Vouchers, Caseworkers, and the Reform: Public Training Programs in Germany*

Ulf Rinne[†]
IZA
Free University of Berlin

Arne Uhlendorff[‡]
IZA
DIW DC

Zhong Zhao[§]
IZA

July 15, 2008

Abstract

This paper studies the role of training vouchers in public training programs. Using a rich administrative data set from the Federal Employment Agency in Germany, we apply regression and propensity score matching procedures to measure the effect of the Hartz reform in 2003, which introduced training vouchers and imposed more selective criteria on the applicants. Besides estimating the total reform effect, we isolate the effect induced by changes in the composition of program participants (selection effect) from the effect based on the introduction of vouchers (voucher effect). Our results indicate that the selection effect plays—at the best—only a minor role in explaining the generally positive impacts of the reform. The most striking picture emerges with respect to the most important program type in terms of participants, for which we find a clearly negative selection effect. We thus conclude that the reform effect for this type would have been even more positive if the composition of participants had not changed after the reform.

Keywords: Active Labor Market Policy; Program Evaluation; Propensity Score Matching; Voucher; Hartz Reform

JEL Classification: J64, J68, H43

*The data used in this paper originate from the evaluation of public training programs as part of the evaluation of the proposals of the Hartz Commission. Schneider et al. (2007) contains details. We would like to thank Hilmar Schneider, Marc Schneider, Marton Csillag, participants at the IZA Conference on the Evaluation of Labor Market Programs, the EALE Conference in Oslo, the EEA Meeting in Budapest, the CAPE in Nuremberg, the Conference of the German Statistical Society in Kiel, the VfS Conference in Munich, and the ESPE Congress in London for valuable discussions and helpful comments. All remaining errors are our own.

[†]Ulf Rinne, IZA, P.O. Box 7240, D-53072 Bonn. Email: rinne@iza.org

[‡]Arne Uhlendorff, IZA, P.O. Box 7240, D-53072 Bonn. Email: uhlendorff@iza.org

[§]Zhong Zhao, IZA, P.O. Box 7240, D-53072 Bonn. Email: zhao@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. Importantly, an innovative voucher scheme was introduced. 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. Competition among providers should thereby increase, and also program quality. While vouchers are a common instrument in other fields of public services, this approach is novel in the context of delivering ALMP.

From a theoretical point of view increased consumer choice and provider competition are the main arguments in favor of the new system. Consumers of training programs should be able to express their preferences optimally and thus maximize their utility. But there may be obstacles in case of public services which prevent from reaching higher levels of efficiency. For example, information asymmetries can lead to choices which do not truly reflect consumers' preferences. On the other hand, increased competition among providers is supposed to arise because of the opportunity for potential competitors to enter the market at little or no costs—the market then resembles a contestable market (Baumol, 1982). This could be the case in the post-reform regime, although for example reputation and brand names are potential barriers to entry for training providers which may remain.

Vouchers are a common instrument in other fields of public services—in particular in the field of education—and are quite extensively studied in the literature.[1] There are also studies on vouchers for pre-school education (e.g., Viitanen, 2007), but most of the examples focus on school education. For instance, Manski (1992) provides a theoretical model and presents simulation results on school vouchers and social mobility. Using a general equilibrium model, Nechyba (2000) studies private-school vouchers and residential mobility. Angrist et al. (2002) analyze the academic outcomes of students who win voucher lotteries. Ladd (2002) presents a review of

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

major studies on school vouchers. More recently, Krueger and Zhu (2004) reexamine the New York City school voucher experiment, and Brunello and Checchi (2004) analyze the impact of vouchers for private schools introduced in some Italian regions. So far the overall picture that can be drawn from these studies is rather inconclusive, and the results are not very robust. For instance, studies with U.S. data typically indicate to have insufficient data to draw clear conclusions about the net effects on, e.g., 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. On the other hand, 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 are only few studies of vouchers for adult education—mainly because vouchers are not very common in this context. Exceptions are mentioned in Dohmen (2003, 2007) which include the U.S. experience with the GI Bill and the Walloon training voucher scheme in Belgium. Another example is a World Bank initiative launched in Kenya in which training vouchers are given to entrepreneurs in the informal sector (see, e.g., Oosterbeek, 1998). The major difference between school vouchers and vouchers for adult education is that in the former case parents make the decision for their children, while in the latter situation participants make their own decision on how to use the voucher. The results on vouchers for adult education are in general rather positive: The GI Bill increased university enrollment (Levin, 2003), preliminary evidence suggests a successful start of the Belgian program (Geers, 2001), and vouchers had a beneficial impact on participants in Kenya although the impact on the training market was only modest (Hallberg, 2006).

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. An increase in consumer sovereignty was also put forward as an argument in support of the voucher system. Winterhager et al. (2006) evaluate the effectiveness of this instrument and find a positive impact on the employment probability

of voucher recipients in Western Germany. However, these vouchers are of little practical relevance as the take-up rate amounts to only about 5 percent receiving a voucher. Bruttel (2005) gives some reasons for the little relevance (e.g., the complicated design, the low incentives for private providers).

This paper focuses on a voucher for adult education in the context of ALMP. We estimate the impact of the introduction of training vouchers on the effectiveness of public training programs. But next to this feature, other elements of the reform may also affect program effectiveness. These elements include a stricter selection rule for participants and a matching process between program types and participants by the caseworkers which is based on the expected reemployment probability. The overall effect of the Hartz reform could thus result from the introduction of the voucher system and/or from a change in the composition of participants because of the new selection criteria. We refer to the former as the voucher effect and to the latter as the selection effect. To decompose the overall reform effect into these two effects, we apply regression analysis and a two-step propensity score matching procedure to a rich administrative data set.

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 data and the program types being analyzed in Section 3, we outline the analytical framework in Section 4. Section 5 presents our results and, finally, Section 6 concludes.

2 Institutional Background

Germany's ALMP aims to increase the employment prospects of unemployed individuals. For this purpose, the Federal Employment Agency (FEA) spends a substantial amount of money on measures 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

On the other hand, Figure 2 indicates that direct program costs did not decrease to a similar extent. Also the amount of benefit claims during program participation—as important indirect costs of participation, did not markedly decrease in 2003 compared to 2002. However, we do observe a substantial drop in this regard between December 2002 and January 2003.

Figure 2 about here

The effectiveness of public training programs in Germany before the Hartz reform has been subject to a number of studies. For a recent review of the results see, e.g., Caliendo and Steiner (2005).[2] The results are quite heterogeneous—depending on the investigation period and the underlying data set. While earlier studies often find insignificant or even negative effects (Lechner, 1999, 2000; Hujer and Wellner, 2000), most of the recent studies which are based on rich administrative data sets find at least for some sub-groups positive treatment effects (Lechner et al., 2005, 2007; Fitzenberger et al., 2006; Biewen et al., 2007; Rinne et al., 2007). But 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 (2008) finds that estimated treatment effects differ considerably across different definitions of non-participation. Overall, the major lesson from the evaluation studies conducted for 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.

