End-of-Year Spending and the Long-Run Effects of Training Programs for the Unemployed

Bernd Fitzenberger*, Marina Furdas**,
Olga Orlanski**, Christoph Sajons**

This version: February 2015

Abstract: This study re-estimates the employment effects of training programs for the unemployed in West Germany using exogenous variation in participation caused by budget rules in local employment offices. As funds could not be transferred to the next year or to other programs, a budget surplus (deficit) after the first half of the year increases (decreases) training participation at the end of the fiscal year ("end-of-year spending") irrespective of the situation of the unemployed. For all programs affected by this budget mechanism, our instrumental variables estimates imply sizeable lock-in effects on employment in the short run, but heterogeneous effects later-on. While retraining and longer programs providing specific professional skills and techniques improve employment in the medium and long run, our findings imply long-lasting negative effects of practice firms.

Keywords: Training for the unemployed, budgetary conditions, administrative data, Germany

JEL-Classification: J64, J68, H43

---

* University of Freiburg (Germany), IFS, IZA, ROA, and ZEW. E-mail: bernd.fitzenberger@vwl.uni-freiburg.de (corresponding author).
** University of Freiburg (Germany).

We are very grateful to Stefan Bender, Karsten Bunk, Theresia Denzer-Urschel, Bärbel Höltzen-Schoh, Else Moser, and Georg Uhlenbrock for helpful information and to participants at the Café Workshop 2014 in Børkop, Denmark for helpful comments. This paper is part of the project "Policy change, effect heterogeneity, and the long-run employment impacts of further training programs" ("Politikänderung, Effektheterogenität und die längerfristigen Beschäftigungswirkungen von Fortbildung und Umschulung", IAB project number: 1213–10–38009). Financial support by the IAB is gratefully acknowledged. The responsibility for all errors is, of course, ours.

1 Introduction

The effectiveness of training programs for the unemployed has been a debated issue (Card et al., 2010). Proponents argue that training programs are a good investment, since enhancing the abilities and skills of unemployed individuals would lead to a quicker reintegration in the labor market and a win-win situation for the government, the companies, and the unemployed. Critics claim that the resources spent on training programs are mostly wasted, arguing that the programs themselves would not have any positive impact on later employment and that good actual outcomes may only reflect a positive selection of training participants (e.g., Wunsch and Lechner, 2008). Any empirical examination of this controversy gets complicated by the possibility that the assignment to a training program depends upon characteristics of the unemployed which are not observed by the researcher but which influence the individual's employment chances (selection on unobservables). In particular, caseworkers decide both whether to send an individual to a training program at all and if so, to which type of program, ranging from short-term training to two to three year long retraining. If either decision depends upon the caseworker's assessment of the chances of this person to find a job, which is typically not observed in the data, then even a sophisticated regression or matching analysis may lead to overly optimistic results. However, causal effect estimates accounting for selection on unobservables is scarce (Card et al., 2010).[1]

This paper estimates the causal effects of training exploiting the spending incentives created by strict budget rules within the Federal Employment Office (FEO) in West Germany during the 1980s and early 1990s which caused an exogenous variation in program entries. In particular, regional employment offices were subject to two rules: First, their annual budgets were determined primarily based on their spending needs in the previous year. And second, funds allocated to training programs in one year could not be transferred in any way, i.e., neither to finance training programs in the following year nor towards other purposes like job creation schemes in the same year. Combined, these budget rules created a strong incentive for employment offices to use their whole budget for training programs, thus assigning more unemployed to training in years with a budget

[1]Notable exceptions are analyses based on the timing-of-events approach of Abbring and van den Berg (2004), see e.g. Osikominu (2013). This approach assumes time-invariant random effects independent of the covariates which govern the selection process. Conditional on these random effects, program participation at any point in time is random and only affects exits from unemployment in the future. The approach requires data in continuous time, however, which is mostly not available. Our paper therefore focuses on exogenous variation to identify the effect of training in discrete time.

surplus early on than they otherwise would have done. Thus, we use the variation in the need to spend remaining funds in the final months of the year to instrument the individual participation in a training program. Doing so allows us to both reduce the potential influence of selection of participants and program and come closer towards estimating the causal effect of further training on the future employment prospects of participants.

