DISCUSSION PAPER SERIES

IZA

Vouchers and Caseworkers in Public Training Programs: Evidence from the Hartz Reform in Germany

Ulf Rinne
Arne Uhlendorff
Zhong Zhao

Vouchers and Caseworkers in Public Training Programs: Evidence from the Hartz Reform in Germany

Ulf Rinne
IZA and Free University of Berlin

Arne Uhlendorff
IZA and DIW Berlin

Zhong Zhao
IZA and Renmin University of China

ABSTRACT

Vouchers and Caseworkers in Public Training Programs: Evidence from the Hartz Reform in Germany[*]

This paper studies the role of training vouchers and caseworkers in public training programs. Using a rich administrative data set, we apply matching and regression methods to measure the effect of the Hartz reform in Germany, which introduced training vouchers and imposed more selective criteria on participants. Besides estimating the overall reform effect, we isolate the effect induced by changes in the composition of program participants due to stricter selection by the caseworkers (selection effect) from the effect based on the introduction of vouchers (voucher effect). Analyzing the most important type of training in Germany, we find a slightly positive impact of the reform. Our decomposition results suggest that the selection effect is − if at all − slightly negative, and that the voucher effect increased both, the employment probability and earnings of the participants.

Corresponding author:

Ulf Rinne
IZA
P.O. Box 7240
D-53072 Bonn
E-mail: rinne@iza.org

---

1 Introduction

Germany reformed its active labor market policy (ALMP) in a series of reforms which are commonly known as the Hartz reforms. When the first wave of these reforms (Hartz I/II) came into force on January 1, 2003 the provision of public training programs was substantially changed. The most important change was the introduction of a voucher scheme. The former contracting-out system was abandoned and replaced by a system in which job seekers are free to select their training provider in the market. Previously this choice was made by the caseworker. However, the participants are not completely free in their choice, because the content of the training is still assigned by the caseworker. Vouchers are a common instrument in other fields of public services, but this approach is rather novel in the context of delivering ALMP. In addition to the voucher scheme, a stricter selection of the participants by the caseworkers was introduced.

This reform provides us a valuable opportunity to study not only the overall training impact, which is the main parameter of interest in most of the literature on training, but also to investigate the effect of different components inside the training "black box". In this paper, we estimate the impact of the reform on the effectiveness of the most important training program type for the unemployed in Germany. We disentangle the effect induced by the introduction of vouchers (voucher effect) from the effect induced by a stricter selection by the caseworkers (selection effect).

Increased consumer choice and provider competition are the main arguments in favor of the introduction of vouchers (see, e.g., Steuerle, 2000). It is argued that the possibility to choose the training provider in accordance to their preferences will lead to better matches between the unemployed and training providers, which will increase the effectiveness of participation. In addition to that, greater freedom of choice may encourage more competition among the providers. Training providers may have to compete more if they must regularly face the demand of participants instead of having a longer-term contract with the employment agency. This could lead to a further increase of the match quality and therewith of the effectiveness of training programs. On the other hand, there may be obstacles in case of public training programs which counteract the potential positive impacts of vouchers. Generally, it is argued that the consumer—in our case the unemployed—may lack competence or resources to optimally choose, and that information asymmetries may lead to choices which do not truly reflect consumers' preferences. For example, the caseworkers may know more about the availability of training courses and the

quality of training providers than the unemployed, because of their experience with previous participants. For a discussion of advantages and disadvantages of training vouchers see Barnow (2000, 2008).

Vouchers have been widely used in other fields of public services—in particular in the field of education—and are quite extensively studied in the literature.[1] There exist some studies on vouchers for pre-school education (e.g., Viitanen, 2007), but most studies focus on school education (e.g., Manski, 1992; Nechyba, 2000; Angrist et al., 2002; Krueger and Zhu, 2004). Ladd (2002) presents a review of major studies on school vouchers. She concludes that the overall picture that can be drawn from these studies is rather inconclusive, and that the results are not very robust. For instance, studies with U.S. data typically indicate to have insufficient information to draw clear conclusions about the net effects on student achievement or social and racial segregation. What can be learned—e.g., from large-scale programs in Chile or New Zealand—seems to be that large-scale universal school voucher programs do not generate substantial gains and could even be detrimental to sub-populations. More narrowly targeted programs seem to be more promising, but should be carefully implemented and only serve as one element of a broader strategy.

While school vouchers are quite extensively studied in the literature, there exist only few studies of vouchers for job training programs. Barnow (2008) gives an overview of studies on vouchers in U.S. vocational training programs. The empirical evidence for the effectiveness of training vouchers for dislocated workers is mixed. To the best of our knowledge there exist no econometric study evaluating the effectiveness of the introduction of vouchers for training programs targeted at unemployed individuals.[2]

In this paper we focus on a voucher for training programs in the context of ALMP, and estimate the impact of the introduction of training vouchers on the effectiveness of the most important program type in Germany: occupation-related or general training. Participants in this program either learn specific skills required for a certain vocation or receive qualifications which are of general vocational use. In the pre-reform period about 60 percent of all unemployed participants in public training programs were assigned to this particular type, and it became even more

---

[1]The discussion about vouchers in the educational context started with Friedman (1962).

[2]A recent example for vouchers in the context of ALMP—although not in the field of education— is the job placement voucher. It was introduced in Germany in 2002 in order to end the public placement monopoly and subsidize private competitors. Winterhager et al. (2006) evaluate the effectiveness of this instrument and find a positive impact on the employment probability of voucher recipients in West Germany.

important after the reform in 2003 as this share increased to more than 70 percent.

In addition to the training voucher, a stricter selection of participants and program types by the caseworkers based on the expected reemployment probability was introduced. This implies that the caseworkers tend to select individuals with higher reemployment probabilities for participation—independent of the individual gain resulting from participation. Thus, the overall effect of the reform could result from the introduction of the voucher system (voucher effect) and from a change in the composition of participants (selection effect). To decompose the overall reform effect into these two effects, we apply a two-step propensity score matching procedure using a rich administrative data set. This approach allows a comparison between pre- and post-reform participants who have similar observable characteristics. Furthermore, we apply regression analyses to the matched data to adjust for possible remaining unbalanced covariates, and to take into account changes in the general economic conditions.

We find a slightly positive impact of the reform. The decomposition of this overall effect shows that the selection effect is—if at all—slightly negative. This finding is consistent with Lechner and Smith (2007). Although the selection effect in our paper differs from the effect of caseworkers in their paper, the major part of our selection effect is also due to the caseworker. Our findings imply that using post-training outcomes as a performance standard is not a good strategy to improve the effectiveness and efficiency of public training programs, see Heckman et al. (1997) for an insightful discussion on performance standards in the context of the Job Training Partnership Act. Furthermore, we find evidence that the voucher effect increased both, the employment probability and earnings of the participants. This effect becomes substantially positive after around 6 months of training, and decreases slightly at the end of our observation period (1.5 years after program entry).

The remainder of this paper is organized as follows. Section 2 describes the institutional background of public training programs in Germany with a particular focus on changes between the pre- and post-reform period. After describing the analytical framework in Section 3 and the data in Section 4, we discuss the matching quality in Section 5 and present the results in Section 6. Finally, Section 7 concludes.

2 Institutional Background

ALMP aims to increase the employment prospects of unemployed individuals. For this purpose, the Federal Employment Agency (FEA) in Germany spends a substantial amount of money on programs such as job creation schemes, public training programs, or employment subsidies. For instance, about 20.5 billion Euros were spent in 2002 (Eichhorst and Zimmermann, 2007). The most important part of ALMP in Germany are public training programs. With almost 7 billion Euros, these programs account for more than 32 percent of the expenditures. However, the number of participants decreased over the last years (see Figure 1). While more than 500,000 unemployed individuals entered a training program in 2000, this number approached only around 130,000 individuals in 2005. In 2006, it increased again to nearly 250,000 persons entering such programs.

Figure 1 about here

The effectiveness of public training programs in Germany before the Hartz reforms has been subject to a number of studies. For a recent review of the results see Caliendo and Steiner (2005).[3] The results are quite heterogeneous—depending on the investigation period and the underlying data set. Recent studies which are based on rich administrative data sets often find at least for some sub-groups positive treatment effects (Lechner et al., 2005, 2007; Fitzenberger et al., 2008; Biewen et al., 2007; Rinne et al., 2007). However, there are also recent studies finding insignificant or negative effects (Hujer et al., 2006; Lechner and Wunsch, 2008). Besides differences in the investigation period and the underlying data set, the mixed results may also be due to different methodological approaches. For instance, Stephan (2009) finds that estimated treatment effects differ considerably across different definitions of non-participants. Overall, the major lesson from the evaluation studies analyzing the pre-reform period—i.e., before 2003—seems to be that positive effects mainly occur in the longer run, and that studies which find positive medium- or long-term effects are also reporting negative short-term effects.

Germany's ALMP has undergone a series of reforms during the past years. Figure 2 summarizes the most important legislative changes in this context. Although these reforms are commonly referred to as the Hartz reforms, the first effort was made when the *JobAQTIVE Law* came into force on January 1, 2002. Next

---

[3]The international literature on the evaluation of ALMP is summarized by LaLonde (2003) and Kluve (2006), among others.

to changes which affected job placement procedures, registration standards and job offer requirements, this law also altered the rules for benefit claims during and after participation in public training programs.

Figure 2 about here

Major changes which affected the provision of public training programs came into force on January 1, 2003 under the first two reform packages (Hartz I/II). Additional changes affected temporary employment as Personal Service Agencies (PSA) were introduced and moreover unemployment assistance levels were cut. Registration standards and job offer requirements were altered (again), and the rules for exclusion from benefits were tightened. Finally, the legislation concerning marginal employment was extended and a new start-up subsidy for the unemployed was introduced.

The third package of Hartz reforms (Hartz III) came into force on January 1, 2004. Its main objective was a reorganization of the Federal Labor Agency. Additional changes affected job creation schemes, the unemployment benefit legislation, and the rules for exclusion from benefits. As a final step of the series of reforms, the fourth package (Hartz IV) was introduced in the beginning of 2005. Its most important feature constitutes the combination of the former systems of unemployment assistance and social assistance into the new means-tested unemployment benefit II system. Besides that, One-Euro-Jobs were introduced. An overview of the impact of the reform on ALMP is given for example by Jacobi and Kluve (2007). A broader picture is provided by Ebbinghaus and Eichhorst (2009) who describe institutional provisions and reforms of employment protection, active and passive labor market policies in Germany between 1991 and 2005.

Prior to the Hartz I/II package, i.e., before 2003, the provision of public training programs in Germany was organized as follows. After consultation with the job seeker, the caseworker in the local office of the FEA decided whether or not the unemployed individual should receive training. Courses were operated by private providers which were approved beforehand. The system is considered as a *de facto* contracting-out, although there were no legal contracts between providers and local FEA offices. Legally, job seekers paid the courses and were reimbursed, but usually the local offices paid the course fees directly to the providers in order to facilitate administration. The degree of competition among providers was limited since approvals were granted only to exactly the number of providers needed to meet regional demand. A public tendering procedure was not in place. This informal

procedure entailed a potential for collusive behavior between local FEA offices and private providers. For instance, there was an informal guarantee that the capacity approved by the local office would be fully used. It was often reported that approved courses were simply filled up, even though the training provided was inappropriate for some individuals.

After January 1, 2003 the provision of public training programs substantially changed. The most prominent feature of the reform marks the introduction of the training voucher (*Bildungsgutschein*) which abandoned the former *de facto* contracting-out system. A training voucher is granted if the caseworker considers participation in a given type of public training program as a successful strategy to reintegrate the job seeker in the primary labor market—without taking into account the *relative* gain compared to the counterfactual situation without participation. The selection criteria for participants thus became stricter after the reform; and the matching between program types and participants by the caseworkers which is also based on the expected reemployment probability is completely novel. As Figure 3 shows, the voucher—once it is granted—prescribes the program's maximum duration, its intended educational target, its geographical scope, and the maximum course fee which will be reimbursed by the local FEA office. It is valid for at most three months. Within this period, job seekers are completely free to chose among approved training providers and courses in the market—subject to the requirements stated in the voucher.[4] Local FEA staff are not allowed to make recommendations, but can provide, e.g., a list of approved courses. There was, however, a transitional arrangement when the reform was introduced: The allocation of participants into public training programs was *exclusively* based on vouchers only from March 2003 onwards (Schneider et al., 2007).[5]

Figure 3 about here

Although the innovative voucher system should both increase consumer sovereignty and competition among training providers, Bruttel (2005) presents initial evidence that there are practical obstacles to fully achieve this positive effects. For instance, information asymmetries constrain consumer sovereignty. In particular low-skilled job seekers lack the abilities to navigate the training market and to take

---

[4]The approval of providers and courses is subject to a new quality management system which adopts a two-level approach. For details see, e.g., Bruttel (2005).

[5]The official transitional arrangement was as follows: "Individuals who were counseled before January 1, 2003 and participation in a public training program was agreed upon do not receive a training voucher if they enter the program until February 28, 2003."

an active role in searching for an appropriate course. This argument is supported by Kruppe (2008) who finds that low-skilled individuals are significantly less likely to redeem a granted voucher than persons with higher qualifications. However, the overall redemption rate is comparatively high with 85 percent in the period from 2003 to 2006 (Kruppe, 2008). On the supply side, a potential obstacle for competition among providers is their unequal distribution across German regions. Providers also reacted to the reform and increased co-operation and collusive behavior, for example by not offering the same courses anymore.