[2]The international literature on the evaluation of ALMP is summarized by Martin and Grubb (2001) and Kluve (2006), among others.

As mentioned above Germany's ALMP has undergone a series of reforms during the past years. Figure 3 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 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 3 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). We will thus refer to this legislative changes as 'the reform' throughout the course of this paper and discuss these changes in detail below. 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, One-Euro-Jobs were introduced.

As stated above, the provision of public training programs was substantially altered in the beginning of 2003. Prior to the Hartz I/II package 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 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 4 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.[3] 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

[3]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).

onwards (Schneider et al., 2007).[4]

Figure 4 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-qualified job seekers lack the abilities to navigate the training market and to take an active role in searching for an appropriate course. This argument is supported by Kruppe (2008) who finds that low-qualified individuals are significantly less likely to redeem a granted voucher than persons with higher qualifications.[5] 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 (e.g., they do not offer 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 generally increases after the reform, while the magnitude of the increased efficacy considerably differs across program types. Nonetheless, the question which features of the reform cause this increase—and to what particular

[4]The official transitional arrangement was as follows: "Individuals who were counselled 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."

[5]However, the overall redemption rate is comparatively high with 85 percent in the period from 2003 to 2006 (Kruppe, 2008).

extent—remains unanswered. We answer this question by decomposing the reform effect into two components: *a*) the selection effect, and *b*) the 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-qualified job seekers lack the abilities to navigate the training market and to take an active role in searching for an appropriate course. Low-qualified individuals thus exhibit a lower voucher receipt rate (intended) 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.

3 Data

We use a sample of a particularly rich administrative data set, the Integrated Employment Biographies (IEB) of the FEA.[6] 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, job search, 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.

Since the public training programs currently in place in Germany are quite heterogenous, we concentrate on and differentiate between four particular types: *a*) type 1: occupation-related or general training, *b*) type 2: practice training in key qualifications, and *c*) type 3: practice firm.

[6]The IEB 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.,Hummel et al. (2005) for details on the IEBS. The IEB consists of 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). For a detailed description see, e.g., Schneider et al. (2007).

Participants in type 1 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). Type 2 is a predominantly practically oriented program with only few theoretical parts. It follows the principle 'learning by doing' and is often combined with internships. Within type 3 the simulation of real operations is conducted, and most of the times technical training is provided. For example, participants are endowed with practical skills of wood working and processing at work benches and machines under the supervision of instructors.

Figure 5 shows that type 1 is by far the most important program type. In the pre-reform period, about 60 percent of all participants in public training programs were assigned to this particular type. It became even more important after the reform in 2003 as this share increased to more than 70 percent. Moreover, the three types together account for roughly 85–90 percent of all participants in public training programs over the period from 2000 to 2004.

Figure 5 about here

As Figure 6 indicates, the program types under consideration are—in comparison to other ALMP measures in Germany—rather shorter measures. Both in the pre-reform and in the post-reform period, after one year more than 90 percent of the participants have left each type. However, the program duration decreased after the reform was introduced as far as type 1 is concerned. While the median program duration is about 8 months for this program type in the pre-reform period, it amounts to about 6 months after the reform. The median duration for types 2 and 3 deceased only slightly and remained rather stable at about 6 months.

Figure 6 about here

Our sample of participants consists of 275 unemployed persons per quarter and program type for the year 2003, i.e., we observe around 1,100 participants for each program type after the introduction of the training vouchers. In order to apply the matching approach as described below (see Section 4), 20 participants from the

period before the reform were drawn per participant in 2003, i.e., we draw a sample of 22,000 participants in 2002 per program type.

Beyond the matching of post-reform participants with pre-reform participants, we need to match participants with non-participants. Therefore, we also draw 80 non-participants per participant. These individuals had to be unemployed for the same duration as the corresponding participants. Moreover, they are required to not having participated in the respective type of training before and in the quarter of the participant's program entry, but we do not condition on future non-participation.[7]

In what follows, the success of program participation is evaluated by looking at the probability of being employed starting at the (fictitious) program entry 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 June 30, 2005. 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 our outcome measure. Moreover, the administrative data set only includes employment that is subject to social security contributions.[8]

4 Analytical Framework

The aim of this paper is to estimate the effect of vouchers for publicly financed training programs, and to test the hypothesis that their introduction improves the quality (i.e., the effectiveness) of the programs under consideration. However, it is difficult to isolate the impact of vouchers since along with the introduction of the voucher system, there are other major changes going on as described above—especially a stricter selection rule for participants. The reform effect could thus result from a change of program quality and from a change of the composition of

[7]In contrast, e.g., Lechner and Wunsch (2008) require non-participation in the follow-up period after the (fictitious) program entry for comparison individuals. Although we opted for the above stated definition of non-participation and do not exclude future participants, the alternative approach clearly has the advantage of employing a very straightforward definition of non-participation.

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

participants.[9]

4.1 Regression Analysis

A simple model to capture these two aspects and to isolate the voucher effect is the following:

$$Y_i = \alpha + X_i\beta + \delta D_i + \gamma R_i + \eta(D_i \times R_i) + \epsilon_i \quad (1)$$

where D_i is a dummy variable that takes the value one if individual i participates in the program and zero otherwise. R_i indicates the pre- and post-reform period similarly. Assuming that *a*) there is no selection on unobservables, *b*) the treatment effect is homogeneous, and *c*) the outcome and covariates have a linear relationship, the coefficient η in equation (1) represents the impact of the voucher. In this case, controlling for X_i is equivalent to controlling for (changes in) the composition of training participants before and after the reform.[10]

4.2 Two-Step Matching Approach

However, in order to avoid parametric assumptions and to accommodate heterogeneity of treatment effects, we also apply propensity score matching procedures to isolate the impact of vouchers. Mueser et al. (2007) present evidence that if administrative data is used to measure the performance of training programs, propensity score matching is generally most effective.

Using the potential outcome framework as in 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 again an indicator for participation, we can define different treatment effects in a similar way as Heckman and

[9]Changes in the general economic situation may be another component of the reform effect. Although we generally argue that we control for such changes as participants and matched non-participants are subject to the same cyclical environment, we will explicitly address this issue in Section 5.3.

[10]We run this regression on the raw sample of participants and non-participants separately for each program type twice: *a*) *without* controlling for individual characteristics X_i, and *b*) controlling for individual characteristics X_i. The results of these regression are discussed in Section 5.1 below.