For the empirical analysis, we implement our instrumental variables strategy using the control function method for a random-coefficients model proposed by Wooldridge (2014), which allows us to account in a flexible way for the discrete nature of the outcome variables. Our instrumental variation on program participation proves highly significant for a number of important programs. Our main source of data source involves administrative records from the German Federal Employment Office containing the complete employment information up until 2004 for a random sample of 50% of all participants in training for the unemployed in Germany between 1980 and 1993, as well as a 3% random sample of all unemployed who did not participate in any such program over the same period. Combined with information about actual and planned spending from the annual books of the FEO, our rich data allow us to both calculate the size of first-semester deficits or surpluses for each regional employment office (REO) and follow individual employment histories over up to nine years.

Our findings imply that a budget surplus in a regional employment office in the first half of the year results in an increase in the probability of an unemployed individual to enter a long training program or a practice firm after the summer holidays. The reverse holds for people living in regions in which the employment office runs a deficit in the first half of the year. This provides plausible evidence that spending the entire assigned budget was a goal of the management of the employment offices. With respect to the long-run impact of training programs on the employability of participants, we find heterogeneous effects. On the one hand, long programs with a strong focus on acquiring new skills seem to increase the chances of their participants to find a job afterwards. On the other hand, there are negative effects of shorter programs (practice firms) in which the unemployed work in simulated firms in order to maintain their general work skills. These findings show the importance of accounting for differences in the effects of different training programs. It is not warranted to pool all training programs together.

Our paper proceeds as follows: In section 2, we provide detailed information on the conduct and institutional setup of active employment promotion in Germany during the time of our study and relate our work to existing literature on the effectiveness of further training programs. Section 3 describes the data. In section 4, we present our instrumental variable, the accumulated budget surplus in the first half of each fiscal year, and report results on the plausibility of our identification strategy. The econometric approach is presented and described in section 5. Section 6 reports and discusses the empirical results on the effects of training on subsequent employment. Section 7 concludes.

2 Training programs for the unemployed in Germany

2.1 Background and institutions

The conduct of training programs for the unemployed has a long history in Germany, dating back to the enactment of the Employment Promotion Act (EPA, *Arbeitsförderungsgesetz*) in 1969. This legislation introduced a variety of instruments of active labor market policy (ALMP), with further training programs (*Fortbildung und Umschulung*) as the most important component at that time.[2] These programs vary strongly with respect to the intended aim of qualification and their duration, ranging from only few weeks for short-term training to a maximum of three years for complete retraining programs. Their overarching goal, however, is the same: to provide general or specific occupational skills in order to improve the labor market prospects of unemployed individuals and those at risk of unemployment. To achieve this aim, the FEO provided financial support for participants which could contain both income maintenance payments and the costs of the program, including money for travel, childcare, and accommodation expenditures. The overall budget available for training programs and income maintenance for program participants totaled around 3.4 billion Deutsche Mark in the mid-1980s (close to 1.8 billion Euros), representing about 11.4% of the annual budget of the FEO at the time.[3]

We use the information available on the five main public-sponsored training programs: Practice Firms (*Übungsfirmen*, PF), Provision of Specific Professional Skills and Techniques (*Bereitstellung von spezifischen Kenntnissen und beruflichen Fähigkeiten*, SPST), short-term training programs according to §41a of the EPA (*Kurzzeitmaßnahmen*, STT), Retraining (*Umschulung*, RT), and Wage Subsidies (*Einarbeitungszuschüsse*, WS). For their construction, we follow the classification developed in Fitzenberger and Speckesser

[2] Other major labor market policy instruments are employment creation schemes (*Arbeitsbeschaffungsmaßnahmen*), promotion of vocational training (*Förderung der beruflichen Ausbildung*), occupational rehabilitation (*Berufliche Rehabilitation*), and short-time work (*Kurzarbeit*).

[3] Own calculations based on figures from the reports of the FEO (Bundesanstalt für Arbeit, 1993).

(2007). In the following, we briefly describe each of these programs in turn. A detailed description of the types of FuU training programs can be found in Bender et al. (2005) and Fitzenberger et al. (2008).

Practice Firms (PF) are simulated firms in which individuals are trained on everyday work activities, focusing either on technical or commercial tasks. The program usually lasts six months and aims at providing participants with general skills appropriate for a wide range of jobs. Additionally, PFs are used to assess a participant's ability for particular professions. Since the program concentrates on exercising existing skills rather than learning new ones, participants do not obtain official certificates.