This initial evidence is supplemented by Schneider et al. (2007) who analyze the implementation of the reform as a whole. Accordingly, the impacts of the reform primarily materialize in two dimensions. First, the composition of participants is affected. Participants in the post-reform period exhibit on average better employment prospects than in the pre-reform period. Second, the structure of program types is affected. The focus shifts towards regions with lower unemployment rates, courses with comparatively shorter durations, and courses providing qualifications and skills which fit regional short-term market demand.

Given the practical obstacles and the actual implementation process, the overall impact of the reform on the effectiveness of public training programs is anything but clear. However, Schneider and Uhlendorff (2006) and Schneider et al. (2007) find that the effectiveness increases after the reform. Nonetheless, the question which features of the reform cause this increase—and to what particular extent—remains unanswered.

Compared to previous studies, the most significant difference of our paper is the decomposition of the overall reform effect into a "selection effect" and a "voucher effect". We refer to the selection effect as the effect resulting from a different composition of participants between the pre- and the post-reform period. This effect is due both to stricter selection criteria and to the *unintended* consequence of the voucher that low-skilled job seekers lack the abilities to navigate the training market and to take an active role in searching for an appropriate course. Low-skilled individuals thus exhibit a lower voucher receipt rate (intended by the caseworkers) as well as a lower voucher redemption rate (unintended). On the other hand, the voucher effect comprises the *intended* impacts of the introduction of training vouchers according to our taxonomy. These consequences include a potentially better match between participants and courses, an apparently more market-oriented (i.e., demand-oriented) approach of the local FEA offices, and quality enhancements which could be due to increased competition among training providers.

In order to isolate the above two effects, and to avoid complications of other components of Hartz III and IV reform discussed previously, we restrict our sample to 2002 (pre-reform period) and 2003 (post-reform period).

3 Analytical Framework

In order to disentangle the effects of the reform arising from a change of program quality due to a better match between participants and providers and/or an improved quality of the offered training programs (voucher effect) and the change of the composition of participants (selection effect), we apply a two-step matching approach.

Using the potential outcome framework as in Neyman (1923), Roy (1951), or Rubin (1974), we assume that each individual has two potential outcomes for the program: Y_{1i} is the outcome if individual i participates, and Y_{0i} if not. Let D_i be an indicator for participation, we can define different treatment effects in a similar way as Heckman and Vytlacil (1999, 2005):

$$TE_i = Y_{1i} - Y_{0i} \qquad \text{(Treatment effect for individual } i\text{)}$$
$$ATE = E[TE_i] \qquad \text{(Average treatment effect for the population)}$$
$$ATT = E[TE_i | D_i = 1] \qquad \text{(Average treatment effect on the treated)}$$

In this paper, we are interested in the treatment effect on the treated and its change induced by the reform. R_i is an indicator which takes on the value 0 if we observe an individual before the reform and 1 if we observe an individual after the reform. The average treatment effects on the treated before and after the reform are given by:

$$ATT_b = E[TE_i | D_i = 1, R_i = 0] \qquad \text{(ATT pre-reform period)}$$
$$ATT_p = E[TE_i | D_i = 1, R_i = 1] \qquad \text{(ATT post-reform period)}$$

A simple comparison of treated and non-treated individuals may be biased if participants and non-participants differ with respect to characteristics having an impact on the outcome Y. If treatment assignment is *strongly ignorable*, i.e. if selection is based on observable characteristics X (conditional independence) and if observable characteristics of participants and non-participants overlap, the matching approach is an appealing choice to estimate treatment effects. Formally, these

assumptions are given by:

$$(Y_{0i}, Y_{1i}) \perp D_i | X_i \qquad \text{(Conditional independence assumption)}$$
$$0 < prob(D_i = 1 | X_i) < 1 \qquad \text{(Overlap assumption)}$$

Rosenbaum and Rubin (1983) show that if the matching assumptions hold, i.e., treatment assignment is strongly ignorable *given* X, it is also strongly ignorable *given any balancing score that is a function of* X.[6] One possible balancing score is the propensity score $P(X)$, i.e., the probability of participating in a given program. Mueser et al. (2007) present evidence that if rich administrative data is used to measure the performance of training programs, propensity score matching is generally most effective.

We thus estimate ATT_b (ATT_p) from pre-reform data (post-reform data) by propensity score matching methods.[7] However, the difference between ATT_b and ATT_p does *not* equal the effect of the introduction of vouchers, since the participants before and after reform may have different characteristics. As mentioned above, compared to the pre-reform period, the post-reform programs are more selective (possibly leading to a selection effect, SE) and vouchers are introduced (which may cause a voucher effect, VE). If we assume additive separability of the two components, ATT_p is given by:

$$ATT_p = ATT_b + VE + SE \tag{4}$$

and the overall reform effect (RE) can be written as:

$$
\begin{aligned}
RE &= ATT_p - ATT_b \\
&= VE + SE
\end{aligned}
\tag{5}
$$

To isolate the voucher effect, we apply a two-step propensity score matching procedure. In the first step, pre-reform participants are matched with post-reform participants. As a result, the obtained pairs of participants only differ with respect to the timing of participation. Importantly, observable characteristics do not differ anymore. In the second step, the matched pre-reform participants in 2002 are matched with non-participants of the same year. The corresponding treatment effect

---

[6]When there are many covariates, it is impractical to match directly on covariates because of the curse of dimensionality. See, e.g., Zhao (2008) for some comments on this problem.

[7]Heckman et al. (1997, 1998) present a weaker version of the conditional independence assumption: $E[Y_{0i}|D_i = 0, X_i] = E[Y_{0i}|D_i = 1, X_i]$.

ATT_{bp} is the effect *only* for those participants under the pre-reform regime who are comparable to participants after the reform. This step controls for the changes in the composition of participants before and after the reform, i.e., the selection effect.

With ATT_{bp} we can calculate the difference in differences of the treatment effects to estimate the voucher effect:

$$VE = ATT_p - ATT_{bp} \tag{6}$$

Finally, the comparison of the voucher effect with the reform effect gives us an estimate of the selection effect:

$$\begin{aligned} SE &= RE - VE \\ &= (ATT_p - ATT_b) - (ATT_p - ATT_{bp}) \\ &= ATT_{bp} - ATT_b \end{aligned} \tag{7}$$

There are several propensity score matching methods suggested in the literature, see, e.g., Imbens (2004), Caliendo and Kopeinig (2008) and Imbens and Wooldridge (2008) for overviews. Based on the characteristics of our data and particularly because of the two-step matching approach which is pursued, we opt for nearest-neighbor matching without replacement. This matching method has the advantage of being the most straightforward matching estimator: a given participant is matched with a non-participant or participant who is closest in terms of the estimated propensity score. We avoid an increased variance of the estimator as we match without replacement (Smith and Todd, 2005a), which is justified since the ratio between participants and (non-)participants—i.e., potential matching partners—is comparatively high in our data. Hence, the constructed counterfactual outcome is based only on distinct (non-)participants.

More specifically, the probability of participation is estimated conditional on a number of observable characteristics using binary probit models with participation as the dependent variable.[8] These characteristics include socio-demographic-characteristics (e.g., age, nationality, marital status), regional information (region, unemployment rate), educational and vocational attainment, the employment history (four years prior to program entry), and information on the last employment

---

[8]The matching algorithms are implemented using the `PSMATCH2` Stata ado-package by Leuven and Sianesi (2003).

spell (duration, income, business sector).[9] We run these regressions separately for women and men from East and West Germany, respectively. After estimating the propensity score, we find a suitable matching partner by exact covariate matching combined with propensity score matching. The variables used for exact matching are region, previous unemployment duration (in months), and quarter of program entry. Therefore, we stratify the four sub-samples of women and men in East and West Germany by these variables first, and then implement propensity score matching for each cell without replacement. This procedure ensures that matched participants and non-participants are *a*) previously unemployed for the same duration at program entry, and *b*) entering the program in the same quarter. The latter condition makes sure that seasonal influences are constant. Furthermore, we do not condition on future non-partition. This is important in the context of dynamic assignment processes. Following the argumentation of Sianesi (2004), in countries like Sweden or Germany in principle any unemployed individual will join a program at some time, provided he remains unemployed long enough. Hence, a restriction on future outcomes—i.e., to require non-participation in the follow-up period after the fictitious program entry—is supposed to effect estimated treatment effects negatively, since a substantial fraction of the 'never treated'-individuals would *de facto* be observed to leave the unemployment register.

In order to assess the impact of the reform on the employment probability and earnings of participants, we estimate linear probability and ordinary least squares models. To test the robustness of our results with respect to potential differences in observable characteristics X, which may remain after the matching, we run additional regressions controlling for observable characteristics X.[10]

These additional regressions also allow us to control for changes in the general economic situation and changes in the extent and the composition of ALMP; these changes may be additional components of the reform effect.[11] Although we generally argue that we control for such changes as participants and matched non-participants are subject to the same environment, we will explicitly address this issue in our sensitivity analyses and control for potential changes in our regressions.

For the variance of the estimated treatment effects, we base our inference on

---

[9]Selected variables are listed in Table 1. The exact specifications and a list of all variables are not reported here, but available upon request.

[10]This procedure can correct a possible remaining bias as discussed in Rubin (1973), Imbens (2004), and Abadie and Imbens (2002).

[11]For instance, Lechner and Wunsch (2006) present evidence for a clear positive relation between the effectiveness of the programs and the unemployment rate over time.

bootstrapping procedures. More specifically, we bootstrap the whole estimation process. This allows us to calculate standard errors based on the distribution of the estimated treatment effects. The standard errors of the reform effect, the voucher effect and the selection effect are based on the distribution of the respective differences in treatment effects across the replications.

4 Data

We use a sample of a particularly rich administrative data set, the Integrated Employment Biographies (IEB) of the FEA.[12] It contains detailed daily information on employment subject to social security contribution including occupational and sectoral information, receipt of transfer payments during periods of unemployment, and participation in different programs of ALMP. Furthermore, the IEB comprises a large variety of covariates—e.g., age, marital status, number of dependent children, disability, nationality and education.

In Germany, public training programs for the unemployed are quite heterogenous. Thus, we concentrate on the most important program type: occupation-related or general training. Participants either learn specific skills required for a certain vocation (e.g., computer-aided design for a technician/tracer) or receive qualifications that are of general vocational use (e.g., MS Office, computer skills). The program does *not* aim to provide a certificate, i.e., an officially recognized vocational degree. In contrast to other program types it focuses on classroom training and is neither provided in combination with internships nor is the simulation of real operations conducted. In the pre-reform period about 60 percent of all participants in public training programs were assigned to this particular type, and it became even more important after the reform in 2003 as this share increased to more than 70 percent. Figure 4 indicates that the program is—in comparison to other ALMP measures in Germany—a rather short measure. Both in the pre-reform and in the post-reform period, after one year more than 90 percent of the participants have exited from the program. However, the program duration decreased after the reform. While the median program duration is about 8 months in the pre-reform period, it amounts to about 6 months after the reform.

---

[12]The IEB combines four different administrative data sources: the employees' history (BeH), the benefit recipients' history (LeH), the job seekers' data base (ASU/BewA), and the program participants' master data set (MTH). However, it is in general not publicly available. Only a 2.2 percent random sample (the Integrated Employment Biographies Sample, IEBS) can be obtained for research purposes. See, e.g., Jacobebbinghaus and Seth (2007) for details on the IEBS.

To evaluate the impact of the reform and its features on the effectiveness of this type of public training program, our data includes participants as well as non-participants from the pre- and post-reform period, respectively. More specifically, we have information on: *a*) participants who entered the program in 2002, *b*) participants who entered the program in 2003, *c*) non-participants in 2002, and *d*) non-participants in 2003. We do not have information on individuals who received a voucher but did not make use of it. Our sample of participants who entered the program in 2003 consists of more than 1,200 individuals. In order to apply the matching approach as described above (see Section 3) roughly 20 participants from the period before the reform were drawn per participant in 2003. Therefore, we have information on about 23,000 participants who entered the program in 2002. Beyond the matching of post-reform participants with pre-reform participants, we need to match participants with non-participants. For both years (2002 and 2003) our sample of non-participants—i.e., potential controls—consists of more than 500,000 individuals. Non-participants are required to not having participated in the given type of training before and in the quarter of the participant's program entry, but we do not condition on future non-participation.

Table 1 displays descriptive statistics of selected variables for the samples of participants and non-participants in 2002 and 2003, respectively. Focusing firstly on individuals participating in training, we find evidence for a change in their composition between the pre- and the post-reform period in our data. The most remarkable change can be observed with respect to the previous employment histories which differ considerably between these two groups. Considering a period of four years prior to program entry, participants who entered after the reform show a higher labor market attachment in terms of un-/employment rates and a slightly higher income from last employment than earlier program entrants. The average age of a participant also dropped by more than one year, while other characteristics remain on average rather stable between the two years. In particular, differences with respect to the educational or vocational attainment do not appear to be substantial. On the other hand, the groups of non-participants are very different from the groups of participants in both years. They are on average older and less educated. Moreover, their employment histories reveal a higher incidence of unemployment as well as a lower incidence of employment relative to participants in training.