Vytlacil (1999, 2005):

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

and the average treatment effects on the treated before and after the reform are

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

Under the matching assumptions of Rosenbaum and Rubin (1983)

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

ATT_b (or ATT_a) can be estimated from pre-reform data (or post-reform data) by propensity score matching methods and it corresponds to δ (or $\delta + \eta$) in equation (1).[11]

However, the difference between ATT_b and ATT_a 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).

Hence,

$$ATT_a = ATT_b + VE + SE \tag{5}$$

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

$$\begin{aligned} RE &= ATT_a - ATT_b \\ &= VE + SE \end{aligned} \tag{6}$$

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

To be able 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 is the effect *only* for those participants under the pre-reform regime who are comparable to participants after the reform (we henceforth refer to this effect as ATT_{bQ}), so this step controls for the selection effect.

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

$$VE = ATT_a - ATT_{bQ} \tag{7}$$

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

$$\begin{aligned} SE &= RE - VE \\ &= (ATT_a - ATT_b) - (ATT_a - ATT_{bQ}) \\ &= ATT_{bQ} - ATT_b \end{aligned} \tag{8}$$

5 Results

In this section, we report the results based on the two approaches outlined above: *a*) regression analysis, and *b*) two-step matching procedure. Estimates of the differences in employment probabilities are presented for a period of two years after the (fictitious) program entry.[12]

[12]We thus follow the prevailing approach in the recent evaluation literature. A different approach concentrates on treatment effects only after the end of the program. For advantages and disadvantages of both approaches see, e.g., Caliendo and Kopeinig (2008).

5.1 Regression Analysis

We run standard regressions according to equation (1) on the raw sample of participants and non-participants separately for each program type twice: *a*) *without* controlling for individual characteristics X_i, and *b*) controlling for individual characteristics X_i. We present estimates on the probability of being employed 6, 12, and 18 months after the program entry, respectively. The results of these regressions are displayed in Table 1.

Table 1 about here

As mentioned above, our primary interest lies on the estimate for the coefficient η which represents the impact of the voucher. For all program types being analyzed, we consistently find significantly positive voucher effects on the outcome variable 6 and 12 months after the program entry, respectively. Moreover, these results are quite robust to controlling for individual characteristics X_i.[13] However, the voucher effects turn out to be in general significantly negative 18 months after program entry, respectively.

5.2 Two-Step Matching Approach

By applying the described two-step matching procedure, we are able to decompose the reform effect into two separate components: *a*) the selection effect, and *b*) the voucher effect. For this purpose, the treatment effects for the respective matched samples are calculated as the difference in mean outcomes between the matched groups. Below, we present estimates of differences in employment probabilities for a period of 18 month after program entry, calculated every fortnight.[14]

Firstly, we display estimates of the reform effect in Figure 7. The underlying ATT_a and ATT_b are based on differences in mean outcomes between program entrants in 2002 for the pre-reform period (in 2003 for the post-reform period) and

[13] Type 2 constitutes an exception in this context. After controlling for for individual characteristics X_i, the voucher effect becomes negative and/or insignificant.

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

matched non-participants *without* controlling for potential changes in the composition of participants between the two periods. The reform effect is then simply the difference between the two depicted lines.

Figure 7 about here

We observe that participants in all analyzed program types face a substantial lock-in effect.[15] In the first months after entering the program, the employment probabilities of participants are considerably lower than those of matched non-participants. The duration and the extent of these lock-in effects vary by program type. While we observe similar lock-in effects for program type 1 and 3 for which treatment effects become positive about one year after program entry, the lock-in effect for program type 2 is comparatively moderate and of shorter duration. Treatment effects become positive already about 8 months after program entry.

The comparison of the estimates between the pre- and the post-reform period in Figure 7 reveals—if anything—positive reform effects for the program types being analyzed. In particular for program type 3 the post-reform treatment effect lies consistently above our pre-reform estimates. The magnitude of this difference is about 5 percentage points for essentially the whole observation period. In case of program type 1, we observe a positive reform of a similar magnitude between 6 and 12 months after program entry. Afterwards, this difference decreases; and 18 months after program entry the estimated treatment effects are virtually the same for the pre- and post reform period. Finally, also type 2 shows a positive reform effect, but only at the end of the observation period. Up to about 14–15 months after program entry we do not observe any sizeable differences between ATT_a and ATT_b.

In Figure 8 we present estimates of the voucher effect. The underlying ATT_a and ATT_{bQ} are based on differences in mean outcomes between *matched* program entrants in 2002 for the pre-reform period or program entrants in 2003 for the post-reform period and matched non-participants, respectively. By doing so, we take potential changes in the composition of participants between the two periods into account. The voucher effect is then the difference between the two depicted lines.

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

Figure 8 about here

Although Figure 8 generally looks quite similar to Figure 7, some differences can be observed. Regarding program type 1, we observe a consistent and comparatively large difference between ATT_b and ATT_{bQ} starting about 6 months after program entry. This points to a voucher effect which is more positive than the reform effect discussed above. The difference between pre- and post reform treatment effects also turns out to increase for type 2 once we take potential changes in the composition of participants between the two periods into account. This is particularly the case at the end of the observation period, i.e., one year after program entry and afterwards. On the other hand, our estimates for program type 3 do not change much compared to the previous results.

Figure 9 reveals more insights about the extent and magnitude of reform effects, voucher effects, and selection effects. It displays the decomposition of the reform effect, and thus summarizes the graphs discussed so far.

Figure 9 about here

The decomposition shows that for all program types being analyzed that the reform effect seems to be almost exclusively based on the voucher effect. The most striking picture emerges with respect to program type 1. For this type, starting about 6 months after program entry the reform effect and the voucher effect become substantially positive. Moreover, the voucher effect is about twice as positive than the reform effect; and while the former lies constantly at about 10 percentage points or above, the latter decreases from about 5 percentage points between 6 and 12 months after program entry to almost 0 percentage points 18 months after program entry. This can be explained with the selection effect being the residual effect: it constantly decreases throughout the observation period and becomes more and more negative. At the end of the observation period, it amounts to less than –10 percentage points. As far as types 2 and 3 are concerned, we observe selection effects that lie around 0 percentage points throughout the observation period. While the decomposition for type 3 shows virtually no difference between the voucher effect and the reform effect, the voucher effect is clearly the driving force behind the

positive reform effect which can be observed for type 2 at the end of the observation period.

The results of the two-step matching procedure thus indicate that the selection effect plays virtually no role in explaining the overall positive impacts of the reform in 2003—in particular as far as the most important program type is concerned, the selection effect is clearly negative and the overall reform effect would have been more positive if the composition of participants would not have changed. In this regard, 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.