In contrast, *Specific Professional Skills and Techniques* (SPST) programs focus on the provision of more specific skills like computer or accounting courses. A completed vocational training degree is usually required for participation in this type of further training. The courses mainly have a theoretical focus, but may also provide some practical experience. In case of successful completion, participants usually earn a certificate describing the content of the course and the newly acquired knowledge and experience. The aim of SPST is to facilitate the reintegration of unemployed individuals into the labor market by improving their skills and providing signals to potential employees. Due to the wide variety of courses offered, SPST represent the largest share within all FuU training programs.

Short-Term Training (STT) programs were created in 1979 in order to fill the gap within the already existing programs by focusing on hard-to-place and low-skilled individuals. In these short-term training courses, job seekers were informed about employment options as well as possibilities for participation in more comprehensive programs. Furthermore, STT participants were taught some limited skills on how to act on the labor market, including job counseling, application and communication training, and contact to potential employees. In general, the maximum length was six weeks and participants did not have to take an exam at the end of the course (Schneider, 1981). Due to tight budgets, STT programs existed only until the end of 1992, before they gained importance again with another program design starting in 1997 (Fitzenberger et al., 2013).

The most expensive programs of the further training programs organized by the FEO are *Retraining* (RT) courses. The difference between retraining and the programs described above is that participants actually complete a full vocational training degree (certificate). Most of its participants already hold a different vocational training degree for a specific occupation but they have low chances to find a job for they are qualified. Retraining is also an option for individuals without any vocational degree provided they meet additional eligibility criteria. Retraining combines both theoretical and practical training, with a duration of up to three years. After successful completion, participants obtain a widely accepted formal certificate, which serves as a signal for new job qualification and which aims at fostering the skill match in the labor market.

Finally, *Wage Subsidies* (WS) with initial skill adaptation training by the employer aim at the reintegration of the unemployed with placement barriers into the labor market. This program aims to overcome the competitive disadvantages of the unemployed by reducing the cost of hiring them and by preparing them for the job requirements in the long run through on-the-job training and work experience.

To qualify for the aforementioned programs, individuals have to fulfill certain requirements, e.g., having worked for at least one year prior to getting unemployed or being entitled to unemployment benefits or subsequent unemployment assistance (Fitzenberger and Völter, 2007; Lechner et al., 2011). In the case of full-time enrollment, participants receive income maintenance payments throughout the duration of their training. The only exception are wage subsidies, where the employers pay participants the usual wage rate, for which they are partially compensated through the subsidy.

2.2 Related literature

Given the costs and visibility of training programs for the unemployed, it is not surprising that many researchers have tried to examine their effects on the later employment of participants. In a meta-analysis of 97 international studies conducted between 1995 and 2007, Card et al. (2010) report that training programs seem to be ineffective in the short-run, but tend to have positive medium-run effects (corroborating evidence for West Germany up to the 1990s can be found in Fitzenberger et al. (2008) and Lechner et al. (2011)). However, the majority of studies for Germany typically rely on selection on observables assumptions for identification and use flexible matching methods for estimation in either a static or dynamic evaluation setting. Unlike the static evaluation approach, the dynamic treatment assignment accounts for the fact that training might start at different points of time during unemployment, so that individuals might belong to the treated or non-treated group based on their elapsed unemployment duration.[4]

[4]See Sianesi (2004) and Fredriksson and Johansson (2008) for a comprehensive discussion of the problems arising from a static approach in the evaluation of training programs for unemployed individuals.

Following Sianesi (2004, 2008), Fitzenberger et al. (2008) estimate the long-run employment effects of further training programs in West Germany in a dynamic context conditional on the starting date of the treatment (treatment vs. waiting). Based on administrative data from the FEO for two unemployment inflow samples, 1986/87 and 1993/94, their results confirm both the negative lock-in effect after program start and significantly positive employment effects in the middle- and long-run. Using the same evaluation approach, Fitzenberger et al. (2013) analyze the effectiveness of short-term training in two different time periods, 1980-1992 and 2000-2003, and find positive and often significant employment effects. These findings vary somewhat by the particular type of short-term training, as programs which focus on testing and monitoring search efforts seem to cause smaller effects compared to those concentrating completely on labor market related training. Biewen et al. (2014) evaluate and compare the employment effects of different further training programs in the early 2000s in West Germany using rich administrative data on individual employment histories from the FEO (the study does not analyze retraining). The results suggest that the employment effects of short-term training are positive and of a similar magnitude as those of medium-term training, but the positive effects of short-term training appear much earlier because of its shorter duration. Furthermore, the authors emphasize the importance of methodological and data issues in the evaluation of public-sponsored training programs and provide a comprehensive sensitivity analysis of their empirical results.

