 6 shows the strongest increase for the low skilled during that time. These results suggest that the low skilled voucher recipients eventually do much better in finding a job compared to the medium skilled. To substantiate this finding, Figure 7 shows the employment rates of the treated and matched controls by skill level. After 36 months, the treated low skilled exhibit almost 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 is in contrast to results reported in the literature. Rinne et al. (2013) 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 few 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 1990's are an exception as they have an extraordinary 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.

[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 online appendix of Biewen et al. 2014).

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 (mostly retraining). 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 show 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 almost 36 ppoints and earnings by over 600 Euro per month. But after 48 months the employment effect is 8 ppoints and earnings gains are relatively large with over 100 Euro per month (Figure 11). So 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 about 12 months. This lock-in period is comparable to Rinne et al. (2013). But after this pronounced lock-in period, the estimated effects remain negative for the whole observation period even though the effect size is reduced over time. In contrast to our results, Rinne et al. find a positive employment effect of about 7 ppoints at the end of their observation period of 1.5 years after program start. In Rinne et al. 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 where comparable the time of the voucher award (among them some of those who did not redeem a voucher) may have found a job in the meanwhile and are thus excluded from the control group. This may induce an upward bias in the effect estimates.

Figures 12 and 13 compare the effect estimates for long-term training and degree courses obtained by different estimators. IPW is again very close to OLS. Interestingly, the difference to the descriptive effect is a little stronger for long-term courses than for degree courses (Figure 14) suggesting that the effect of cream skimming is stronger for long-term training. Correspondingly, also 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 (last two columns in Table 1 to 3) 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 show more positive long-term effects than long-term training. To shed further light on these findings, Figure 15 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 materialize 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 through allowing the individual to participate in a training program, but it may also have an effect on labor market outcomes itself. Figures 16 to 18 shows the effect estimates by the redemption decision. These OLS and IPW 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 (with 83% this is the majority among the treated) show the same pattern as for the effect for all treated. But 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 to be assigned to a mandatory active labor market program, like for example a job creation scheme, or to loose their entitlement to unemployment benefits, if they do not redeem the voucher and remain unemployed. But the positive effect may also be due to those individuals who receive a job offer quickly and who therefore do no redeem the voucher. 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. Even though 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. In addition, we use an instrumental variable strategy which exploits exogenous variation in the regional award intensity which we take as differences in the policy style of the employment offices.

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 almost the same results. A comparison to raw employment differences shows that with regard to observables voucher recipients represent a strong positive selection both regarding employment and earnings. The IV estimates are quite noisy and not significantly different from OLS. Thus, we focus on the OLS and IPW effect estimates. However, the IV point estimates would consistently imply further positive selection based on unobservables. 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 shows 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 which 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 do not do 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 which 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.

Brenke, K., A. Uhlendorff, A. Werwatz, F. Aust, R. Cramer, R. Gilbert, D. Hess, B. Jesske, K. Marwinski, A. Prussog-Wagner, H. Schröder, M. Smid, A. Steinwede, J. Steinwede, L. Kaiser, U. Rinne, H. Schneider, M. Schneider, and Z. Zhao (2006): "Evaluation der Massnahmen zur Umsetzung der Vorschäge der Hartz-Kommission * Modul 1b: Förderung beruflicher Weiterbildung und Transferleistungen.," *Bericht 2006 für das Bundesministerium für Arbeit und Soziales. Abschlussbericht und Kurzfassung.*

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.

Doerr, A., and T. Kruppe (2012): "Bildungsgutscheine und Zertifizierung aus Sicht der Arbeitsverwaltung: Umfrageergebnisse aus der Bundesagentur für Arbeit.," *IAB Forschungsbericht*, 6.

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.

Heinrich, C., P. Mueser, K. Troske, K. Jeon, and D. Kahvecioglu (2010): "New Estimates of Public Employment and Training Program Net Impacts: A nonexperimental Evaluation of the Workforce Investment Act program," *Working Paper No. 1003, Department of Economics, University of Missouri.*

——— (2011): "The Workforce Investment Act: Implementation Experiences and Evaluation Findings," in *The Workforce Investment Act: Implementation Experiences and Evaluation Findings*, ed. by D. J. Besharov, and P. H., pp. 371–404. Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.

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.

Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, 86(1), 4–29.

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.

Kruppe, T. (2009): "Bildungsgutscheine in der aktiven Arbeitsmarktpolitik," *Sozialer Fortschritt*, (58), 9–19.

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 C. Wunsch (2011): "Sensitivity of matching-based Program Evaluations to the Availability of Control Variables," *Unpublished Work*.

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

McConnell, S., P. Decker, and I. Perez-Johnson (2011): "The Role of Counseling in Voucher Programs: Findings from the Individual Training Account Experiment," *Unpublished Manuscript, Mathematica Policy Research, Princeton, NJ.*

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., 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 consist 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. In order 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 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 showing 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.

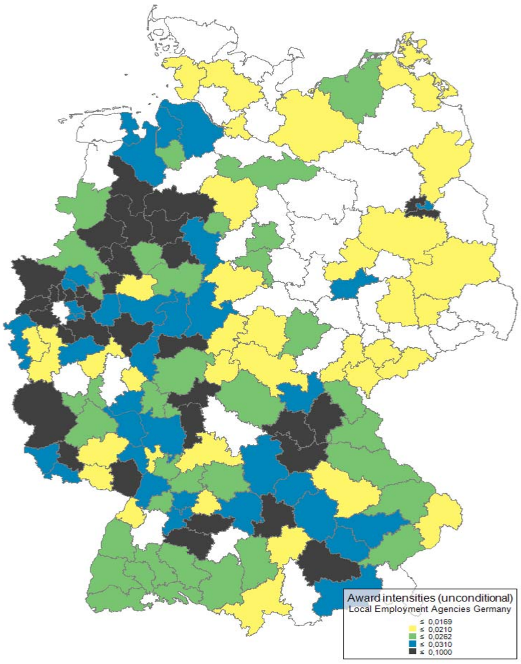
We also apply a second balancing test following an approach of Smith and Todd (2005). Therefore, we run the regression

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

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 6 we report the summarized results of the test for each of the twelve treatment times. Overall we run 1,272 regressions whereof 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 IPW estimations. Since 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.