The success of program participation is evaluated by looking at a) the probability of being employed, and b) earnings. Our observation period—i.e., the period in which outcomes are observed—starts at program entry and it ranges over a period of 18 months. This period is based on the fact that we focus on program participation in the years 2002 and 2003, and we can observe reliable data for all employment states until December 31, 2004. Individuals are regarded as employed if they hold a job in the primary labor market. For instance, participation in job creation schemes is not included in this outcome measure. Moreover, the administrative data set only includes employment that is subject to social security contributions.[13] Additionally, we evaluate the effect of program participation on monthly earnings in the primary labor market. We apply the described definition of employment and consider remunerations associated with these spells in terms of monthly earnings.

In order to control for changes in the general economic situation which may constitute another component of the reform effect, we consider a number of economic and labor market characteristics available for each labor market district. We use monthly information on the share of unemployed, the share of vacancies, the share of participants in various ALMP measures (including public training programs) as well as GDP growth rates.[14] Table A1 in the Appendix reports the change of these variables between 2002 and 2004. For example, the unemployment rate slightly increased on average from around 10 percent in 2002 to around 10.7 percent in 2004, while the share of unemployed individuals participating in training programs decreased during this period.

Furthermore, the implementation of the reform may have varied across local FEA districts. We address this issue by using information about the subjective judgement of the Hartz reforms obtained through a survey conducted in the beginning of 2005 in the management departments of the local FEA districts. They are asked about the change of the job placement, the benefit granting, the administrative effort and the co-operation with third parties like training providers and employers, and the subjective judgment is on average rather positive. However, we observe heterogeneity in the judgements, and we will control for this in our regressions. The included items are reported in Table A2 in the Appendix.

---

[13]This means that, e.g., we do not observe self-employment.

[14]We include annual GDP growth rates for the 16 federal states since more disaggregated data is not available.

5 Matching Quality

We apply different strategies to evaluate the balancing of observable characteristics between the different groups after the matching.

One way to assess the matching quality is to compare the standardized bias before matching, SB^b, to the standardized bias after matching, SB^a. The standardized biases are defined as

$$SB^b = \frac{(\overline{X}_1 - \overline{X}_0)}{\sqrt{0.5 \cdot (V_1(X) + V_0(X))}} \quad ; \quad SB^a = \frac{(\overline{X}_{1M} - \overline{X}_{0M})}{\sqrt{0.5 \cdot (V_{1M}(X) + V_{0M}(X))}} \quad , \quad (8)$$

where X_1 (V_1) is the mean (variance) in the treated group before matching and X_0 (V_0) the analogue for the comparison group. X_{1M} (V_{1M}) and X_{0M} (V_{0M}) are the corresponding values after matching (Rosenbaum and Rubin, 1985). The mean standardized bias should be reduced after matching.

Following the suggestion of Sianesi (2004) we also re-estimate the propensity score on the matched sample to compute the pseudo-R^2 before and after matching. The pseudo-R^2 indicates how well the observable characteristics X explain the probability of being treated. After the matching the pseudo-R^2 should be low because there should be no systematic differences between the treated and not treated individuals.

In a third approach we test the balancing following a suggestion by Smith and Todd (2005b) and estimate the following regression for each observable characteristic x included in our preferred specification:

$$\begin{aligned} x_k = \beta_0 &+ \beta_1 \widehat{PS(X)} + \beta_2 \widehat{PS(X)}^2 + \beta_3 \widehat{PS(X)}^3 \\ &+ \alpha_0 D + \alpha_1 D\widehat{PS(X)} + \alpha_2 D\widehat{PS(X)}^2 + \alpha_3 D\widehat{PS(X)}^3 + \varepsilon_k \end{aligned} \quad (9)$$

D is the treatment indicator, $\widehat{PS(X)}$ the estimated propensity score, and x_k is the observable characteristic k. For each x we perform an F-test of the joint null hypothesis that all coefficients on terms involving D equal zero. If the balancing score satisfies the balancing condition, D should not provide any information about x_k.

Table 2 summarizes the results from the three balancing tests. We tested the balancing for different sub-samples: women and men in East and West Germany, respectively. Altogether, we perform five matching procedures: *a)* pre-reform participants are matched with post-reform participants, *b)* *unmatched* pre-reform participants are matched with pre-reform non-participants, *c)* *unmatched* post-reform

15

participants are matched with post-reform non-participants, *d) matched* pre-reform participants are matched with pre-reform non-participants, and *e) matched* post-reform participants are matched with post-reform non-participants. *Unmatched* and *matched* participants may differ because we do not find for every participant after the reform a suitable match from the period before the reform, i.e., the *matched* participants are a subset of the *unmatched* participants. Overall, the balancing of the different matching procedures is quite satisfactory: the percentage biases are apparently reduced. More specifically, mean standardized biases in the matched samples are—with one exception—noticeably smaller than in the unmatched samples and are are mostly below five percent after matching. Likewise, the pseudo-R^2 after matching are fairly low and decrease substantially compared to before matching. Moreover, in most of the matching procedures our third test indicates that D does not provide any information about any observable characteristics. However, some of our matching procedures perform better than others. We get the worst performance for our matching of participants before the reform with participants after the reform, especially for females in East Germany—although the third test indicates no problems for the participant-participant matching for any of our sub-samples. Therefore, we will check the sensitivity of our results to the inclusion of observable characteristics in our regressions based on the matched samples, and we have to be careful in the interpretation of our results for females in East Germany.

Table 2 about here

As an additional check of the matching quality, we plot the fraction of individuals being employed in the primary labor market before and after matching for a period of four years prior to participation. This approach follows, e.g., Heckman and Hotz (1989) and Mueser et al. (2007). Figures 5 and 6 display that the one-step as well as the two-step matching procedures generate comparison groups with employment probabilities prior to participation which are pretty close to those of the treatment groups. Although they are still not identical, the substantial differences before matching disappear across all matched samples. Moreover, Figure 6(a) points out very clearly that—before matching—participants after the reform have more favorable employment histories than participants before the reform.

Figures 5 and 6 about here

6 Results

In this section, we present the effects on different outcomes for a period of 1.5 years after program entry. First, we compare the employment probabilities. Second, we present results for the effects on earnings, and finally we conduct several sensitivity analyses.

6.1 Employment Probabilities

The estimates of the average treatment effects on employment probabilities of the participants are reported in Figure 7. We observe that participants before and after the reform face a substantial lock-in effect, both treatment effects are significantly negative in the first months.[15] After around 6 months of training, both treatment effects diverge and the treatment effect for participants after the reform constantly lies above the treatment for participants before the reform. At the end of our observation period, i.e., 1.5 years after program entry, the point estimates of the treatment effects amount to about 3 percentage points before the reform and about 7 percentage points after the reform.

> Figure 7 about here

The differences between these two treatment effects describes the reform effect. We thus find a positive impact of the reform, which may occur due to the voucher effect or due to the change in the composition of participants. Figure 8 displays the decomposition of the reform effect and reveals insights about the extent and magnitude of reform effect, voucher effect, and selection effect. The upper part reports the point estimates of the three effects, while the effects including the corresponding confidence intervals are reported separately in the lower part.

> Figure 8 about here

The decomposition shows that the positive reform effect seems to be almost exclusively based on the voucher effect. Similar to the reform effect, the voucher effect becomes substantially positive after around 6 months. The voucher effect is significantly positive from month 7 until month 13 after entering the program.

---

[15]While participating—or being 'locked-in' in the program—individuals probably reduce their search activities for new jobs (van Ours, 2004).

However, for the last 5 months of our observation period, the effect is still positive, but not significantly different from zero.

The point estimates of the selection effect are almost always negative, but never significantly different from zero. This indicates that there is no evidence for a positive impact of a stricter selection of participants on the average treatment effect. In contrast to that, our results—if at all—seem to indicate that the overall reform effect would have been more positive if the composition of participants had not changed. Our finding is consistent with Lechner and Smith (2007) who present evidence that caseworkers are not the best choice to allocate unemployed individuals into programs. Although their results are based on Swiss data, the situation in which caseworkers select the training providers and programs on behalf of the unemployed precisely describes the pre-reform situation in Germany. This changed under the new regime; and after the reform job seekers are free to choose their provider on their own by means of training vouchers.

Additionally, we estimate the corresponding effects for four subgroups separately: men and women in West and East Germany, respectively. The results are reported in the Appendix, Figures A1–A16. We find some evidence for heterogeneity of the effects, but the general picture is similar. We find rather negative and never significant selection effects and rather positive voucher effects. The voucher effect is significantly different from zero only for men in East Germany. However, because the number of observations clearly drops if we analyze the effects separately for the four subgroups, these results have to be interpreted with caution.

6.2   Earnings

Corresponding to the effects on employment probabilities, we present the average treatments effects on monthly earnings before and after the reform in Figure 9. Again, we observe substantial lock-in effects for both periods and clearly higher point estimates for the post-reform period after around 6 months of treatment. 18 months after entering the program, the point estimates of the treatment effects are about € 40 and roughly € 150 per month in the pre- and post-reform period, respectively.

Figure 9 about here

Figure 10 displays the decomposition of the reform effect. Similar to the employment probabilities, the positive reform effect seems to be almost exclusively

18

based on the voucher effect. We find no significant impact of the selection effect and a positive impact of the voucher effect, although not always significant. The similarity to the employment probabilities is not surprising, because the positive earnings effects reflect at least partly the increased employment probabilities.

<div style="text-align:center">| Figure 10 about here |</div>

In order to get additional insights in the effects on earnings, we estimate the treatment effects using realized earnings only, i.e., we compare earnings conditional on being employed. The results are reported in Figures 11 and 12.

<div style="text-align:center">| Figures 11 and 12 about here |</div>

The point estimates of the voucher effect are always positive, and the selection effect is always negative. Both effects are not significantly different from zero, but the voucher effect steadily increases during the observation period. Altogether, this indicates that the introduction of the voucher—next to an increased employment probability—also leads to better job matches for the participants, measured by on average higher monthly earnings in the new job.

6.3 Sensitivity Analysis

We address the robustness of our previous results in this section. For this purpose, we perform a sensitivity analysis in two steps. We assess the robustness of our results with respect to *a)* the inclusion of additional control variables, and *b)* the transitional arrangement for the training voucher in the beginning of 2003.

Inclusion of additional control variables

One may argue that changes in the general economic situation constitute another component of the reform effect. Therefore, we additionally control for a number of economic and labor market characteristics which are available for each local FEA district. These variables are changing over time. In addition to that, we include observable individual characteristics measured before entering the treatment and include—only for the post-reform period—indicators describing the implementation of the Hartz reform on the FEA district level.

The results are presented in Figures 13 and 14. In general, the picture is very similar to the results presented above. The point estimates of the voucher effect

are slightly lowered, while the selection effect is slightly less negative. However, the voucher effect is still significantly positive between month 7 and month 13 after entering the program, and the selection effect is still almost always negative.

<div style="text-align: center; border: 1px solid black; display: inline-block;">Figures 13 and 14 about here</div>

The results are also very similar for earnings, as reported in Figures 15 and 16. The voucher effect marginally lowers and selection affect slightly increases. Our results thus appear to be robust to the inclusion of additional control variables.

<div style="text-align: center; border: 1px solid black; display: inline-block;">Figures 15 and 16 about here</div>

Transitional Arrangement

We mentioned above that there has been a transitional arrangement in place until March 2003 (see Section 2). Unfortunately, the administrative data set does not allow us to identify those participants who actually received and redeemed a training voucher. We thus perform a sensitivity analysis and exclude participants who entered public training programs in the first quarter of 2003.[16] The results of this analysis are depicted in Figures 17 and 18.

<div style="text-align: center; border: 1px solid black; display: inline-block;">Figures 17 and 18 about here</div>

Also after taking account of the transitional arrangement in the beginning of 2003, we still observe the main result of a positive impact of the voucher. The selection effect is virtually zero. However, at the end of the observation period we estimate both voucher and selection effect to be positive, although not significantly different from zero. The results are again very similar for earnings, as reported in Figures 19 and 20.

<div style="text-align: center; border: 1px solid black; display: inline-block;">Figures 19 and 20 about here</div>

---

[16]According to Schneider et al. (2007) who analyze survey data, the fraction of participants in public training programs actually receiving a voucher was about 30 percent in the first quarter of 2003, but sharply increased subsequently. Of course excluding participants who entered public training programs in the first quarter of 2003 implies that we also exclude participants who entered public training programs in the first quarter of 2002 as well as corresponding non-participants based on our matching algorithm.

7 Conclusion

This paper analyzes the impact of the labor market reform in 2003 on the effectiveness of the most important type of public training program in Germany. This reform had two main features: a) it introduced training vouchers, and b) it imposed more selective criteria on participants. Next to estimating the overall impact, we decompose the reform effect into a) the voucher effect, and b) the selection effect.

Our results confirm previous findings that the reform slightly increased the effectiveness of public training programs in Germany. The decomposition shows that the positive reform effect seems to be almost exclusively based on the voucher effect. The selection effect is—if at all—slightly negative, although never significantly different from zero. This indicates that there is no evidence for a positive impact of a stricter selection of participants on the average treatment effect. In contrast to that, our results seem to indicate that the overall reform effect would have been even more positive if the composition of participants had not changed.

Our results moreover indicate that the introduction of the voucher—next to an increased employment probability—also leads to better job matches for the participants, measured by on average higher monthly earnings in the new job.