5.3 Sensitivity Analysis

We address the robustness of our previous results in this section. For this purpose, we perform a sensitivity analysis in three steps. We assess the robustness of our results with respect to *a*) changes in the general economic situation, *b*) the transitional arrangement for training voucher in the beginning of 2003, and *c*) changes in program duration.

General Economic Situation

One may argue that changes in the general economic situation constitute another component of the reform effect. Although we think that we control for such changes above as participants and matched non-participants are subject to the same cyclical environment, we explicitly address this issue. 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. Therefore, we additionally control for a number of economic and labor market characteristics which are available for each local FEA district. More precisely, we include monthly information on the

share of unemployed, the share of vacancies, the share of participants in various ALMP measure (including public training programs) as well as GDP growth rates.[16] Figures 10–12 display the results of this procedure.

Figures 10–12 about here

Comparing these results with our previous results without controlling for economic and labor market characteristics reveals that there are only marginal differences. Importantly, the conclusions that can be drawn do not change at all.

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.[17] The results of this analysis are depicted in Figures 13–15.

Figures 13–15 about here

Again, no substantial changes in our results can be observed. More specifically, for types 2 and 3 virtually no changes appear. But with respect to program type 1, the reform effect turns out to be slightly more positive than before. The decomposition then reveals that this increase is on the one hand driven by a more positive voucher effect at the at of the observation period, and on the other hand it is driven by a less negative selection effect about 7–8 months after program entry.

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

[17] 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.

Program Duration

As discussed in Section 3 and displayed in Figure 6, the program duration for types 2 and 3 remains rather stable between the pre- and post-reform period. However, the program duration of program type 1 decreases after the reform. We thus additionally estimate for this program type treatment effects after the (fictitious) program exit which are displayed in Figures 16–18. For obvious reasons the observation period reduces to one year.

Figures 16–18 about here

Basically the estimated treatment effects for program type 1 after the (fictitious) program exit confirm the picture that has been sketched above. We still observe a positive voucher effect which lies around 10 percentage point and a negative selection effect of slightly lower magnitude. Therefore, the reform effect is in general positive throughout the observation period. But it also exhibits an inverse U-shaped pattern: While it is negative directly after the (fictitious) program exit, it increases afterwards, remains rather stable around 2–3 percentage points, and finally it becomes slightly negative again one year after program exit.

6 Conclusion

This paper analyzes the impacts of the labor market reform in 2003 on the effectiveness of publicly financed training in Germany. Using a rich administrative data set from the FEA, we apply regression and propensity score matching procedures to address the question whether the introduction of vouchers raised the quality of the programs by means of an increased competition. Therefore, we decompose the total reform effect into two separate components. By distinguishing between a voucher effect and a selection effect, we disentangle the effects that are based on the introduction of vouchers from those that are based on changes in the composition of program participants.

Regression analysis reveals generally positive voucher effects on the outcome variable 6 and 12 months after program entry. However, voucher effects turn out

to be in general negative 18 months after program entry, respectively. Our results obtained from the two-step matching procedure indicate that the selection effect plays only a minor role in explaining the generally positive impacts of the Hartz reform. More precisely, our decomposition of the reform effect reveals that this effect is almost exclusively based on the voucher effect. The most striking picture emerges with respect to the most important program in terms of participants.

In sum, based on the results of the two-step matching procedure we conclude that the selection effect plays—at the best—virtually no role in explaining the overall positive impacts of the reform in 2003. More specifically, the selection effect is clearly negative as far as the most important program type is concerned. The overall reform effect for this type would thus have been more positive if the composition of participants had not changed after the reform.

References

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.

Baumol, W. (1982). Contestable Markets: An Uprising in the Theory of Industry Structure. *American Economic Review 72*(1), 1–15.

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*.

Brunello, G. and D. Checchi (2004). School Vouchers Italian Style. *Giornale degli Economisti e Annali di Economia 63*(3/4), 357–399.

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. *Zeitschrift für ArbeitsmarktForschung 38*(2-3), 396–418.

Dohmen, D. (2003). Alternative Formen der Finanzierung beruflicher Weiterbildung. *FiBS-Forum 15*.

Dohmen, D. (2007). Current Trends in the of Demand-led Financing of Further Training in Europe: A Synopsis. *FiBS-Forum 40*.

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 (2006). Get Training or Wait? Long-Run Employment Effects of Training Programs for the Unemployed in West Germany. *IZA Discussion Paper 2121*.

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

Geers, F. (2001). Belgium, Wallonia Region: The Training Voucher Scheme. In T. Pukkinen, C. Romijn, and S. Elson-Rogers (Eds.), *Funding Continuing Traning in Small and Medium-Sized Enterprises. Discussion and Case Studies from Across the EU*, Volume 17 of *Cedefop Panorama Series*. Luxemburg.

Hallberg, K. (2006). A Retrospective Assessment of the Kenya Voucher Training Program. *Small Enterprise Development 17*(2), 56–67.

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., 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.

Hujer, R. and M. Wellner (2000). Berufliche Weiterbildung und individuelle Arbeitslosigkeitsdauer in West- und Ostdeutschland: Eine mikroökonometrische Analyse. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 33*(3), 405–419.

Hummel, E., P. Jacobebbinghaus, A. Kohlmann, M. Oertel, C. Wübbeke, and M. Ziegerer (2005). Stichprobe der Integrierten Erwerbsbiographien, IEBS 1.0. *FDZ Datenreport 6*.

Kluve, J. (2006). The Effectiveness of European Active Labor Market Policy. *IZA Discussion Paper 2018*.

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*.

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

Lechner, M. (1999). Earnings and Employment Effects of Continuous Off-the-Job Training in East-Germany after Unification. *Journal of Business & Economic Statistics 17*(1), 74–90.

Lechner, M. (2000). An Evaluation of Public-Sector-Sponsored Continuous Vocational Training Programs in East Germany. *The Journal of Human Resources 35*(2), 347–375.

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

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*.

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.

Levin, H. M. (2003). Post-Compulsory Entitlements: Vouchers for Lifelong Learning. In D. Dohmen and B. A. Cleuvers (Eds.), *Finanzierung von Weiterbildung und lebenslangem Lernen*, Volume 2 of *Schriften zur Bildungs- und Sozialökonomie*. Bielefeld.

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

Martin, J. P. and D. Grubb (2001). What Works and for Whom: A Review of OECD Countries' Experiences with Active Labor Market Policies. *Swedish Economic Policy Review 8*(2), 9–56.

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.