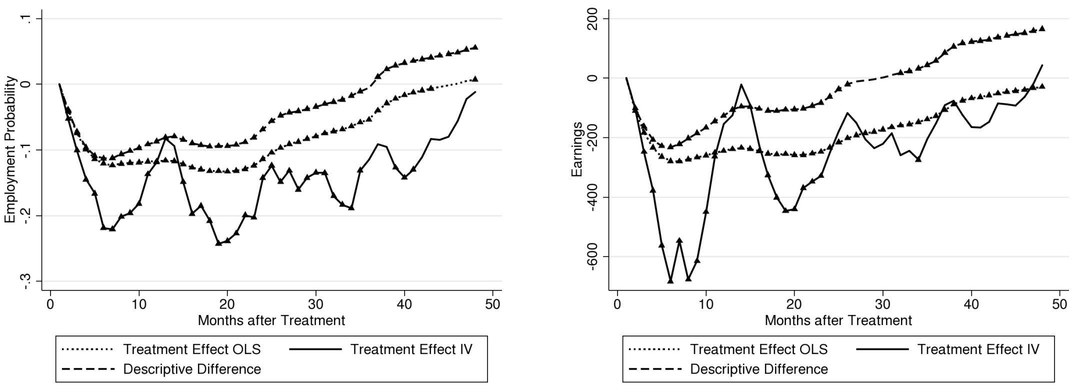
Figures and Tables

Figure 1: Regional Differences in Voucher Awards per Unemployed



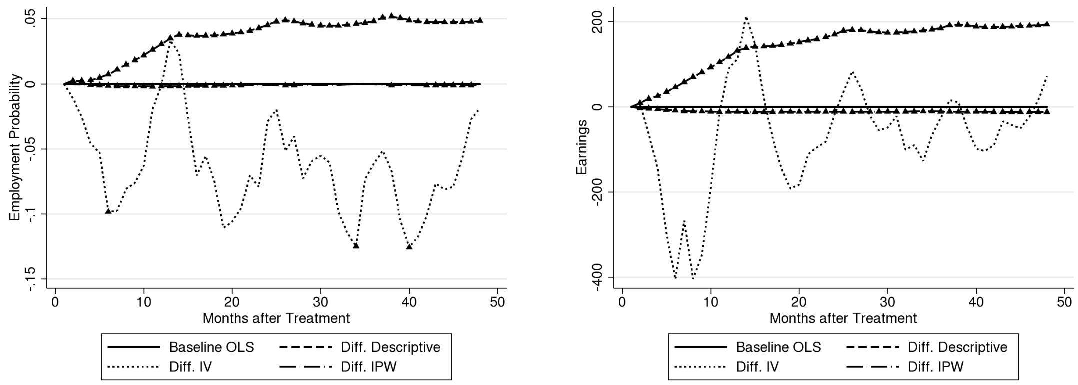
Notes: Differences in unconditional award intensities across employment office districts. Min= 0.08%, Max= 5.59%, Mean= 2.43%, Award Intensity = #Voucher Recipients/#Unemployed by District.

Figure 2: Effect of a voucher award on employment and earnings averaged over elapsed unemployment durations until treatment.



Triangles and diamonds indicate significant effects.

Figure 3: Differences in the estimated effects of a voucher award on employment and earnings averaged over elapsed unemployment durations until treatment.



Triangles and diamonds indicate significant effects.

Figure 4: Fraction of individuals in training after the award of a voucher.

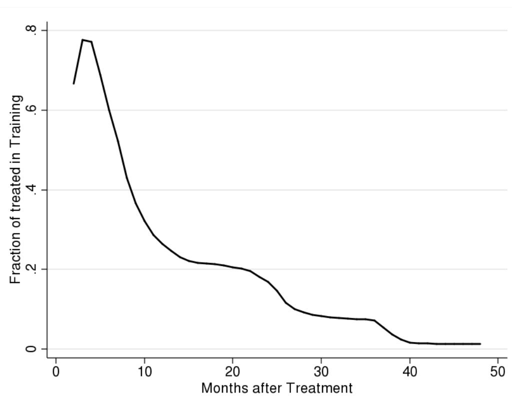


Figure 5: Comparison of average employment and average earnings between treatment and matched control group averaged over elapsed unemployment durations until treatment.

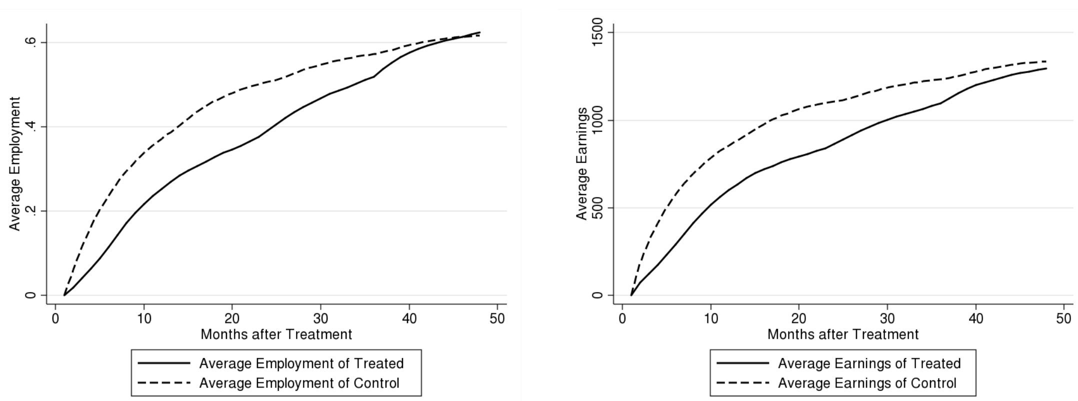


Figure 6: Heterogeneous effects on employment and earnings by skill group (OLS) averaged over elapsed unemployment durations until treatment.