References

Abadie, A. and G. W. Imbens (2002). Simple and Bias-Corrected Matching Estimators for Average Treatment Effects. NBER Technical Working Paper 286, National Bureau of Economic Research, Inc.

Angrist, J., E. Bettinger, E. Bloom, E. King, and M. Kremer (2002). Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment. *American Economic Review 92*(5), 1535–1558.

Barnow, B. S. (2000). Vouchers for Federal Targeted Training Programs. In C. E. Steuerle, V. Doorn Ooms, G. E. Peterson, and R. D. Reischauer (Eds.), *Vouchers and the Provision of Public Services*. Washington D.C.: Brookings Institution Press.

Barnow, B. S. (2008). Vouchers in U.S. Vocational Training Programs: An Overview of What We Have Learned. *mimeo*.

Biewen, M., B. Fitzenberger, A. Osikominu, and M. Waller (2007). Which Program for Whom? Evidence on the Comparative Effectiveness of Public Sponsored Training Programs in Germany. IZA Discussion Paper 2885, Institute for the Study of Labor (IZA), Bonn.

Bruttel, O. (2005). Delivering Active Labor Market Policy through Vouchers: Experiences with Training Vouchers in Germany. *International Review of Administrative Sciences 71*(3), 391–404.

Caliendo, M. and S. Kopeinig (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys 22*(1), 31–72.

Caliendo, M. and V. Steiner (2005). Aktive Arbeitsmarktpolitik in Deutschland: Bestandsaufnahme und Bewertung der mikroökonomischen Evaluationsergebnisse. *Journal for Labour Market Research 38*(2-3), 396–418.

Ebbinghaus, B. and W. Eichhorst (2009). Employment Regulation and Labor Market Policy in Germany, 1991–2005. In P. de Beer and T. Schils (Eds.), *Social Policy and the Labour Market: Achieving an Optimal Policy Mix*. Cheltenham: Edward Elgar.

Eichhorst, W. and K. F. Zimmermann (2007). And Then There Were Four... How Many (and Which) Measures of Active Labor Market Policy Do We Still Need? *Applied Economics Quarterly 53*(3), 243–272.

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 Statistique* (forthcoming).

Friedman, M. (1962). The Role of Government in Education. In *Capitalism and Freedom*, Chapter 6. Chicago: University of Chicago Press.

Heckman, J. and E. Vytlacil (1999). Local Instrumental Variables and Latent Variables Models for Identifying and Bounding Treatment Effects. *Proceedings of the National Academy of Sciences of the USA 96*, 4730–4734.

Heckman, J. and E. Vytlacil (2005). Structural Equations, Treatment Effects and Econometric Policy Evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J., C. Heinrich, and J. A. Smith (1997). Assessing the Performance of Performance Standards in Public Bureaucracies. *American Economic Review 87*(2), 389–395.

Heckman, J. J. and V. J. Hotz (1989). Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training. *Journal of the American Statistical Association 84* (408), 862–874.

Heckman, J. J., H. Ichimura, and P. E. Todd (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies 64* (4), 605–654.

Heckman, J. J., H. Ichimura, and P. E. Todd (1998). Matching as an Econometric Evaluation Estimator. *Review of Economic Studies 65* (2), 261–294.

Hujer, R., S. L. Thomsen, and C. Zeiss (2006). The Effects of Vocational Training Programmes on the Duration of Unemployment in Eastern Germany. *Allgemeines Statistisches Archiv 90* (2), 299–322.

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

Imbens, G. W. and J. M. Wooldridge (2008). Recent Developments in the Econometrics of Program Evaluation. IZA Discussion Paper 3640, Institute for the Study of Labor (IZA), Bonn.

Jacobebbinghaus, P. and S. Seth (2007). The German Integrated Employment Biographies Sample IEBS. *Schmollers Jahrbuch 127* (2), 335–342.

Jacobi, L. and J. Kluve (2007). Before and After the Hartz Reforms: The Performance of Active Labour Market Policy in Germany. *Journal for Labour Market Research 40* (1), 45–64.

Kluve, J. (2006). The Effectiveness of European Active Labor Market Policy. IZA Discussion Paper 2018, Institute for the Study of Labor (IZA), Bonn.

Krueger, A. B. and P. Zhu (2004). Another Look at the New York City School Voucher Experiment. *American Behavioral Scientist 47* (5), 658–698.

Kruppe, T. (2008). Selektivität bei der Einlösung von Bildungsgutscheinen. IAB Discussion Paper 17/2008, Institute for Employment Research (IAB), Nuremberg.

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

LaLonde, R. J. (2003). Employment and Training Programs. In M. Feldstien and R. Moffitt (Eds.), *Means Tested Transfer Programs in the U.S.* Chicago: University of Chicago Press for the National Bureau of Economic Research.

Lechner, M., R. Miquel, and C. Wunsch (2005). Long-Run Effects of Public Sector Sponsored Training in West Germany. IAB Discussion Paper 3/2005, Institute for Employment Research (IAB), Nuremberg.

Lechner, M., R. Miquel, and C. Wunsch (2007). The Curse and Blessing of Training the Unemployed in a Changing Economy: The Case of East Germany After Unification. *German Economic Review 8* (4), 468–509.

Lechner, M. and J. A. Smith (2007). What is the Value Added by Caseworkers? *Labour Economics 14* (2), 135–151.

Lechner, M. and C. Wunsch (2006). Are Training Programs More Effective When Unemployment Is High? IZA Discussion Paper 2355, Institute for the Study of Labor (IZA), Bonn.

Lechner, M. and C. Wunsch (2008). What Did All the Money Do? On the General Ineffectiveness of Recent West German Labour Market Programmes. *Kyklos 61*(1), 134–174.

Leuven, E. and B. Sianesi (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components, Boston College Department of Economics. Available at http://ideas.repec.org/c/boc/bocode/s432001.html.

Manski, C. F. (1992). Educational Choice (Vouchers) and Social Mobility. *Economics of Education Review 11*(4), 351–369.

Mueser, P. R., K. R. Troske, and A. Gorislavsky (2007). Using State Administrative Data to Measure Program Performance. *The Review of Economics and Statistics 89*(4), 761–783.

Nechyba, T. J. (2000). Mobility, Targeting, and Private-school Vouchers. *American Economic Review 90*(1), 130–146.

Neyman, J. S. (1923). On the Application of Probability Theory to Agriculture Experiments. Essay on Principles. Section 9. *Roczniki Nauk Rolniczych Tom X*, 1–51 (Annals of Agricultural Sciences). Translated in: J. S. Neyman, D. M. Dabrowska and T. P. Speed (1990), *Statistical Science 5*(4), 465–472.

Rinne, U., M. Schneider, and A. Uhlendorff (2007). Too Bad to Benefit? Effect Heterogeneity of Public Training Programs. IZA Discussion Paper 3240, Institute for the Study of Labor (IZA), Bonn.

Rosenbaum, P. R. and D. B. Rubin (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika 70*(1), 41–55.

Rosenbaum, P. R. and D. B. Rubin (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score. *The American Statistican 39*(1), 33–38.

Roy, A. D. (1951). Some Thoughts on the Distribution of Earnings. *Oxford Economic Papers 3*, 135–146.

Rubin, D. B. (1973). The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies. *Biometrics 29*, 185–203.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies. *Journal of Educational Psychology 66*(5), 688–701.

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, Institute for the Study of Labor (IZA), Bonn.

Schneider, H. and A. Uhlendorff (2006). Die Wirkung der Hartz-Reform im Bereich der beruflichen Weiterbildung. *Journal for Labour Market Research 39*(3-4), 477–490.

Sianesi, B. (2004). An Evaluation of the Active Labor Market Programmes in Sweden. *The Review of Economics and Statistics 86*(1), 133–155.

Smith, J. A. and P. E. Todd (2005a). Does Matching Overcome LaLonde's Critique of Non-Experimental Estimators? *Journal of Econometrics 125*(1-2), 305–353.

Smith, J. A. and P. E. Todd (2005b). Rejoinder. *Journal of Econometrics 125*(1-2), 365–375.

Stephan, G. (2009). The Effects of Active Labor Market Programs in Germany: An Investigation Using Different Definitions of Non-treatment. *Jahrbücher für Nationalökonomie und Statistik* (forthcoming).

Steuerle, C. E. (2000). Common Issues for Voucher Programs. In C. E. Steuerle, V. Doorn Ooms, G. E. Peterson, and R. D. Reischauer (Eds.), *Vouchers and the Provision of Public Services*. Washington D.C.: Brookings Institution Press.

van Ours, J. C. (2004). The Locking-in Effect of Subsidized Jobs. *Journal of Comparative Economics 32*(1), 37–55.

Viitanen, T. (2007). Public versus Private Provision of Daycare: An Experimental Evaluation. IZA Discussion Paper 3009, Institute for the Study of Labor (IZA), Bonn.

Winterhager, H., A. Heinze, and A. Spermann (2006). Deregulating Job Placement in Europe: A Microeconometric Evaluation of an Innovative Voucher Scheme in Germany. *Labour Economics 13*(4), 505–517.

Zhao, Z. (2008). Sensitivity of Propensity Score Methods to the Specifications. *Economics Letters 98*(3), 309–319.

Figure 1: Entrants into public training programs, unemployment rate (2000–2006)
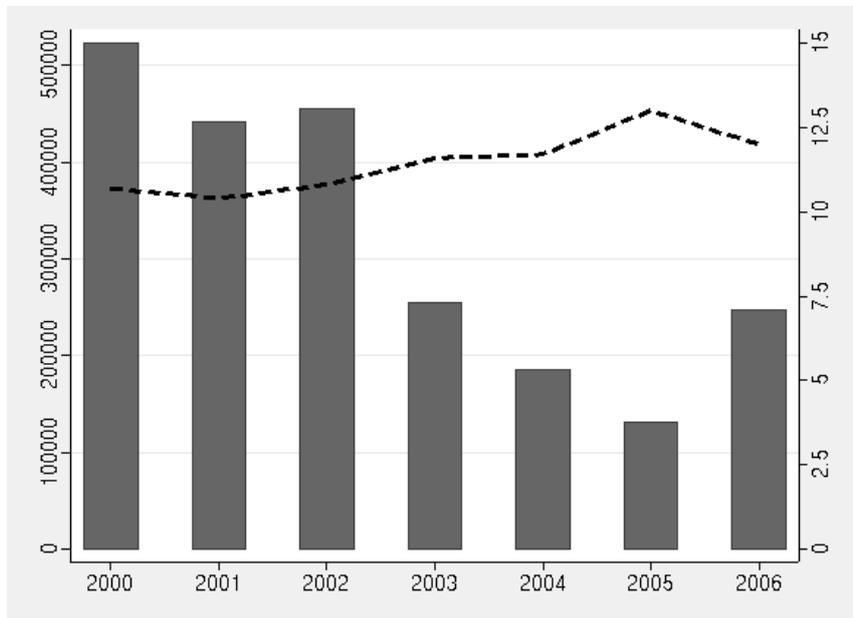
Note: Bars show annual number of entrants into public training programs (left axis). The dashed line represents the average unemployment rate (right axis, in percent).

Figure 2: Chronology of the Hartz reforms

2002

January 1, 2002: JobAQTIVE Law
→ Training: changes in benefit claims during and after participation
→ Job placement: profiling, aptitude tests, assignment of third parties
→ Compulsory registration for those threatened with unemployment
→ Changes in reasonability of job offers

2003

January 1, 2003: Hartz I/II
→ Training: introduction of voucher, stricter selection criteria
→ Temporary employment: Personal Service Agency (PSA)
→ Cut in unemployment assistance
→ Compulsory registration as job seeker at the time of layoff notification
→ Further changes in reasonability of job offers
→ Tightening of rules for exclusion from benefits
→ Extension of marginal employment legislation, start-up subsidies

January 1, 2004: Hartz III
→ Reorganization of the Federal Labor Agency
→ Legislative changes concerning job creation schemes (ABM/SAM)
→ Alteration of unemployment benefit legislation
→ Further tightening of rules for exclusion from benefits

2004

January 1, 2005: Hartz IV
→ Combination of unemployment assistance and social assistance
   to unemployment benefits II (means-tested)
→ Introduction of One-Euro-Jobs

Source: Authors' illustration.

27

Figure 3: Training Voucher

Training Voucher Number.: _____/_____
Customer Number: _____
(In accordance with § 77 para. 3 of the German Social Code III)

Valid until:

Costs are covered
    ☐ according to the certification procedure (lump-sum)
    ☐ according to actual costs (as evidence is provided)

Maximum duration: up to _____ months (including an internship if necessary)

Educational target / qualifications provided:

Type of course: ☐ full-time (35 hours/week) ☐ part-time (12–24 hours/week)
      ☐ on-the-job course   ☐ correspondence course

Provider: ☐ in-firm training   ☐ off-the-job training

Location: ☐ outside a daily commuting distance

Source: Authors' illustration.