Using the same data source as in Fitzenberger et al. (2008), Lechner et al. (2011) evaluate training programs starting from January 1992 to June 1993 in West Germany. By contrast, Lechner et al. (2011) do not estimate the effect of training versus waiting, but the effect of participation versus non-participation in training employing hypothetical starting dates for the non-treated individuals. The results are, however, similar, suggesting negative employment effects in the short and positive employment effects in the long run.[5] Among the examined programs, retraining exhibits the largest effect on later employment with about 20 percentage points after eight years, whereas other training programs seem to cause smaller gains of about 10 percentage points.

An exception to the aforementioned findings of positive long-run effects is the study by Wunsch and Lechner (2008), in which the authors examine several training and employment programs from 2000 to 2002. Applying a matching approach to rich administrative data, the authors find sizable negative effects on the employment of participants within

[5] Defined here as up to eight years after the start of the program.

2.5 years after starting the program. This result is not entirely unexpected, however, as several program types have a duration of up to three years and may thus partly explain the difference to the other findings in the literature.

All the studies discussed attempt to address the problem of selection bias. In short, if only those unemployed participate in a training program who have good chances to find a job later-on anyway, comparing the later employment status of participants and non-participants does not reveal the causal average effect of the training, but simply picks up the difference in the underlying ability to take advantage of the newly gained knowledge, experience, or skill. While dynamic matching approaches in a combination with flexible matching techniques address the problem of dynamic sorting of the pool of unemployed eligible for treatment, they can not account for selection based on unobservables.

To our knowledge there are only few related studies that address treatment endogeneity in the evaluation of training programs and account for selection on unobservables in discrete time (see Abbring and van den Berg (2004) and Osikominu (2013) for studies in a continuous time framework that accounts for selection on unobservables). For instance, Aakvik et al. (2005) investigate the impact of Norwegian vocational rehabilitation programs on employment using discrete choice models in a latent index framework with unobservables generated by a normal factor structure. The authors use the degree of rationing at the local level as an instrument and find negative training effects after controlling for selection on observables and unobservables. Using rich administrative data for Switzerland, Frölich and Lechner (2010) exploit exogenous differences of participation probabilities within local labor markets as an instrument for training participation. Based on a combination of conditional IV and matching methods estimation, their findings suggest that in the short run training programs induce a 15 percentage points increase in the employment probability for compilers. For Germany, Fitzenberger et al. (2010) use Bayesian techniques in a dynamic framework and model selection into and out of training and employment based on observed as well as unobserved characteristics. Their results show an underestimation of the effect of training versus waiting in the medium and long run and an increased long-run employment originated in longer planned duration of training programs.

The present paper contributes to the literature in two ways. First, by trying to examine the effect of training programs from a different methodological angle. In particular, we exploit certain features in the budget policy of the FEO in the 1980s and early 1990s to obtain exogenous variation in the probability to participate in a training program and use this to instrument real participation in a control function approach. Second, by extending the period of examination after program participation to up to nine years, we hope to enable more accurate calculations about the overall benefit of training programs. In investment language, the cost of assigning an unemployed to a training programs is large at the beginning (both because of the direct costs and the foregone taxes and contributions in case the unemployed would have found a job in the meantime) and has to be recouped by positive effects on employment for several time after the treatment. If the potentially positive effects of training programs in the middle and long run hold firm over time, it would provide more support for their efficiency.