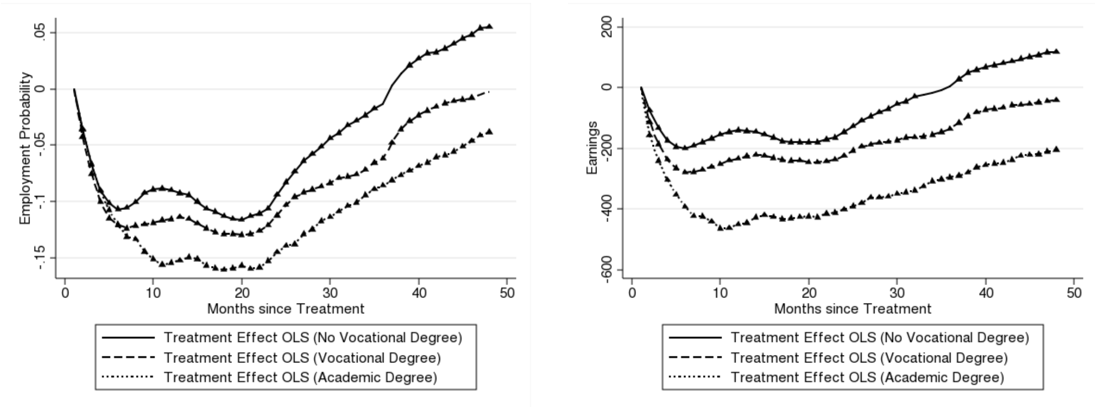


Figure 7: Comparison of average employment of treated and matched control group by skill group averaged over elapsed unemployment durations until treatment.

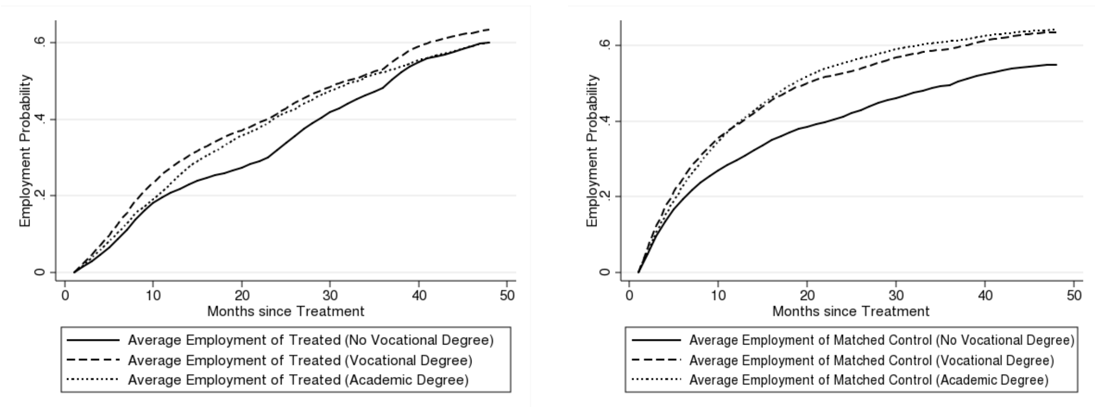


Figure 8: Effect of a voucher award on employment and earnings for individuals without vocational degree averaged over elapsed unemployment durations until treatment.

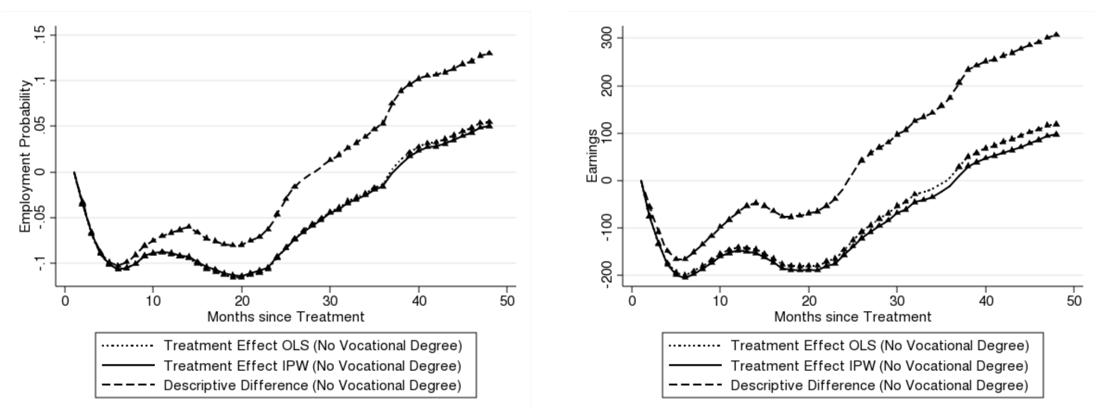


Figure 9: Effect of a voucher award on employment and earnings for individuals with vocational degree averaged over elapsed unemployment durations until treatment.