Figure 4: Actual program duration
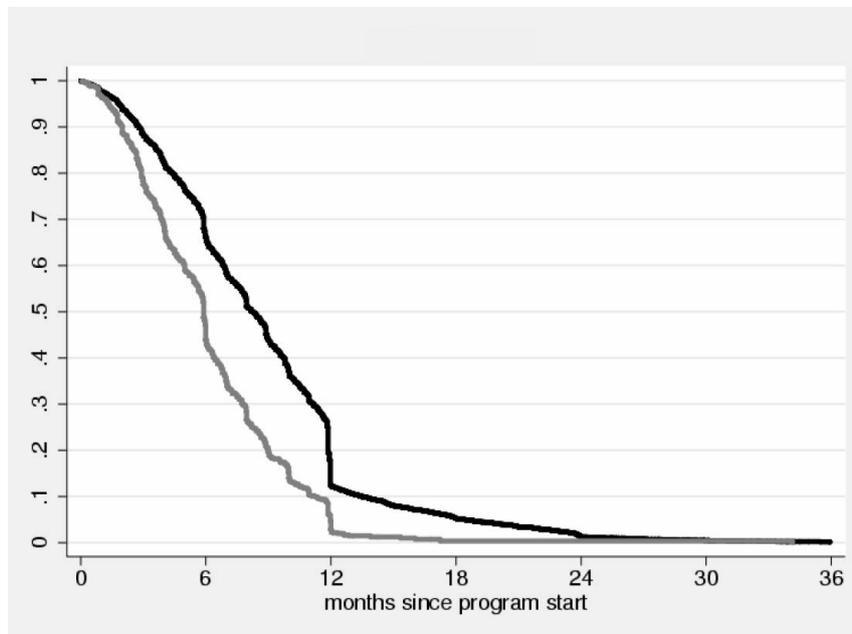
Source: IEB, own calculations.

Note: Kaplan–Meier Estimates. Pre-reform period in black, post-reform period in gray.

28

Table 1: Descriptive statistics (selected variables)

	Participants 2002		Participants 2003		Non-participants 2002		Non-participants 2003	
	Mean	(Std. Dev.)	Mean	(Std. Dev.)	Mean	(Std. Dev.)	Mean	(Std. Dev.)
Socio-demographic characteristics								
Female	0.4156	(0.4928)	0.4224	(0.4941)	0.4271	(0.4947)	0.4293	(0.4950)
East	0.4298	(0.4951)	0.4086	(0.4918)	0.3029	(0.4595)	0.3115	(0.4631)
Age	37.248	(9.5597)	35.871	(9.7840)	39.975	(11.439)	39.547	(11.143)
German citizenship	0.9179	(0.2745)	0.9236	(0.2657)	0.8693	(0.3370)	0.8829	(0.3215)
Educational attainment								
No graduation	0.0545	(0.2270)	0.0601	(0.2378)	0.1147	(0.3186)	0.1140	(0.3179)
First stage of secondary level	0.3263	(0.4689)	0.2965	(0.4569)	0.4922	(0.4999)	0.4722	(0.4992)
Second stage of secondary level	0.4207	(0.4937)	0.4452	(0.4972)	0.2879	(0.4528)	0.3033	(0.4597)
Advanced technical college entrance qualification	0.0664	(0.2490)	0.0772	(0.2670)	0.0365	(0.1875)	0.0399	(0.1957)
General qualification for university entrance	0.1321	(0.3386)	0.1210	(0.3263)	0.0687	(0.2530)	0.0705	(0.2561)
Vocational attainment								
No vocational degree	0.1808	(0.3848)	0.1901	(0.3925)	0.3389	(0.4733)	0.3230	(0.4676)
In-plant training	0.6253	(0.4841)	0.6182	(0.4860)	0.5519	(0.4973)	0.5555	(0.4969)
Off-the-job training, vocational school, technical school	0.0953	(0.2936)	0.0983	(0.2978)	0.0659	(0.2481)	0.0727	(0.2597)
University, advanced technical college	0.0986	(0.2982)	0.0934	(0.2911)	0.0433	(0.2034)	0.0487	(0.2153)
Unemployment history								
Share of unemployment in 1st year before program entry	0.5942	(0.3080)	0.5972	(0.3059)	0.6225	(0.3192)	0.6320	(0.3175)
Share of unemployment in 2nd year before program entry	0.2479	(0.3293)	0.2332	(0.3209)	0.3384	(0.3688)	0.3476	(0.3740)
Share of unemployment in 3rd year before program entry	0.2065	(0.3164)	0.1613	(0.2770)	0.3016	(0.3743)	0.2894	(0.3669)
Share of unemployment in 4th year before program entry	0.1851	(0.3032)	0.1462	(0.2676)	0.2715	(0.3643)	0.2636	(0.3608)
Employment history								
Share of employment in 1st year before program entry	0.2687	(0.3057)	0.2560	(0.2953)	0.2260	(0.2984)	0.2087	(0.2867)
Share of employment in 2nd year before program entry	0.5176	(0.4276)	0.5604	(0.4190)	0.4181	(0.4186)	0.4208	(0.4198)
Share of employment in 3rd year before program entry	0.5287	(0.4380)	0.5897	(0.4279)	0.4438	(0.4329)	0.4648	(0.4314)
Share of employment in 4th year before program entry	0.5264	(0.4379)	0.5840	(0.4282)	0.4568	(0.4344)	0.4888	(0.4303)
ln(Last income from employment)	3.8728	(0.6220)	3.8825	(0.6179)	3.7950	(0.6383)	3.7901	(0.6377)
# Observations	22,839		1,231		520,483		532,893	

Source: IEB, own calculations.

Table 2: Matching quality

Sex	Region	Before Matching		After Matching		
		Mean bias	R^2	Mean bias	R^2	# sign.
a) Two-step matching: Participants 2002 – Participants 2003						
Female	East Germany	8.359	0.2001	8.652	0.1329	0
Female	West Germany	10.211	0.2125	6.760	0.1054	0
Male	East Germany	7.493	0.0990	5.135	0.0595	0
Male	West Germany	8.184	0.1158	5.972	0.0769	0
b) One-step matching: Participants 2002 – Non-participants 2002						
Female	East Germany	11.621	0.0536	1.059	0.0017	4
Female	West Germany	11.991	0.0693	0.982	0.0013	2
Male	East Germany	11.730	0.0562	1.050	0.0015	0
Male	West Germany	11.725	0.0584	0.895	0.0009	0
c) One-step matching: Participants 2003 – Non-participants 2003						
Female	East Germany	13.461	0.0486	4.329	0.0337	0
Female	West Germany	15.383	0.0558	3.714	0.0278	0
Male	East Germany	14.231	0.0535	4.064	0.0271	1
Male	West Germany	13.358	0.0480	2.241	0.0106	0
d) Two-step matching: Matched Participants 2002 – Non-participants 2002						
Female	East Germany	24.504	0.1092	5.811	0.0385	0
Female	West Germany	25.836	0.1068	3.402	0.0247	0
Male	East Germany	20.709	0.0881	2.969	0.0214	0
Male	West Germany	22.185	0.0724	2.629	0.0204	0
e) Two-step matching: Matched Participants 2003 – Non-participants 2003						
Female	East Germany	21.841	0.0859	3.924	0.0333	0
Female	West Germany	23.526	0.0946	2.581	0.0121	0
Male	East Germany	18.842	0.0664	2.916	0.0219	0
Male	West Germany	18.242	0.0714	2.329	0.0151	0

Notes: Mean Bias: Mean standardized bias; R^2: Pseudo-R^2 of propensity score estimation; # sign.: number observable characteristics for which F-test rejects the joint null. Further details are given in the text.

Figure 5: Pre-entry employment shares: one-step matching procedure
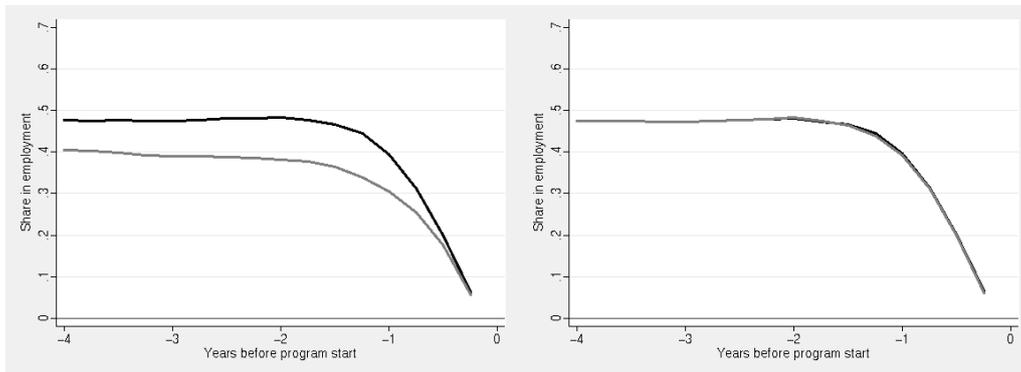
(a) unmatched P 2002 – unmatched NP 2002   (b) matched P 2002 – matched NP 2002
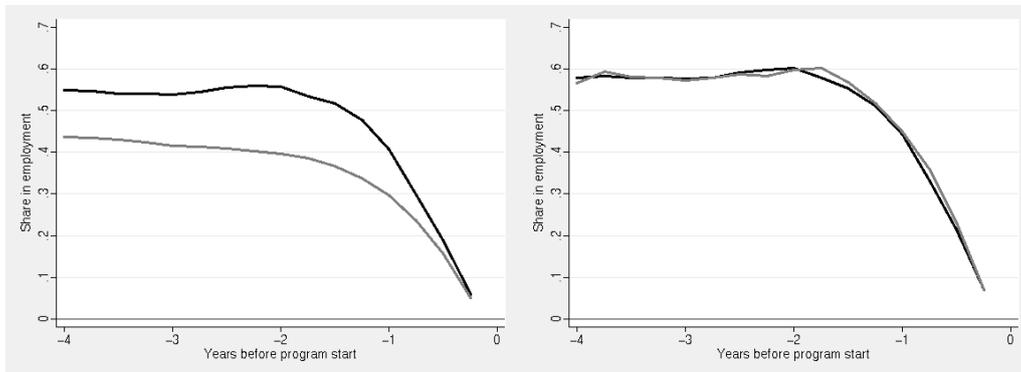
(c) unmatched P 2003 – unmatched NP 2003   (d) matched P 2003 – matched NP 2003

Notes: P: participants (in black); NP: non-participants (in gray).

Figure 6: Pre-entry employment shares: two-step matching procedure
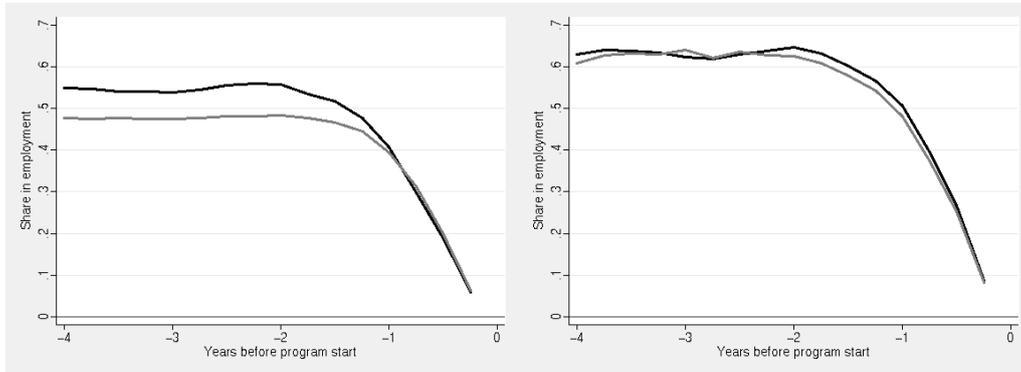
(a) unmatched P 2002 – unmatched P 2003   (b) matched P 2002 – matched P 2003
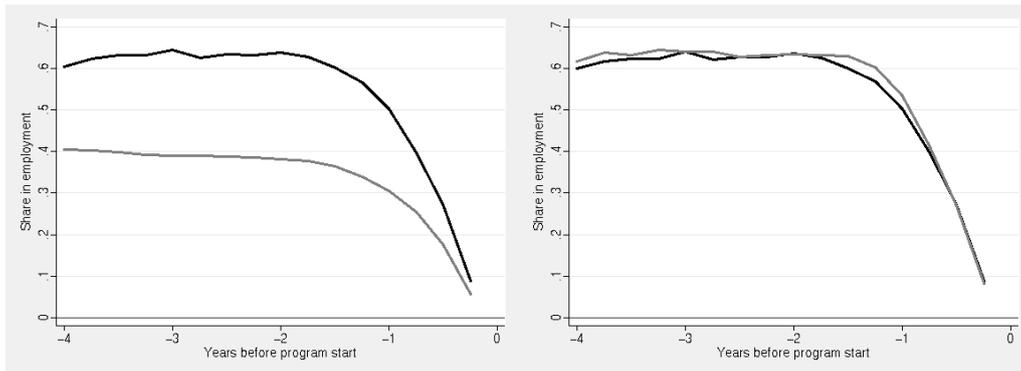
(c) matched P 2002 – unmatched NP 2002   (d) matched P 2002 – matched NP 2002
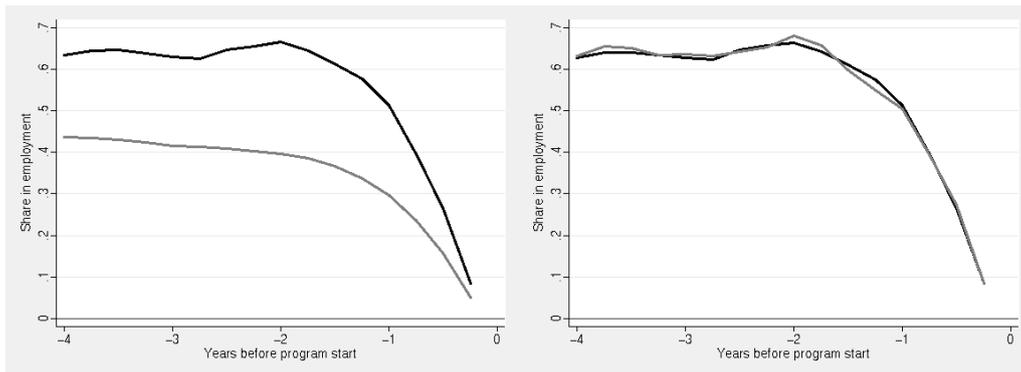
(e) matched P 2003 – unmatched NP 2003   (f) matched P 2003 – matched NP 2003

Notes: P: participants (in black); NP: non-participants (in gray). Matching between participants 2002 and participants 2003: pre-reform participants in gray, post-reform participants in black.