Oosterbeek, H. (1998). Innovative Ways to Finance Education and Their Relation to Lifelong Learning. *Education Economics 6*(3), 219–251.

Rinne, U., M. Schneider, and A. Uhlendorff (2007). Too Bad to Benefit? Effect Heterogeneity of Public Training Programs. *IZA Discussion Paper 3240*.

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.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized 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*.

Schneider, H. and A. Uhlendorff (2006). Die Wirkung der Hartz-Reform im Bereich der beruflichen Weiterbildung. *Zeitschrift für ArbeitsmarktForschung 39*(3-4), 477–490.

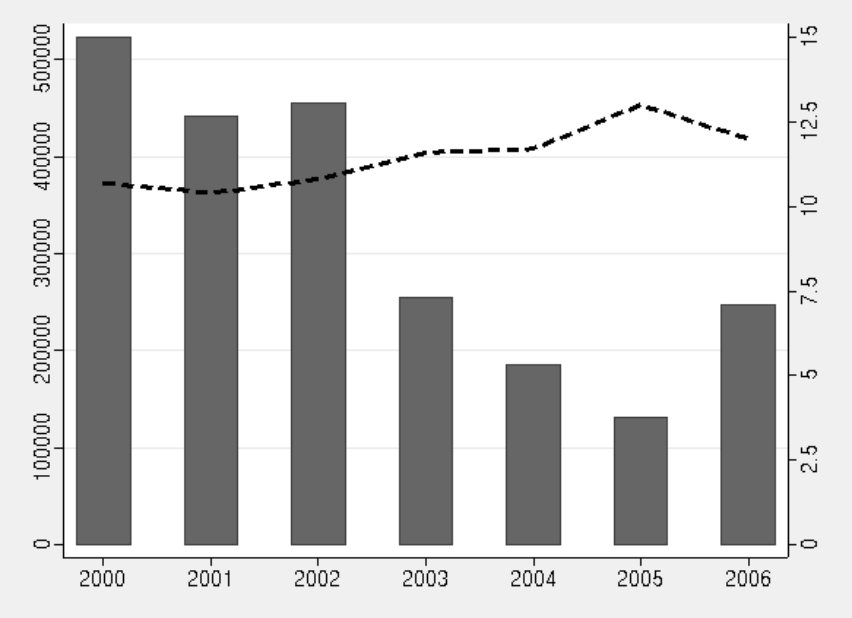
Stephan, G. (2008). The Effects of Active Labor Market Programs in Germany: An Investigation Using Different Definitions of Non-treatment. *IAB Discussion Paper 12*.

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*.

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.

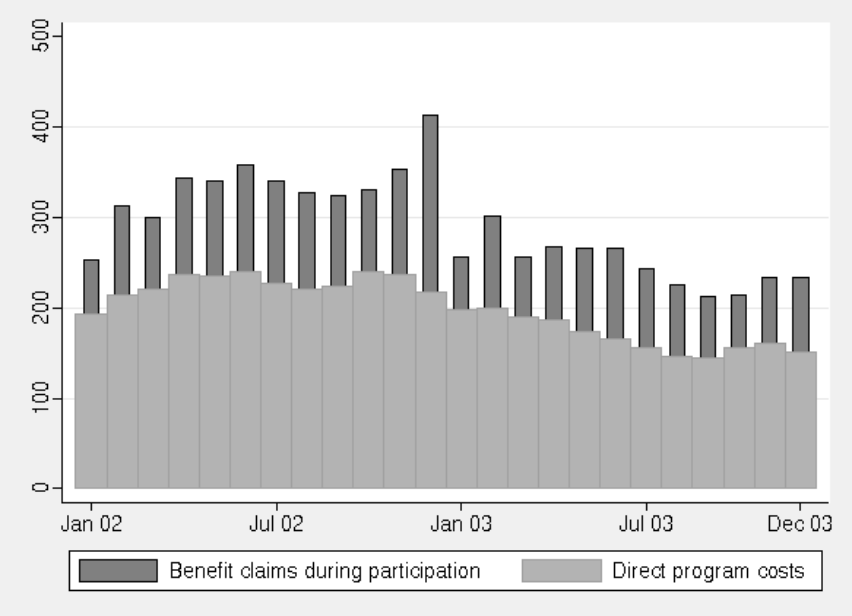
Figure 1: Entrants into public training programs and unemployment rate (2000–2006)



Source: Federal Employment Agency (FEA).

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).

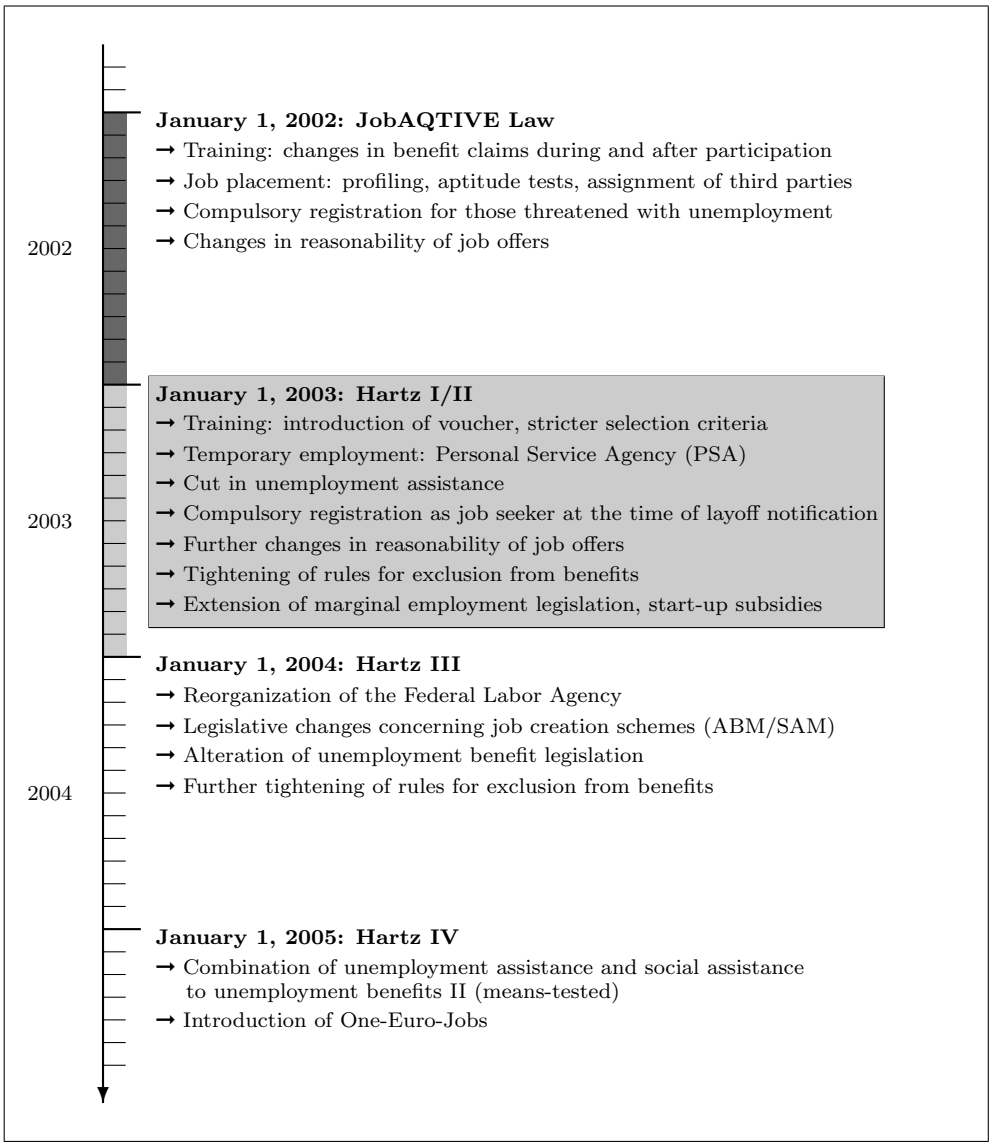
Figure 2: Costs of public training programs (2002/2003)