3 Data description

For this study, we use a new database which combines different administrative data sets collecting information on further training program participation in Germany.[6] The two main sources of information are the IEB data (*Integrated Employment Biographies / Integrierte Erwerbsbiografien*)[7] and the *FuU* data on program participation (also known as *St35* data). The IEB data are based on *daily* records on health insurance from employers and/or the FEO. They contain employment register data for all employees subject to social insurance contributions for the years from 1971 to the end of 2004. Thus, it provides a complete history of employment and unemployment periods for each individual, as well as information on the receipt of transfer payments from the FEO and a wide range of personal and job-specific characteristics. The FuU data set, on the other hand, consists of *monthly* information about participation in public-sponsored training programs between 1976 and 1997, collected by the FEO for controlling and statistical purposes. While it contains less information about the single individual, it provides a more detailed view on participation in training programs and enables a more precise and detailed identification

[6]The data was generated within the scope of the project "Policy Change, Effect Heterogeneity, and the Long-Run Employment Effects of Further Training" (IAB project Nr. 1213-10-38009). Preparing the data, we used the well documented experience of Stefan Bender, Annette Bergemann, Bernd Fitzenberger, Michael Lechner, Ruth Miquel, Stefan Speckesser, and Conny Wunsch (the project group in the previous project "About the Impact of Further Training Programs"). The main advantage of the new data is its large sample size. While Bender et al. (2005) merely used a 1% subsample, the current study is based on two 50% subsamples for training participants.

[7]In this project, we use a special version of the *IEB* data, which is adapted according to the BLH variable structure. In contrast to the conventional IEB version, our data only contain information from BeH (*Beschäftigten-Historik*) and LeH (*Leistungsempfänger-Historik* of the IAB).

of treatment types.[8]

As starting point for the construction of our sample, we combine a 50% subsample of participants of training programs from the FuU data with a 50% subsample of program participants from the IEB data. In addition, we use a 3% subsample from the IEB data without any program participation as control group, together with individuals who only entered a program at a later point in time.[9] In each data source, we only consider information starting in 1980 for reasons of data reliability.

The resulting data set is unique for Germany in three aspects: First, the length of the observation period including information for 25 years from 1980 to the end of 2004. Second, the size of the treatment group which contains 50% of all individuals who participated at least once in a training program within this period.[10] And third, the amount of information about each individual's employment and program participation history.

For our empirical analysis, we restrict attention to individuals entering unemployment at a particular calendar month after experiencing an employment period of at least three months. We follow these cohorts from the time they start their unemployment spell until December, 2004, which is the end of our observation period. If an individual enters unemployment more than once during that period, she is thus part of different unemployment cohorts and may appear several times in the empirical analysis.

The first cohort in our analysis consists of individuals entering unemployment in August, 1981.[11] We define treatment as the first training beginning within the first twelve months after the start of unemployment. Individuals who start training later belong to the control group, together with individuals who did not take part in any training at all. Additionally, we only include individuals in the control group if they received either unemployment benefits or unemployment assistance within the first unemployment quarter at the latest. Furthermore, we exclude individuals who enter unemployment at a particular point of time and remain without employment for more than 72 calendar months due

[8]In the merged data, the identification of treatment status and types of training programs is based on combining participation information from FuU data with transfer payment information from the IEB data, giving priority to the former.

[9]We identify program participants in the IEB data on the basis of transfer payment information, i.e., whether they obtained income maintenance payments of the type that indicates the participation in a training program.

[10]For reason of the computing power constraints, we reduce the control group by drawing a 10% subsample. More information on the construction of the data set, subsample, and weighting is provided in the additional appendix on page 46.

[11]Earlier cohorts are only used for the calculation of our instrument.

to issues of data quality.[12] Finally, we only consider individuals living in West Germany and aged 25 to 50 at the time of becoming unemployed.[13]

4 Budget surplus as instrument

Estimating the causal impact of training participation on subsequent employment must address various potential sources for a selection bias. First, it may be the case that those unemployed participate in training who are more able, motivated, and ambitious, and would therefore find it easier to get a job anyway. Second, caseworkers at the local employment offices could base their decisions about a) whether to offer an unemployed a training opportunity, and b) which type of program seems most appropriate, on their personal assessment of the potential benefits of participation for the respective individual. Third, unemployed may be assigned to training because of their particularly bad employment chances. While the first and second source imply a positive selection of participants, the third source implies a negative selection. When there is selection on unobservables (reflecting e.g. motivation or strive), OLS, Probit, or matching estimates of the effect of training participation on employment, while accounting for selection on observables, would yield biased estimates in all three cases.

We deal with the potential endogeneity of program participation by exploiting budget rules for active labor market policies in Germany which create a source of exogenous variation in training probabilities. We implement this instrumental variable (IV) approach using a version of the control function method for a random-coefficients model proposed by Wooldridge (2014). Our econometric approach will be described in section 5.