32

Figure 7: Reform effect, employment
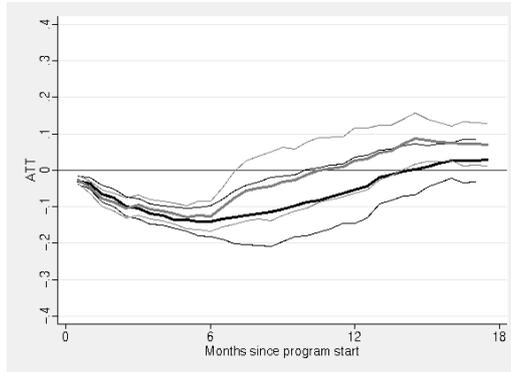
Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure 8: Decomposition, employment
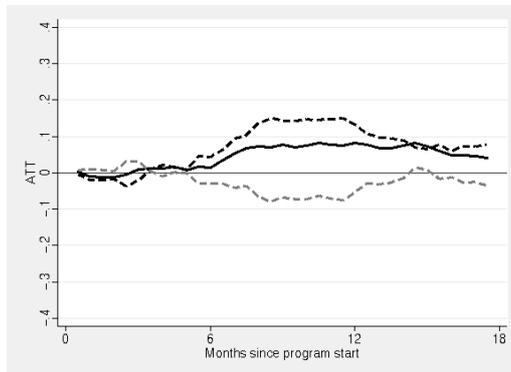
(a) Decomposition
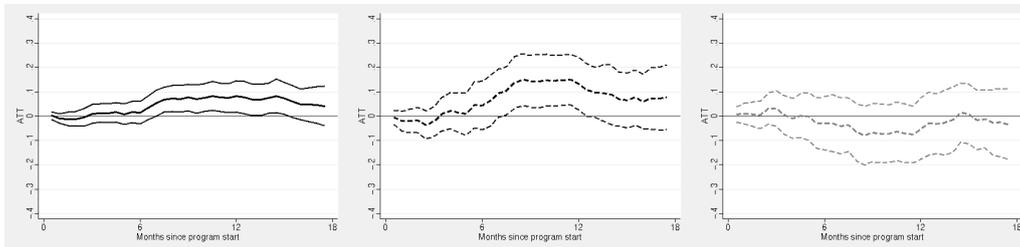
(b) Reform effect   (c) Voucher effect   (d) Selection effect
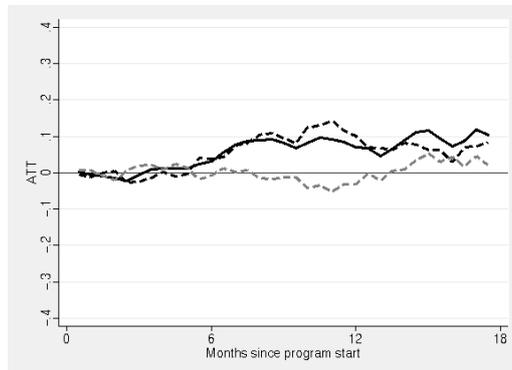
Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

33

Figure 9: Reform effect, earnings (definition A)
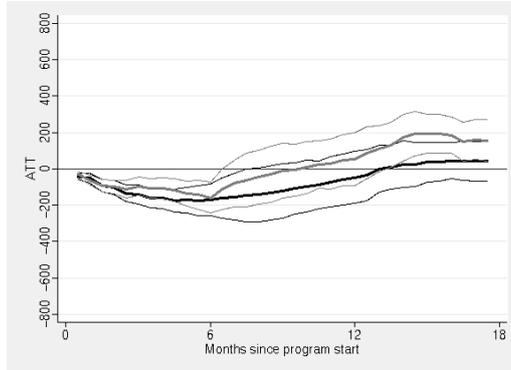
Definition A: Monthly earnings where no earnings are treated as zero.
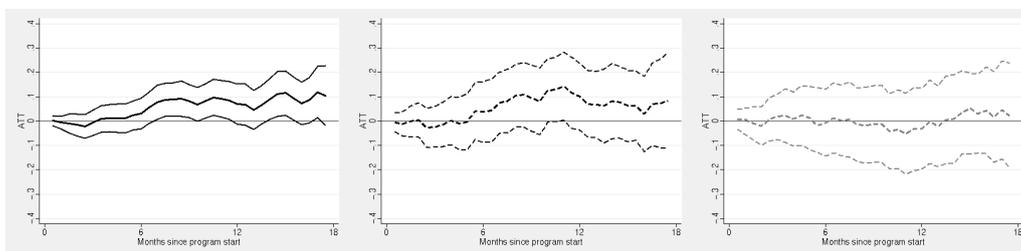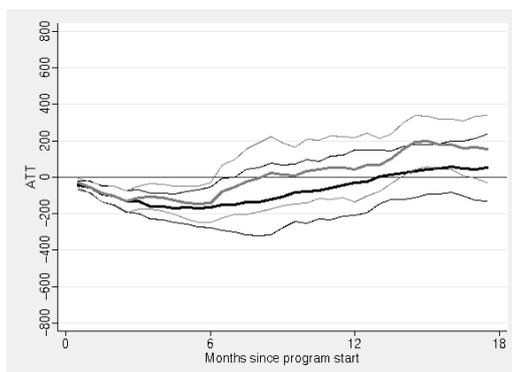
Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure 10: Decomposition, earnings (definition A)
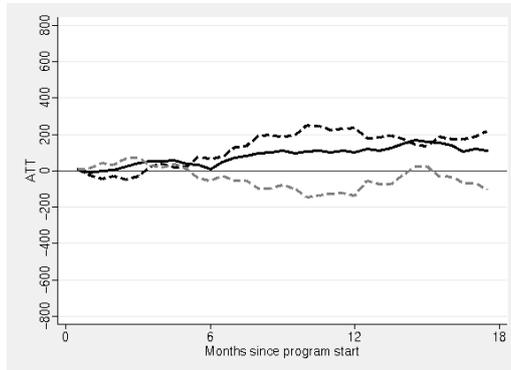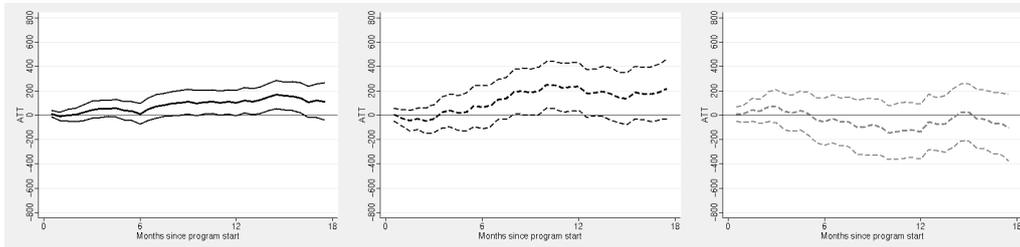
(a) Decomposition

(b) Reform effect (c) Voucher effect (d) Selection effect

Definition A: Monthly earnings where no earnings are treated as zero.

Note: Total reform effect in black (solid), voucher effect in black (dashed), and selection effect in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.
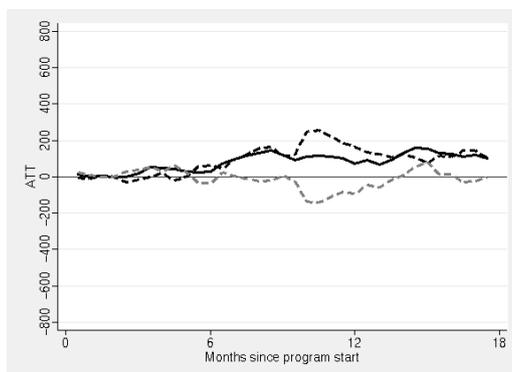
Figure 11: Reform effect, earnings (definition B)
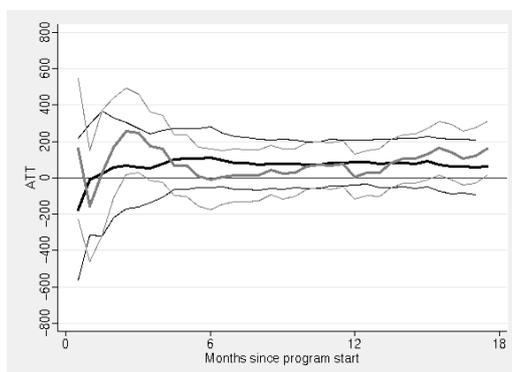
Definition B: Monthly earnings where no earnings are treated as missing.

Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure 12: Decomposition, earnings (definition B)
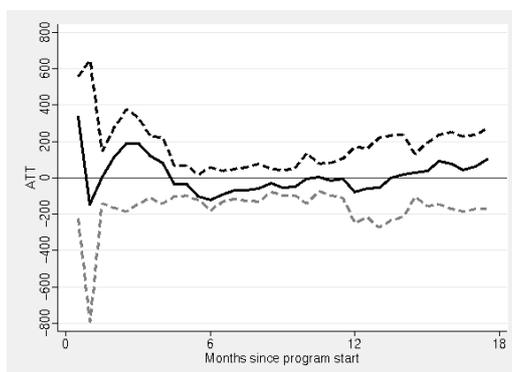
(a) Decomposition
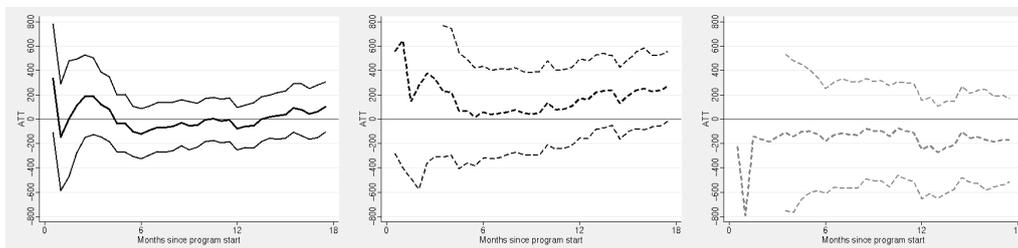
(b) Reform effect   (c) Voucher effect   (d) Selection effect

Definition B: Monthly earnings where no earnings are treated as missing.

Note: Reform effect in black (solid), voucher effect in black (dashed), and selection effect in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

35

Figure 13: Reform effect, employment (including additional control variables)
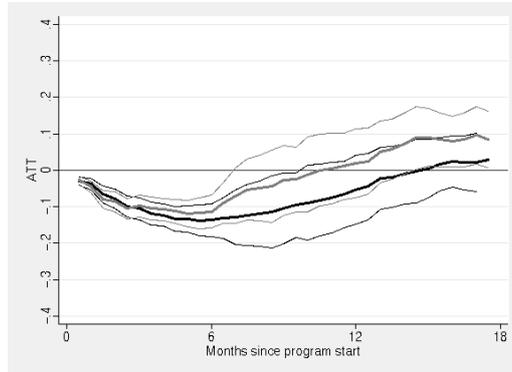
Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure 14: Decomposition, employment (including additional control variables)
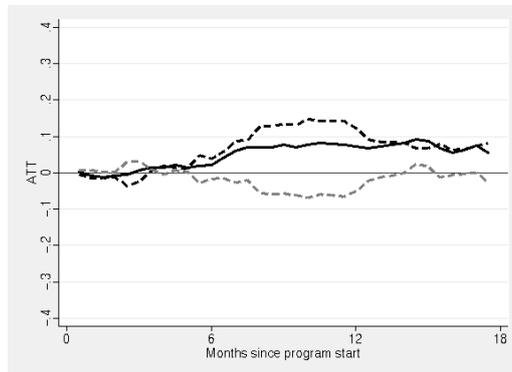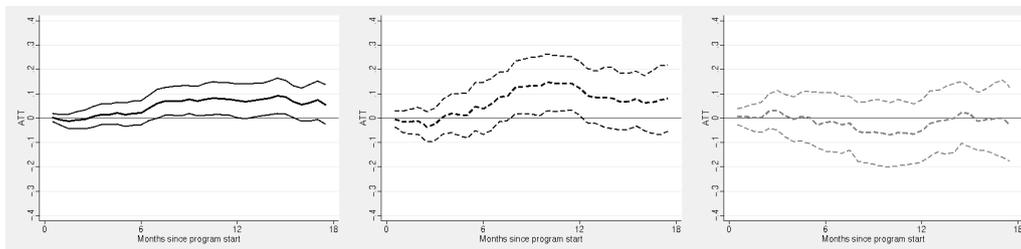
(a) Decomposition

(b) Reform effect   (c) Voucher effect   (d) Selection effect

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure 15: Reform effect, earnings (definition A, including additional control variables)
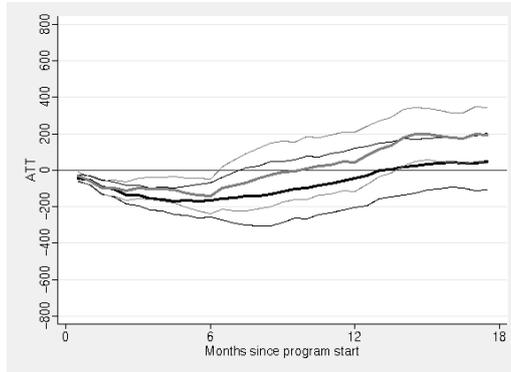
Definition A: Monthly earnings where no earnings are treated as zero.

Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure 16: Decomposition, earnings (definition A, including additional control variables)
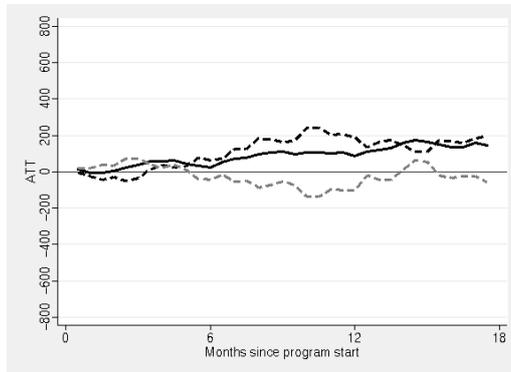
(a) Decomposition

(b) Reform effect　　　　　(c) Voucher effect　　　　　(d) Selection effect

Definition A: Monthly earnings where no earnings are treated as zero.

Note: Total reform effect in black (solid), voucher effect in black (dashed), and selection effect in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

37
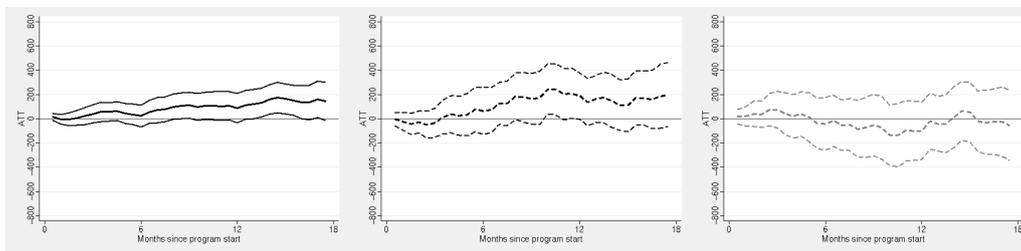
Figure 17: Reform effect, employment (excluding first quarter, including additional control variables)
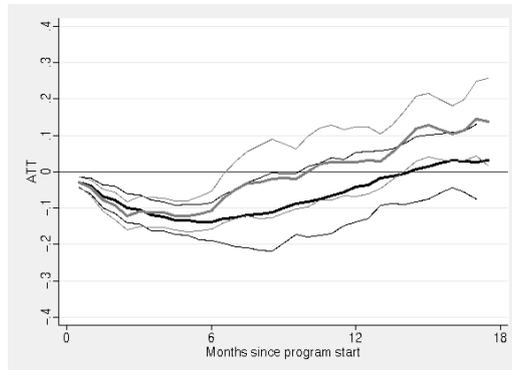
Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure 18: Decomposition, employment (excluding first quarter, including additional control variables)

(a) Decomposition

(b) Reform effect   (c) Voucher effect   (d) Selection effect

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

38

Figure 19: Reform effect, earnings (excluding first quarter, definition A, including additional control variables)

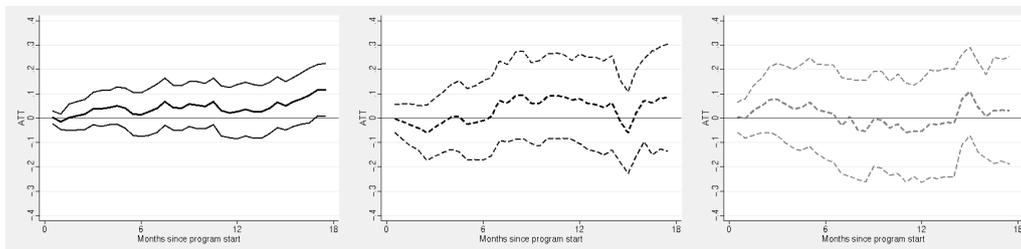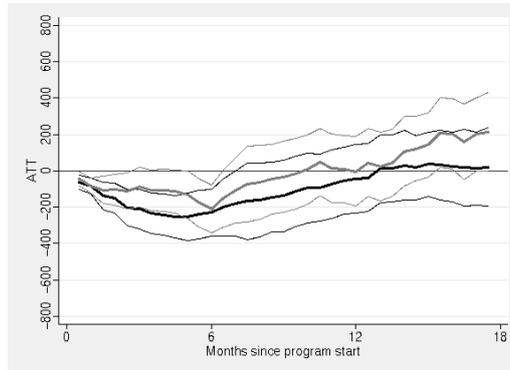*Definition A:* Monthly earnings where no earnings are treated as zero.

Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure 20: Decomposition, earnings (excluding first quarter, definition A, including additional control variables)

(a) Decomposition

(b) Reform effect   (c) Voucher effect   (d) Selection effect

Definition A: Monthly earnings where no earnings are treated as zero.

Note: Total reform effect in black (solid), voucher effect in black (dashed), and selection effect in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

39
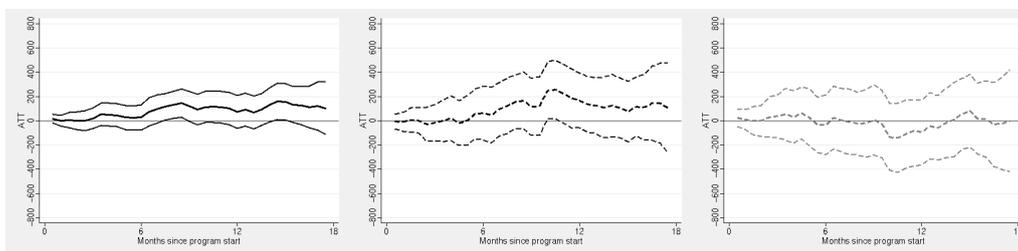
Appendix

Table A1: Economic and labor market variables

	2002 Mean (Std. Dev.)	2003 Mean (Std. Dev.)	2004 Mean (Std. Dev.)
1) Job seekers	0.1238 (0.0630)	0.1329 (0.0603)	0.1441 (0.0617)
2) Unemployment rate	0.0991 (0.0499)	0.1057 (0.0491)	0.1065 (0.0490)
3) Vacancies	0.0107 (0.0050)	0.0082 (0.0044)	0.0065 (0.0038)
Participants in ...			
4) Public training programs	0.0085 (0.0052)	0.0065 (0.0037)	0.0045 (0.0023)
5) Subsidized employment	0.0037 (0.0043)	0.0040 (0.0049)	0.0029 (0.0037)
6) Job creation schemes	0.0050 (0.0076)	0.0037 (0.0059)	0.0029 (0.0047)
7) GDP growth rate	1.5191 (1.3170)	1.0258 (0.6455)	2.2837 (0.7575)

Source: Federal Employment Agency (FEA); Statistical Offices of the Federal States.

Notes: 1)–6) are monthly shares in the civilian labor force in 178 FEA districts. 7) are annual GDP growth rates for the 16 Federal States.

Table A2: Rating of the Hartz reforms by FEA districts

How did the reforms *affect the ...*	negative −2	−1	neutral 0	positive +1	+2	missing .
...effectiveness of job placement	1 0.6	8 4.9	81 49.7	66 40.5	3 1.8	4 2.5
...process of job placement	1 0.6	9 5.5	61 37.4	82 50.3	6 3.7	4 2.5
...efficiency of job placement	1 0.6	7 4.3	64 39.3	83 50.9	5 3.1	3 1.8
...process of benefit granting	2 1.2	14 8.6	67 41.1	64 39.3	13 8.0	3 1.8
...co-operation with third parties	2 1.2	34 20.9	108 66.3	15 9.2	1 0.6	3 1.8
...administration effort	23 14.1	60 36.8	56 34.4	18 11.0	1 0.6	5 3.1
...matching accuracy of job placement	1 0.6	6 3.7	71 43.6	75 46.0	6 3.7	4 2.5

Source: Survey in 163 FEA districts conducted in the beginning of 2005.

Notes: First row: frequencies; second row: percentages.

Figure A1: Reform effect, West German Men, employment
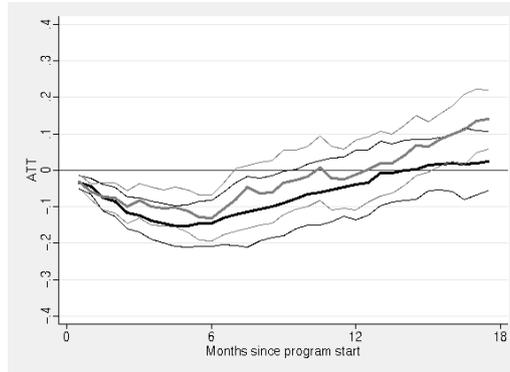
Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

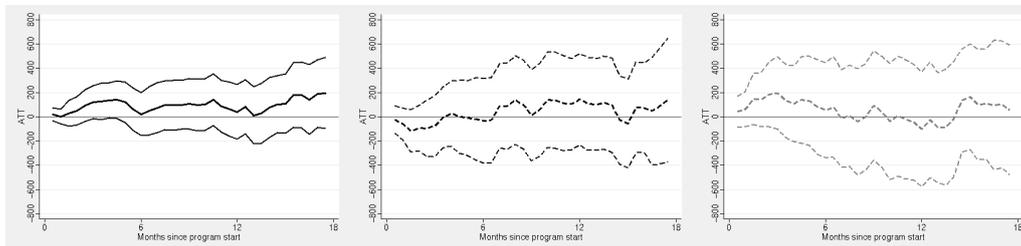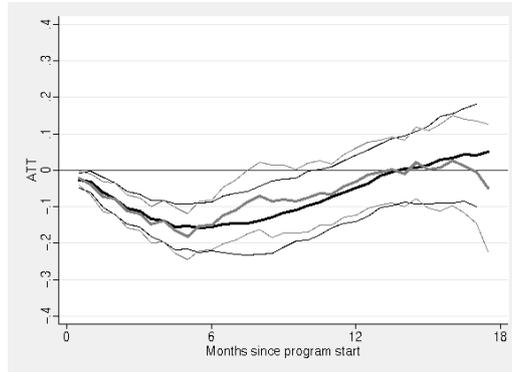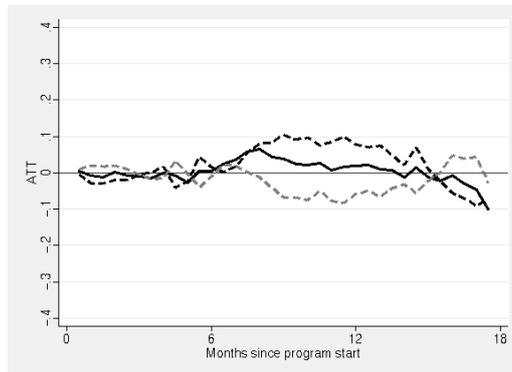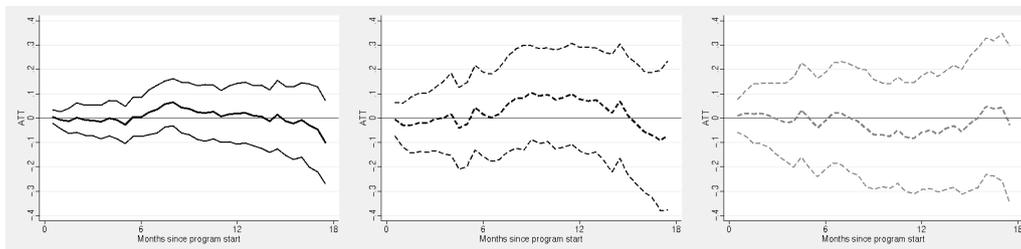Figure A2: Decomposition, West German Men, employment
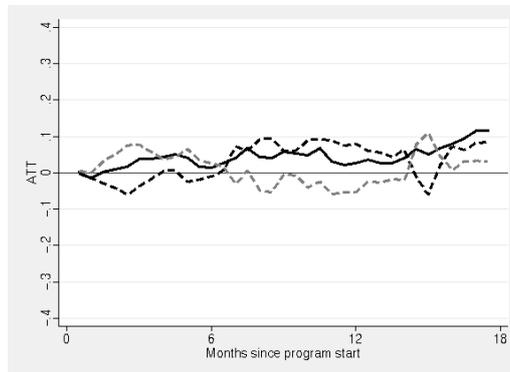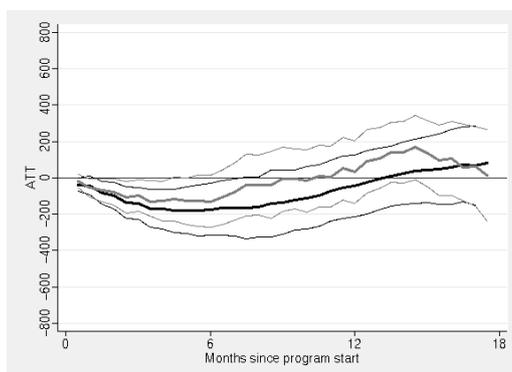