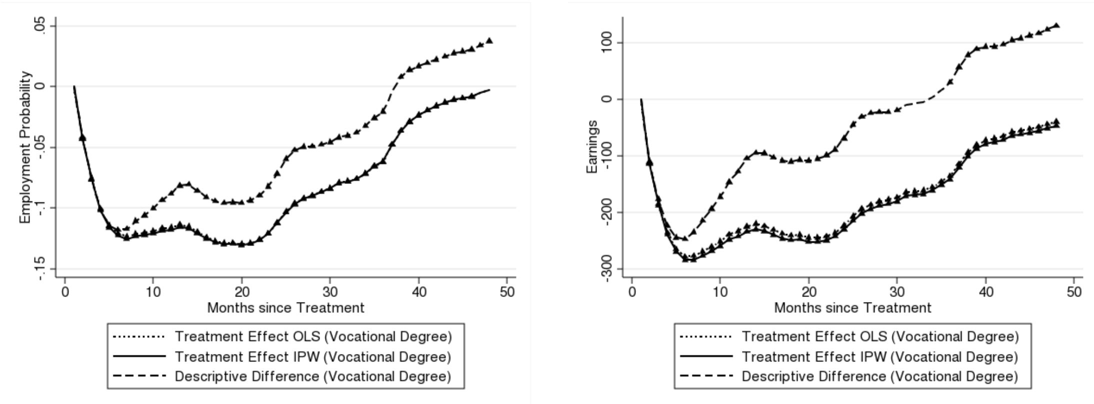


Figure 10: Effect of a voucher award on employment and earnings for individuals with academic degree averaged over elapsed unemployment durations until treatment.

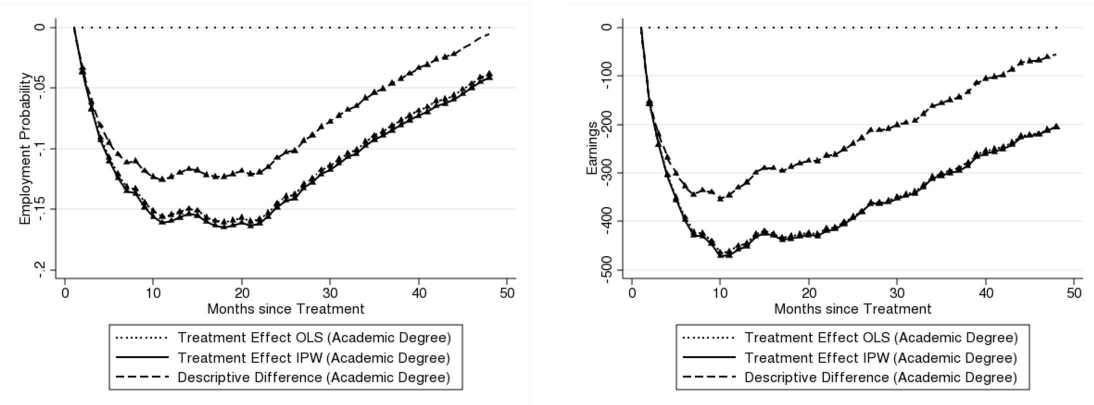


Figure 11: Heterogeneous effects on employment and earnings with regard to the type of training (OLS) averaged over elapsed unemployment durations until treatment