4.1 Institutional background

To explain the idea behind our identification strategy, it is necessary to present the institutional structure of the FEO and its budgeting system with respect to training programs at the time of our analysis. During the 1980s and early 1990s: (a) The FEO was organized in three tiers, with a nation-wide central office in Nuremberg, nine regional employment offices (REOs) at the intermediate level (*Landesarbeitsämter*, today known

[12] In particular, many of these long-term unemployment spells may be caused by gaps in the employment history, which are considered as non-employment in the data.

[13] The imposed age restrictions are necessary in order to avoid biased results in the probability to receive training due to some age-specific labor market programs, as for instance, for the youth or the elderly.

as *Regionaldirektionen*), and 142 local employment offices at the lowest institutional level (*Arbeitsämter*, or today *Arbeitsagenturen*).[14] (b) The total budget was determined and managed largely by the central office of the FEO, especially for all entitlement programs like income maintenance and training programs for the unemployed (Fertig and Schmidt, 2000). Local offices, on the other hand, possessed limited leeway in the use of their allocated funds for training programs. (c) The budget for training programs was planned and allocated separately from other instruments like job creation schemes and could not be transferred to other purposes. (d) The allocation of funds to the regional and then local offices was based primarily on past levels of program participation, but adjusted for anticipated changes in need for the next year. (e) Unused means from one fiscal year could not be carried over to finance measures in the following year. This institutional framework remained stable up until 1994, when new rules were enacted granting a modest level of budget autonomy to the local employment offices.

These features had implications for the management of training programs for the unemployed by the REOs during the time of our study. Most importantly, as the budget for the next year depended on the degree of utilization in the current year, REOs had an incentive to spend their whole budget before the end of the year, since they otherwise ran the risk of losing funds for the following year. However, an REO could not simply overuse its budget every year to secure a continuous rise of its budget. As a consequence, the best outcome in the self interest of local decision makers seemed to be to use the whole budget in a year in order to guarantee a stable and possibly growing budget for the next years. Thus, it seems very likely that the degree to which the budget was spent already at a given point in time may have influenced the decision of at a local level to assign an additional unemployed into training. If budgets were almost exhausted, caseworkers may have thought twice about sending more people, while the chance to get a training course paid by the employment office should have increased if resources were abundant and needed to be deployed.

There is a lot of anecdotal and suggestive evidence for end-of-year spending patterns in government agencies and company divisions (examples include Comptroller General (1980), Douglas and Franklin (2006), and McPherson (2007), for the former, and Merchant (1985), for the latter), but there is hardly any empirical evaluation due to a lack of reliable data (General Accounting Office, 1998). The only study to our knowledge is Liebman and

[14]Except for small states and city-states, the regional employment offices corresponded to the states in West Germany, see figure 1. The exception is the state of Bavaria with two regional employment offices.

Mahoney (2013), who investigate IT procurement decisions of various US federal agencies and find evidence for higher spending on lower quality projects in the last week of the fiscal year.

In the present paper, we use the difference between projected and actual spending in the first half of the year to instrument program participation in the final months of the year. The order of events is illustrated in figure 2, which displays the 12 months of a calendar year in a time line. Based on our understanding of the decision process in the REOs, the first semester review took place in July after all the information of the first six months was available. The accumulated budget surplus or deficit up to this point then influenced the participation decisions in the remaining months of the budget year after the summer holidays. Thus, the "end-of-year" period in our examination starts either in August or September depending on the region and year under consideration and ends in November as the costs for entries in December are taken out of the budget in January.[15]

There are two main reasons that support this strategy, both originating from personal interviews and correspondence with FEO experts and practitioners. First, because of the non-transferability of funds between different years, caseworkers intended to set training program starts as early as possible in the year in order to ensure that available funds were spent in the current fiscal year. Second, holiday periods needed to be taken into account for the planning, as many providers did not offer courses during vacations, since they required a stable group size and an economically viable number of participants. Therefore, the usual point in time for readjustments in the number of participants with respect to budgetary conditions were the months after the summer holidays ending in August or September depending on state and year. This should ensure that as many months of program participation as possible were paid for in the corresponding fiscal year. However, note that a participation in a long program which starts late in a year is mostly paid for out of the budget in future years.