Source: Federal Employment Agency (FEA).

Note: In million Euro.

Figure 3: Chronology of the Hartz reforms



Source: Authors' illustration.

Figure 4: 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)

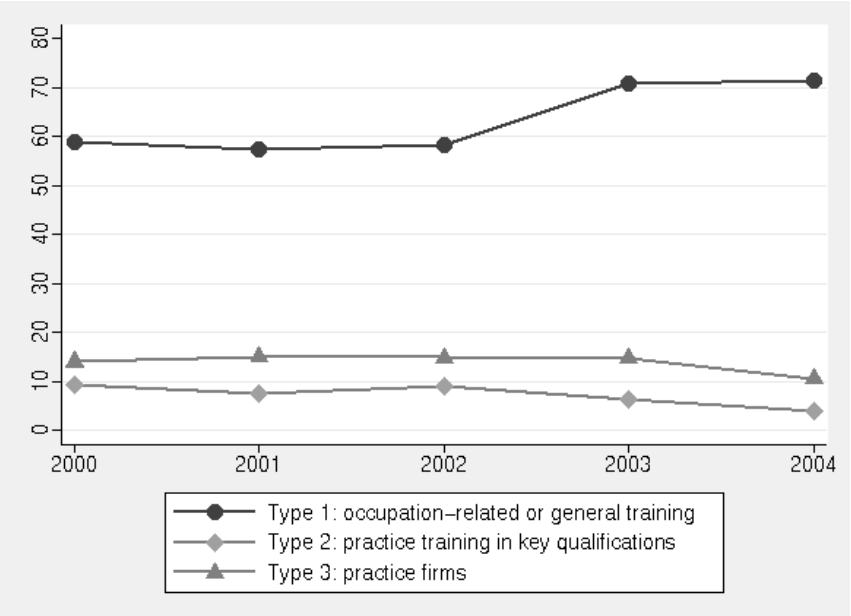
Educational target / qualifications provided:

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

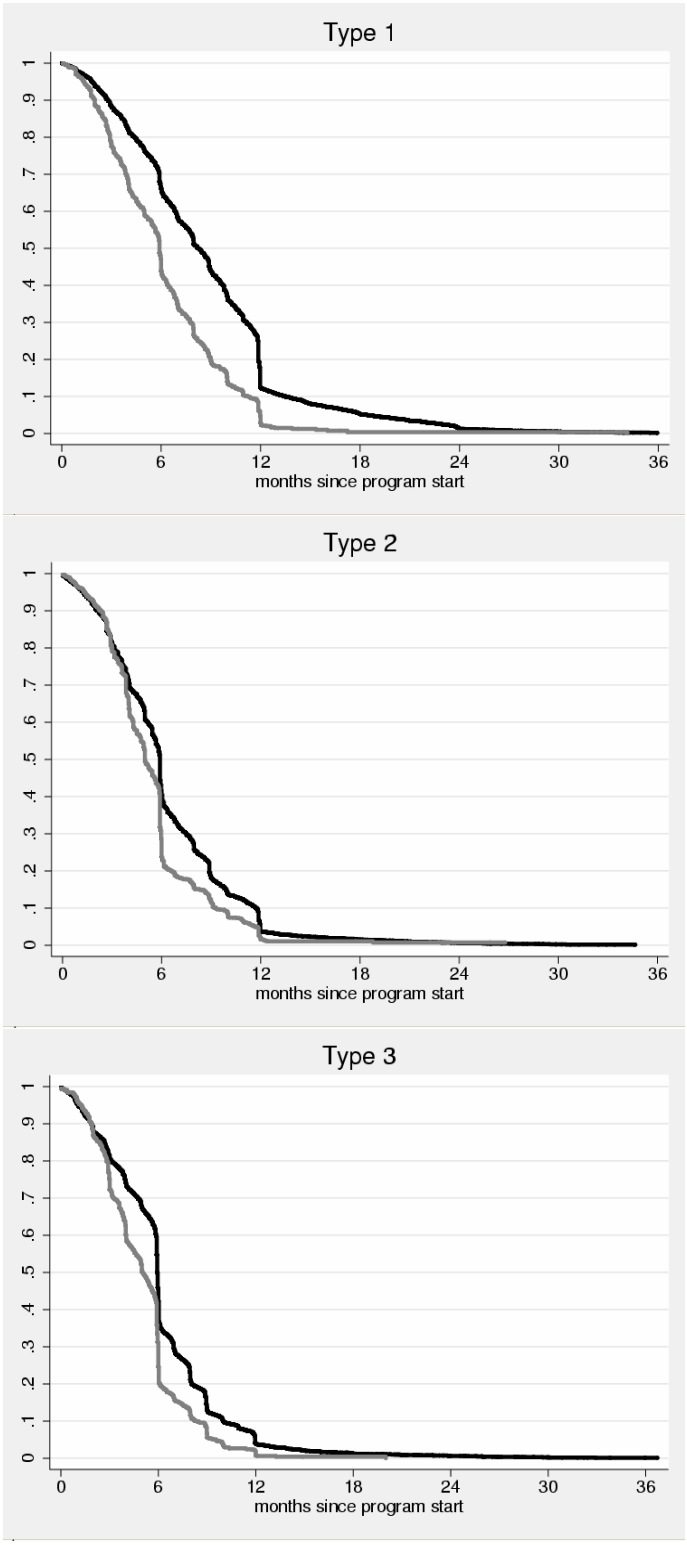
Figure 5: Share of entrants by program type (2000–2004)



Source: Federal Employment Agency (FEA).