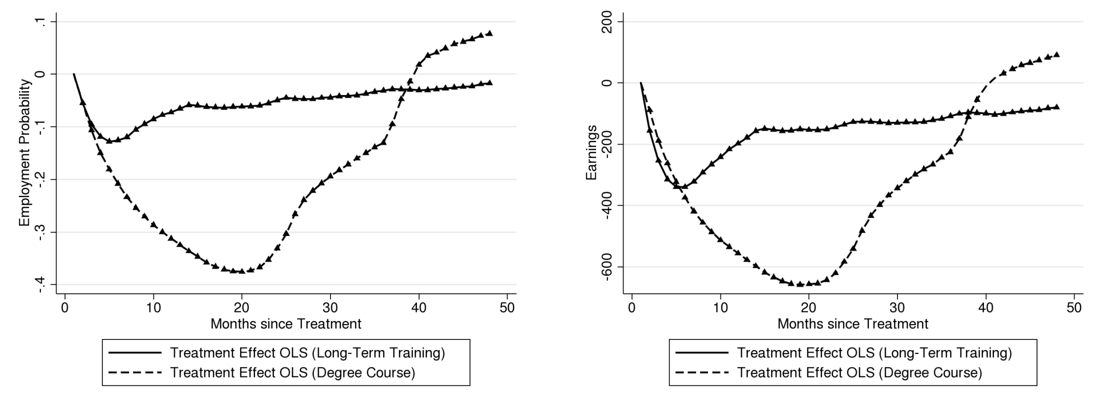


Figure 12: Effect of a voucher award on employment and earnings for individuals participating in long-term courses averaged over elapsed unemployment durations until treatment.

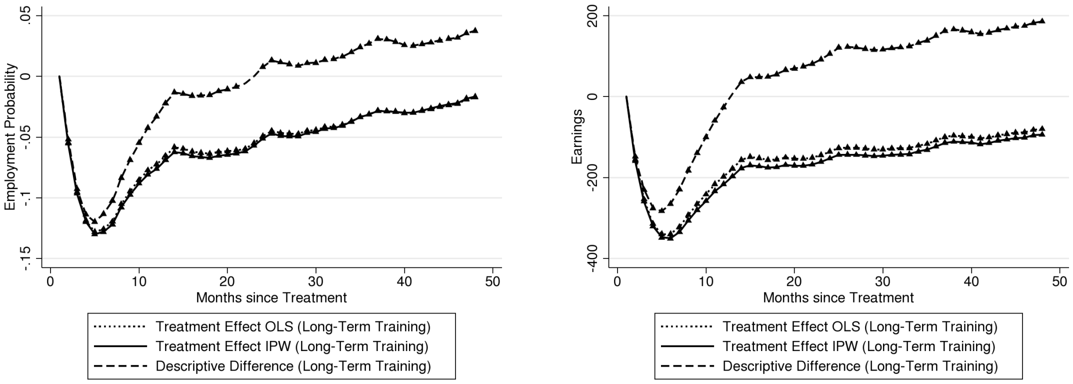


Figure 13: Effect of a voucher award on employment and earnings for individuals participating in degree courses averaged over elapsed unemployment durations until treatment.

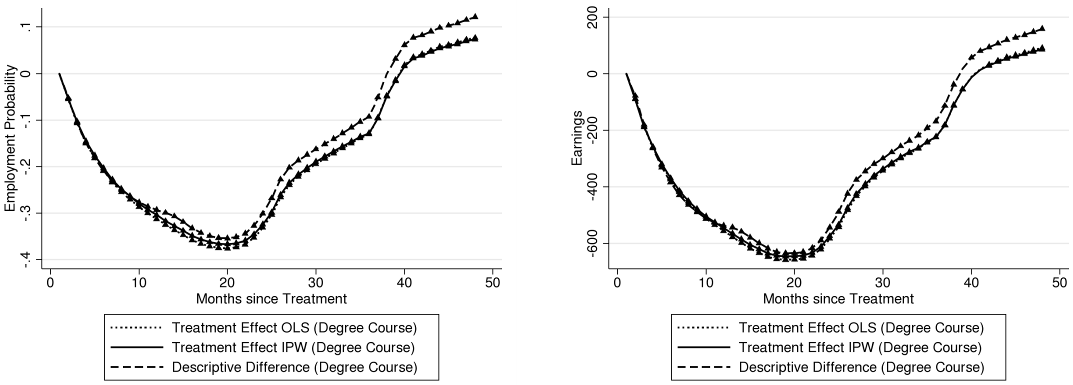


Figure 14: Comparison of average employment of treated and matched control group by course type averaged over elapsed unemployment durations until treatment.