Looking at the monthly shares of the total year entries into training programs in figure 3 supports this reasoning. It shows a pronounced seasonal pattern with relatively many entries in the early months of the year and a large spike after the summer holidays. On the contrary, the shares are lowest in June and July, and at the very end (December) of the calendar year. This confirms that caseworkers take both summer holidays as well as

[15]Due to a rotation system for the summer holidays of all public schools among the German states, the timing of the summer holidays varies from year to year. Thus, we treat August as the first month of our examination period for a region and year if at most half of the workdays in August belonged to the school holidays, and September otherwise.

the end of the budget year into account.

4.2 The derivation of the budget surplus

We follow three steps to construct our instrument "budget surplus", defined as the gap between planned and actual program entries per 1000 eligible unemployed over the first half of a calendar year for each regional employment office: (1) We calculate the intended number of new participants at the regional level for each month. (2) We compute the "budget surplus" for each of the first six months m in year j and region r using the following formula:

$$(1) \qquad surplus_{rmj} = \frac{(planned\ entries_{rmj} - actual\ entries_{rmj})}{1000\ eligible\ unemployed_{rmj}} \qquad (m \leq 6)$$

The resulting measure for monthly surplus is positive if the planned number of new participants exceeds the actual entries in the respective month and negative if the planned entries are lower than the actual ones. (3) To obtain the cumulative budget surplus over the first half of the year, the monthly numbers for each REO are aggregated over the first six months of each calendar. This cumulative surplus serves as our instrument for starting a training program during the last months of the year.

Two issues need to be discussed in more detail when using this approach. The first is our use of program entries to represent the budget. There is no information available on actual *and* planned expenditures for the nine REOs and budget rules at the time allow for a direct link between program entries and available funds. Since funds allocated to training programs could not be transferred to job-creation schemes or vice versa, there was no way to spend it differently than for training programs. This resulted in a close relationship between training budgets and the number of participants. Furthermore, the planning and allocation of budgets was done for training programs as a whole, that is, not distinguishing between the different program types. Thus, regional employment offices were assigned fixed head-count shares rather than monetary resources up until 1994 (Bach et al., 1993).

The second issue is how to get the exact numbers for the respective quantities. Based on our a 50% sample of all program participants, we can compute the number of new participants by region and month using sampling weights with high precision. However, determining a number for planned entries is more complex. Here, we estimate the number of intended entries into all training programs as a whole by month based on separate out-of-sample predictions for each of the nine REOs. That is, we regress the number of monthly entries on a full set of past year and calendar month dummies, as well as on the period number as a continuous variable and the monthly entries of the previous year. We then use the estimated coefficients to predict monthly entry rates for each year between 1982 and 1993 based on actual entries in all previous years covered by our data.

To strengthen the link between our forecasting procedure and the ex ante planned budget, we correct the number of predicted entrants by changes in allocated funds for training programs at the federal level. For this purpose, we calculate for each year the ratio between intended spending in that year and ex ante actual spending in the previous year.[16] A ratio less than one implies a reduction of overall spending on training programs and likewise in the absolute number of entrants compared to the preceding year. We multiply the raw predictions of entries with this ratio. Base on this correction, Figure 4 shows actual and predicted entries by REO during the time between 1980 and 1993. We can see from the graph that our forecasting procedure captures quite well both the trend and the seasonal variation in the entries.

Overall, the resulting surplus variable has a mean of 1.8 and a standard deviation of 8.5. This means that the typical REO could have sent almost 2 out of 1000 eligible unemployed more into training programs in the first half of the year, i.e., the executives on average ran surpluses during the first half of the year as could be expected by the budget rules discussed in section 4.1. Compared to the overall monthly average of 7 entries into training programs per 1000 eligibles, this average surplus over the first half of the year amounts to approximately one week worth of new program participants. There is, however, a large variation in first-semester surpluses across region and year, as the distribution of surpluses in figure 5 indicates. While the large majority of observations lie between -7 and 10, some REOs even experienced surpluses or deficits of more than 20 entries per 1000 eligible unemployed in some years. As a consequence, some REOs are very likely to expand training after the summer holidays in order to spend their whole budget, whereas others needed to cut back training substantially if they did not want to exceed their funds and get reproached for that.