(a) Decomposition

(b) Reform effect   (c) Voucher effect   (d) Selection effect

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.
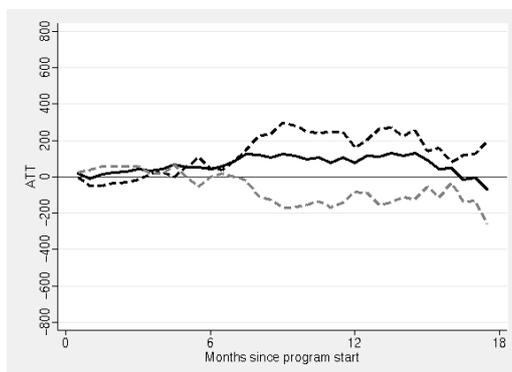
41

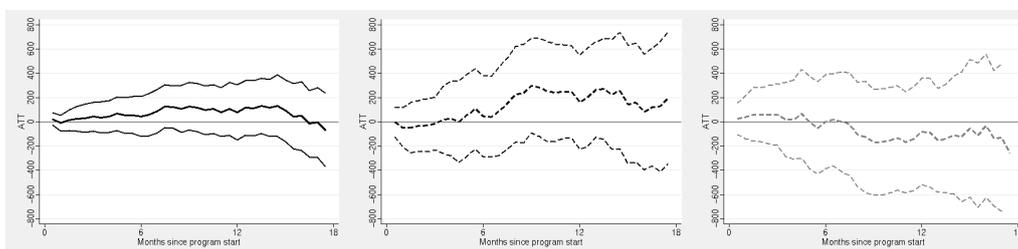Figure A3: Reform effect, West German Men, earnings (definition A)

Definition A: Monthly earnings where no earnings are treated as zero.

Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure A4: Decomposition, West German Men, earnings (definition A)
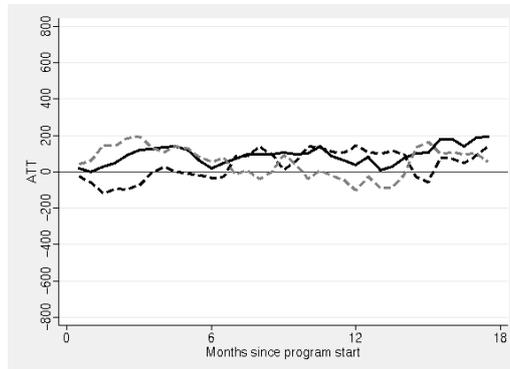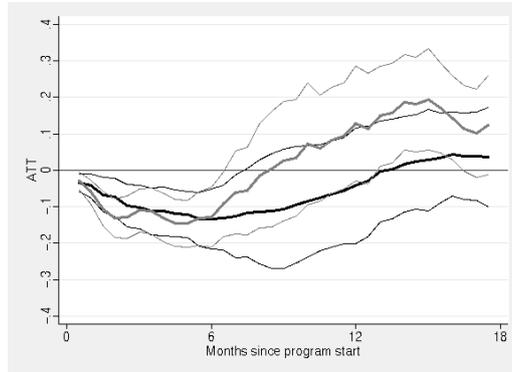

(a) Decomposition

(b) Reform effect (c) Voucher effect (d) Selection effect

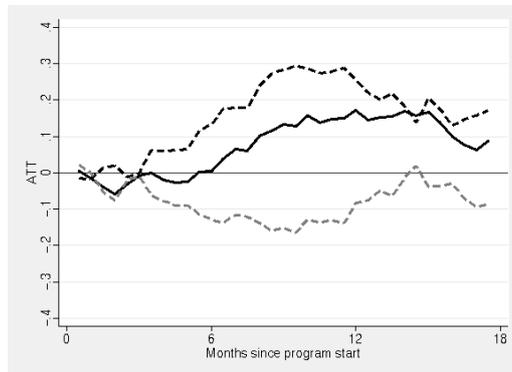*Definition A:* Monthly earnings where no earnings are treated as zero.

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

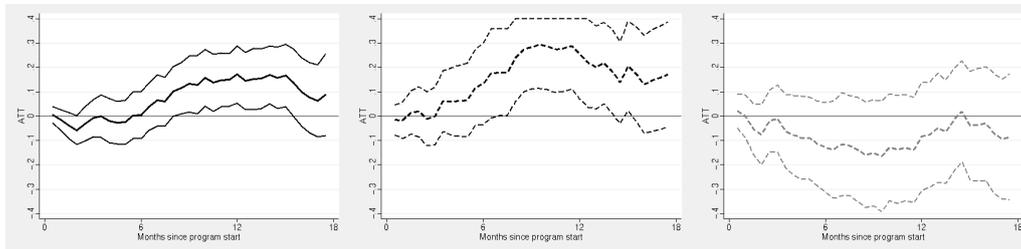Figure A5: Reform effect, West German Women, employment

Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure A6: Decomposition, West German Women, employment
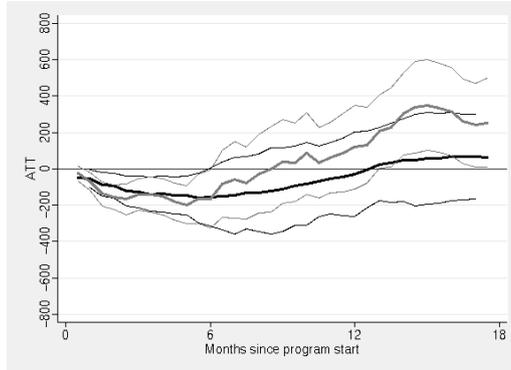

(a) Decomposition

(b) Reform effect   (c) Voucher effect   (d) Selection effect

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure A7: Reform effect, West German Women, earnings (definition A)

Definition A: Monthly earnings where no earnings are treated as zero.

Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure A8: Decomposition, West German Women, earnings (definition A)

(a) Decomposition

(b) Reform effect   (c) Voucher effect   (d) Selection effect

Definition A: Monthly earnings where no earnings are treated as zero.

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.
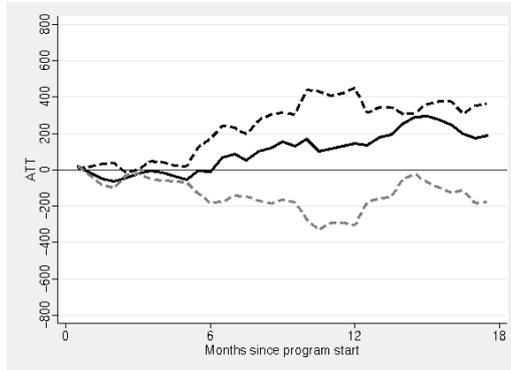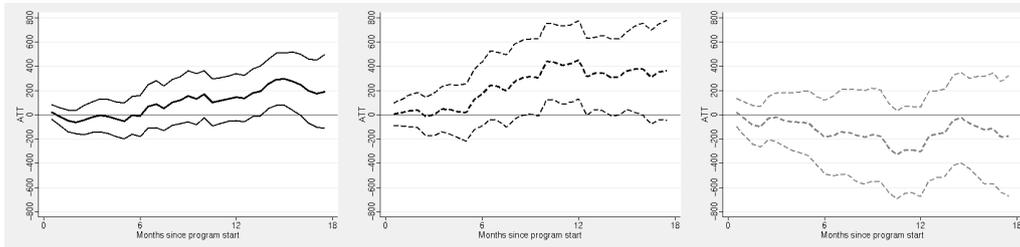
44

Figure A9: Reform effect, East German Men, employment

Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure A10: Decomposition, East German Men, employment

(a) Decomposition

(b) Reform effect   (c) Voucher effect   (d) Selection effect

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

45

Figure A11: Reform effect, East German Men, earnings (definition A)

Definition A: Monthly earnings where no earnings are treated as zero.

Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure A12: Decomposition, East German Men, earnings (definition A)
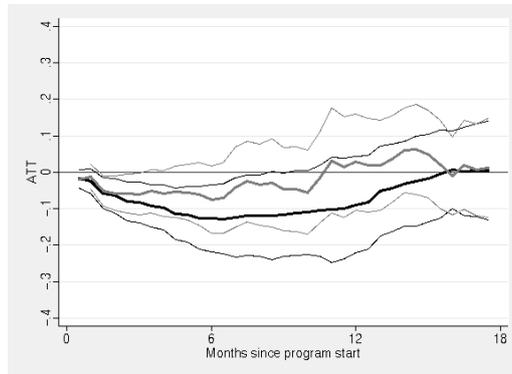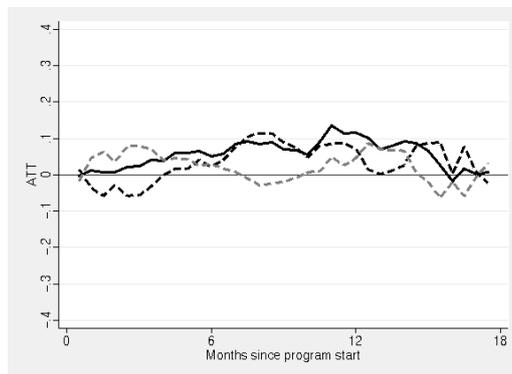

(a) Decomposition

(b) Reform effect   (c) Voucher effect   (d) Selection effect

Definition A: Monthly earnings where no earnings are treated as zero.

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

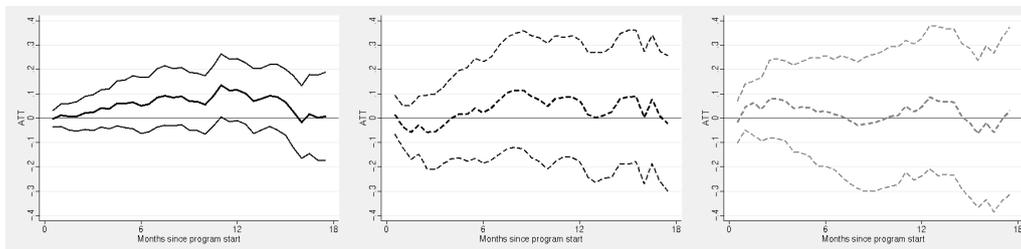Figure A13: Reform effect, East German Women, employment

Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure A14: Decomposition, East German Women, employment
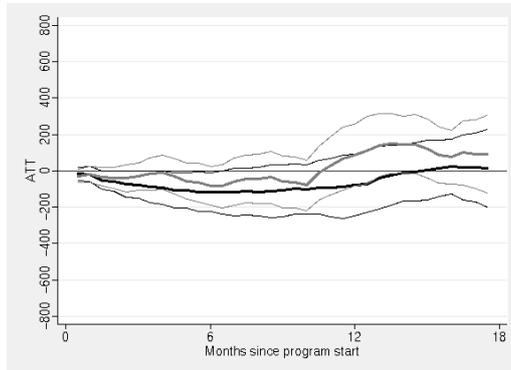

(a) Decomposition

(b) Reform effect   (c) Voucher effect   (d) Selection effect

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.
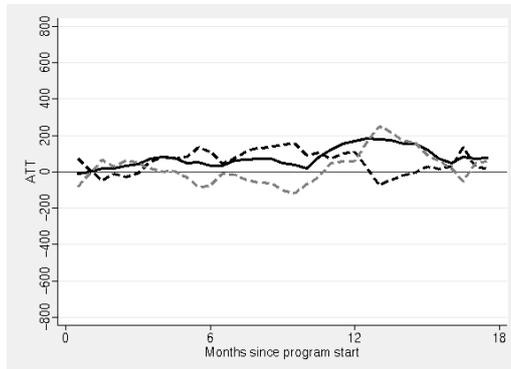
47

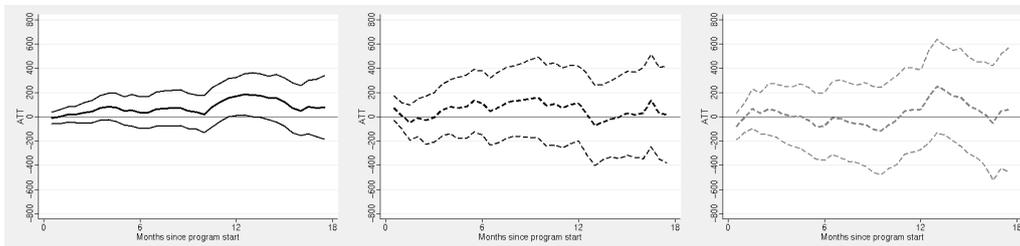Figure A15: Reform effect, East German Women, earnings (definition A)

Definition A: Monthly earnings where no earnings are treated as zero.

Note: Pre-reform period in black, post-reform period in gray. Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure A16: Decomposition, East German Women, earnings (definition A)

(a) Decomposition

(b) Reform effect   (c) Voucher effect   (d) Selection effect

Definition A: Monthly earnings where no earnings are treated as zero.

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed). Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.