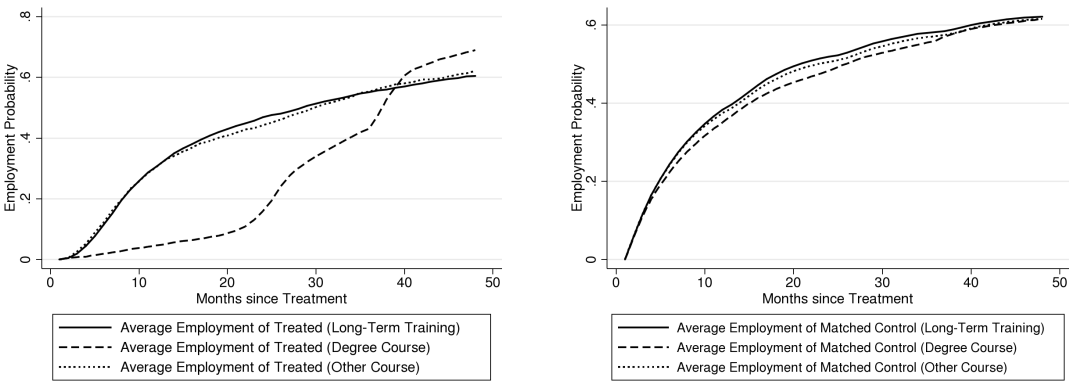


Figure 15: Heterogeneous effects on employment and earnings with regard to the type of training and vocational degree (OLS) averaged over elapsed unemployment durations until treatment

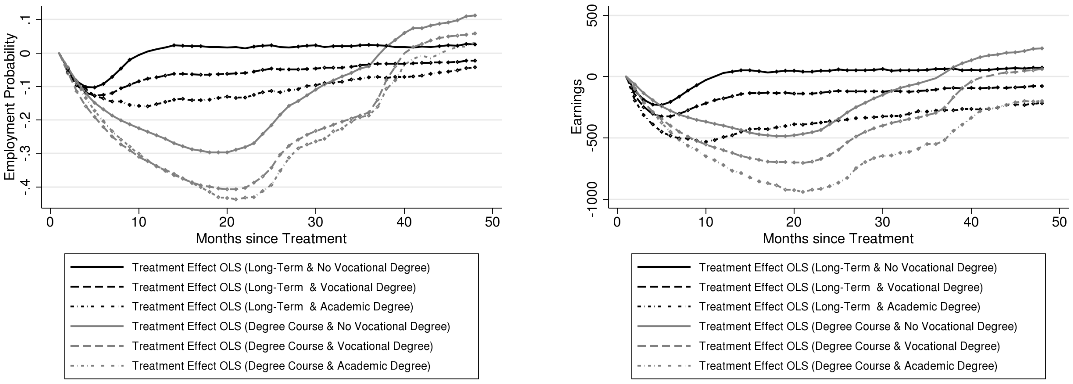


Figure 16: Heterogeneous effects on employment and earnings with regard to the redemption decision (OLS) averaged over elapsed unemployment durations until treatment.

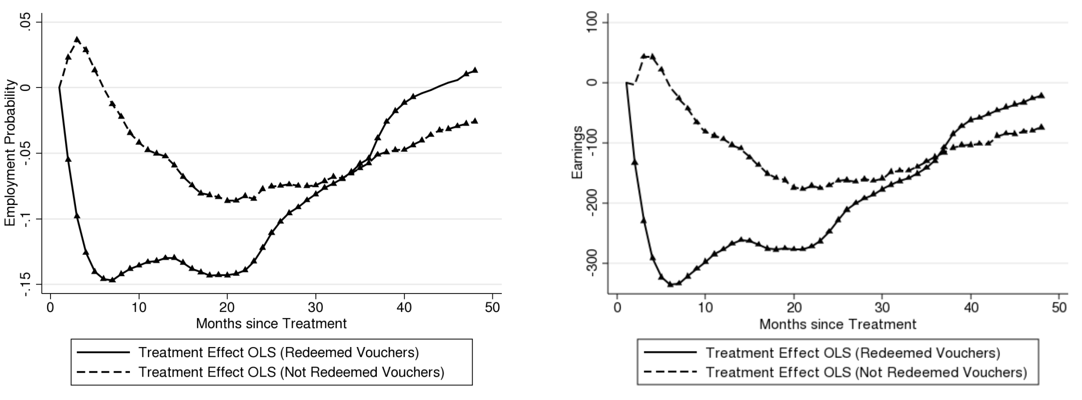


Figure 17: Effect of a voucher award on employment and earnings for individuals who redeem the voucher averaged over elapsed unemployment durations until treatment.