Note: Share (in percent) in annual number of entrants in publicly financed training programs.

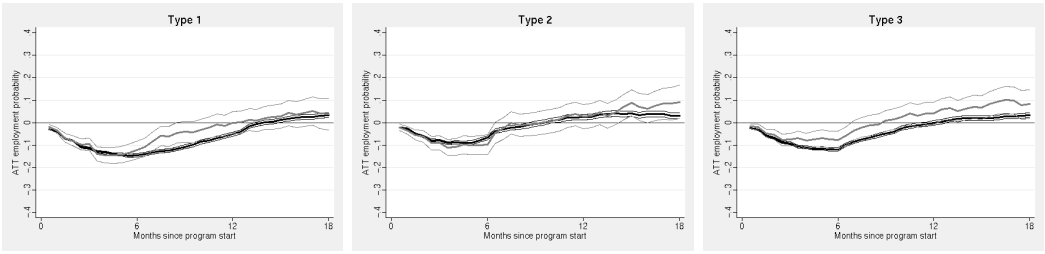
Figure 6: Actual program duration by program type



Source: IEB, own calculations.

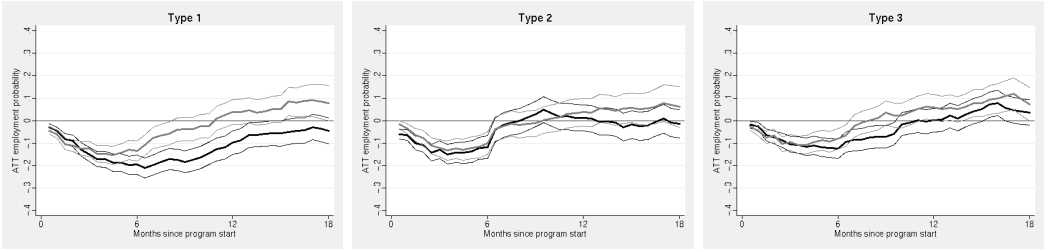
Note: Kaplan–Meier Estimates. Pre-reform period in black, post-reform period in gray.

Figure 7: Reform effect



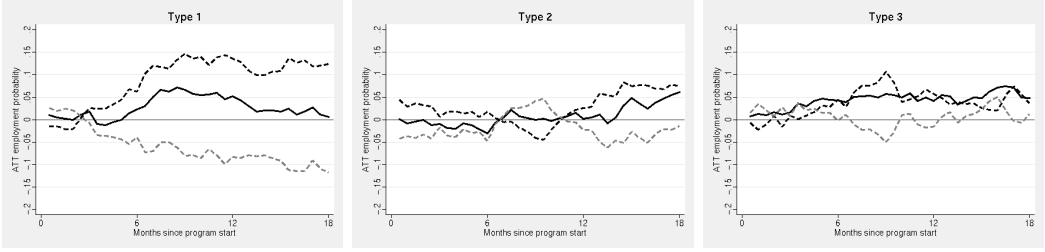
Source: IEB, own calculations.
Note: Pre-reform period in black, post-reform period in gray.
Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure 8: Voucher effect



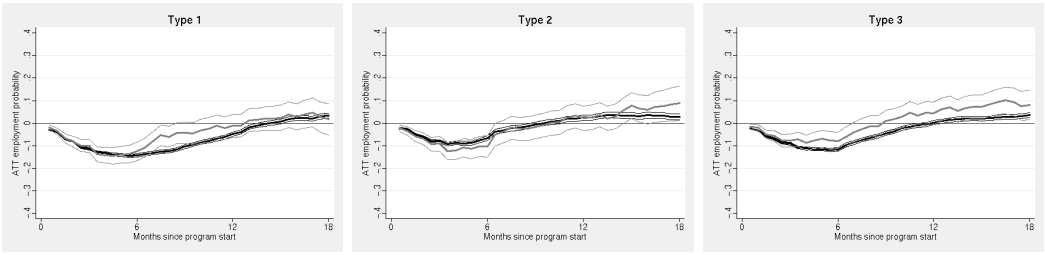
Source: IEB, own calculations.
Note: Pre-reform period in black, post-reform period in gray.
Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

Figure 9: Decomposition (reform effect, voucher effect, and selection effect)



Source: IEB, own calculations.
Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed).

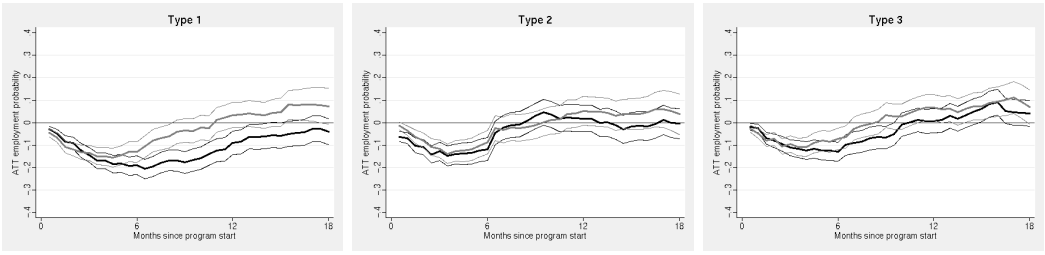
Figure 10: Reform effect *controlling for macro conditions*



Source: IEB, own calculations.
Note: Pre-reform period in black, post-reform period in gray.

Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

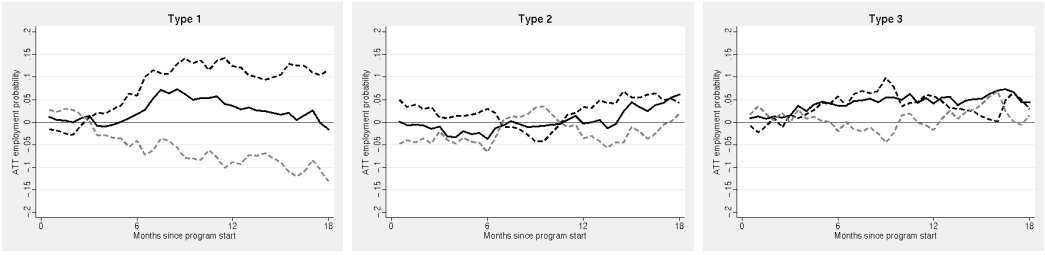
Figure 11: Voucher effect *controlling for macro conditions*



Source: IEB, own calculations.
Note: Pre-reform period in black, post-reform period in gray.

Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

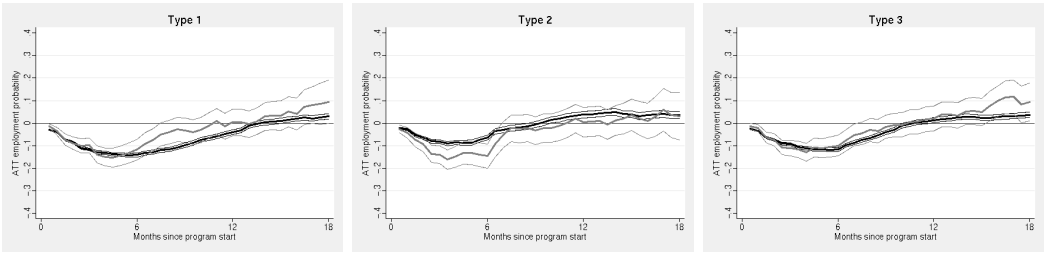
Figure 12: Decomposition (reform effect, voucher effect, and selection effect) *controlling for macro conditions*



Source: IEB, own calculations.

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed).

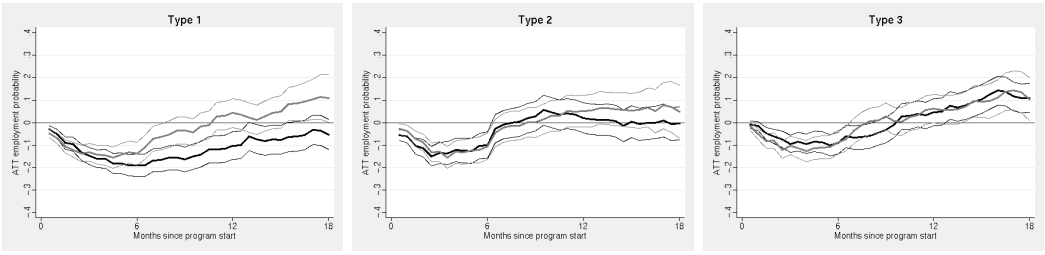
Figure 13: Reform effect *excluding first quarter 2003*



Source: IEB, own calculations.
Note: Pre-reform period in black, post-reform period in gray.

Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

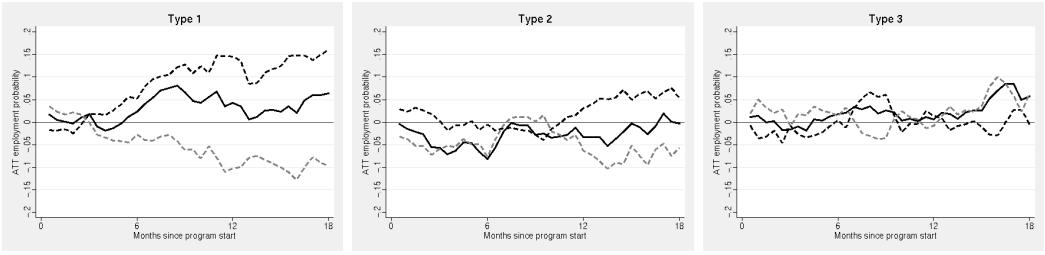
Figure 14: Voucher effect *excluding first quarter 2003*



Source: IEB, own calculations.
Note: Pre-reform period in black, post-reform period in gray.

Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

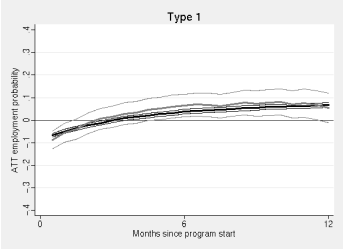
Figure 15: Decomposition (reform effect, voucher effect, and selection effect) *excluding first quarter 2003*



Source: IEB, own calculations.

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed).

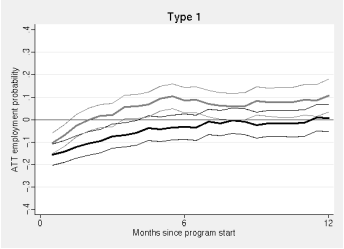
Figure 16: Reform effect *after program exit*



Source: IEB, own calculations.
Note: Pre-reform period in black, post-reform period in gray.

Thick lines refer to point estimates, thin lines indicate 95 percent confidence intervals.

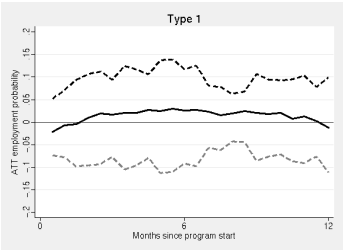
Figure 17: Voucher effect *after program exit*



Source: IEB, own calculations.
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 (reform effect, voucher effect, and selection effect) *after program exit*



Source: IEB, own calculations.

Note: Total reform effect (RE) in black (solid), voucher effect (VE) in black (dashed), and selection effect (SE) in gray (dashed).

Table 1: Regression analysis (OLS)

	Employment after 6 Months		Employment after 12 Months		Employment after 18 Months	
Type 1						
Participant	-.066***	-.108***	0.057***	-.0004	0.13***	0.079***
After	0.024***	0.014***	0.021***	0.009***	-.061***	-.078***
Participant × After	0.036***	0.039***	0.059***	0.058***	-.131***	-.121***
Vector X_i included	No	Yes	No	Yes	No	Yes
Obs.	993572	993572	993572	993572	993572	993572
R^2	0.002	0.109	0.001	0.104	0.009	0.132
Type 2						
Participant	-.002	-.022***	0.066***	0.048***	0.077***	0.043***
After	0.024***	0.013***	0.021***	0.009***	-.061***	-.078***
Participant × After	0.008	-.016	0.054***	0.015	-.100***	-.100***
Vector X_i included	No	Yes	No	Yes	No	Yes
Obs.	983924	983924	983924	983924	983924	983924
R^2	0.001	0.109	0.001	0.105	0.006	0.131
Type 3						
Participant	-.052***	-.086***	0.072***	0.031***	0.098***	0.05***
After	0.024***	0.014***	0.021***	0.009***	-.061***	-.078***
Participant × After	0.031***	0.024**	0.055***	0.037***	-.079***	-.078***
Vector X_i included	No	Yes	No	Yes	No	Yes
Obs.	990314	990314	990314	990314	990314	990314
R^2	0.001	0.109	0.001	0.105	0.007	0.132

Source: IEB, own calculations.
Note: *** significant at 1%, ** significant at 5%; * significant at 10%.
Vector X_i contains similar variables as the matching specifications.