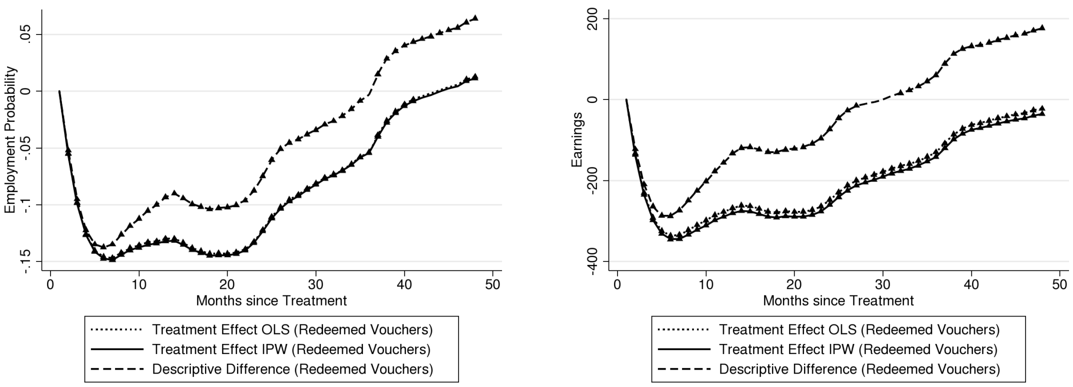


Figure 18: Effect of a voucher award on employment and earnings for individuals who do not redeem the voucher averaged over elapsed unemployment durations until treatment.

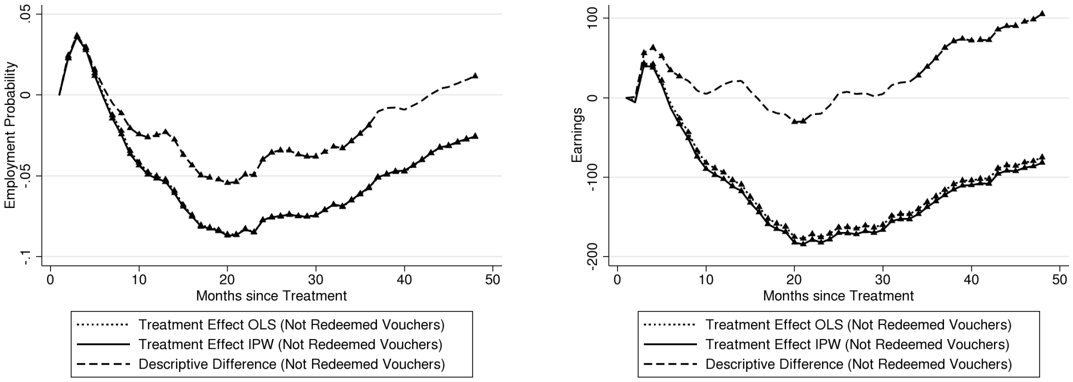


Table 1: Means and Standardized Differences (SD) for Personal Characteristics

	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
Abitur/Hochschulreife	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
Kind 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
Kind 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									
Mostly employed in last period									
Mostly unemployed	0.170	0.223	13.180	0.171	0.290	0.170	0.173	0.228	0.150
3 years employed (close)	0.131	0.095	11.280	0.131	0.100	0.131	0.132	0.135	0.127
3 years employed (far)	0.026	0.055	14.690	0.026	0.190	0.026	0.027	0.023	0.027
3 years unemployed (close)	0.012	0.025	9.969	0.012	0.120	0.012	0.011	0.010	0.012
3 years unemployed (far)	0.099	0.088	3.640	0.099	0.210	0.099	0.095	0.112	0.095
Mixed employment	0.049	0.061	5.430	0.049	0.170	0.049	0.049	0.053	0.047
Mostly unemployed in last period									
Mostly employed	0.014	0.030	10.650	0.014	0.090	0.014	0.014	0.013	0.015
3 years employed (close)	0.004	0.006	2.640	0.004	0.080	0.004	0.005	0.006	0.004
3 years employed (far)	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									
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 defence	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 householdlevel	0.001	0.001	2.630	0.001	0.324	0.001	0.001	0.001	0.001
Extraterritorial organisations 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: 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
	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 6: 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 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.