[16]Data on intended and actual spending for FuU training programs on the federal level in West Germany is available only until the end of 1989.

4.3 Plausibility of the identification strategy

Relevance condition

In order to see whether this reasoning is true, we check first whether our instrument has predictive power for the probability to start a training program, i.e., whether it is actually relevant. We do this by running first stage probit regressions of the probability to enter the respective training program in the months after the summer holidays on the size of a region's budget surplus, controlling for personal characteristics, work biographies, and indicators for each region and year.[17] The resulting average partial effects (APE) are reported in table 1. The corresponding results for the evaluation sample with the reduced control group are very similar (see table 3).

Table 1: First stage probit estimates of treatment (entire sample)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Evaluated program	RT	PF	SPST	STT	WS	Short SPST	Long SPST
Surplus	0.0473***	0.0415***	0.0253	-0.0342***	0.0110	-0.0211*	0.0445***
	(0.0139)	(0.0144)	(0.0173)	(0.0112)	(0.0083)	(0.0127)	(0.0134)
F-statistic	11.91	16.52	2.14	9.29	2.03	2.77	11.74
No. of clusters	540 000	528 649	551 076	455 912	527 238	538 665	536 606

Note: Average partial effects per 1000 unemployed (in percentage points) obtained from a first stage probit regression of the probability to participate in retraining (RT), practice firms (PF), specific professional skills and techniques (SPST), short-term training (STT), wage subsidies (WS), short SPST, and long SPST. Standard errors are clustered at individual level and reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10% level, respectively.

The estimates show that for three out of the five main types of training programs (practice firms, retraining, and short-term training) the probability to participate in the final months of the year reacts strongly to our measure of budget surplus in the first half of the year. In particular, the chances of entering one of the typically more costly training programs (PF and RT) increase significantly with the size of the surplus, while they decrease for short-time training which is cheaper and usually only lasts for a couple of weeks or a few months. To illustrate the magnitude of the effect, a one standard-deviation increase in surplus (i.e., not having filled 8.5 available program spots in the first half of the year) leads to an increase in the individual probability to participate in retraining in the months after the holidays of 0.4 percentage points (ppoints). The corresponding numbers for practice firms and short-term training are +0.35 ppoints and -0.29 ppoints,

[17] A list of description of all explanatory variables used in our empirical analysis is provided in table 2.

respectively. These findings suggest that officials at the employment offices reacted to some extent to the budgetary environment, with the aim of using up remaining funds by the end of the year (or to meet the budget in the case of deficits).

We do not find any effect of budgetary conditions on participation in the wage subsidy program or in courses that provide specific professional skills and techniques (SPST), although the latter are the most common type of further training for the unemployed. In the case of wage subsidies, this seems very plausible, as they are only granted to employers after they submit a detailed application form. That is, adjustments in the size of this program are slow and depend on the labor demand by firms. Therefore, participation in the program can not be changed quickly and unilaterally by the employment office. In the case of the SPST courses, however, it is possible that the relevant adjustments in spending were not made on the extensive, but rather the intensive margin, i.e., that budget surpluses did not affect the absolute number of entrants, but the length of the programs they were sent to.[18] We examine whether this happened in the case of budget surpluses by looking at entries into long and short SPST programs separately, where we defined long courses as having an intended duration of more than six months. The results in columns 6 and 7 of table 1 show that a surplus changes the composition of SPST programs towards longer durations both by reducing entries into a short SPST program and by increasing entries into a long SPST program. This finding matches the results for the other programs, namely zero or negative effects for short programs and positive effects for long programs. Note the increase in the entries into longer programs caused by a budget surplus also absorbs parts of the budget in subsequent years. This can be interpreted as a spending commitment in the future which provides a safeguard against future budget cuts.

Overall, the results of the first stage indicate a significant impact of the surplus from the first half of the year on the probability to participate in individual training program in the months after the summer holidays. To avoid a problem of weak instruments, we estimate treatment effects only for those programs where the first stage F-statistic is larger than 10. In our case, this applies to retraining, practice firms, and long SPST programs.

[18]This possibility was mentioned during our discussions with REO officials.

Exclusion restriction

Apart from having an impact on treatment (relevance), our instrument, the budget surplus, must not directly influence our outcome variable, employment, or, put differently, no omitted variable simultaneously affects the budget surplus and employment. An obvious issue could be persistent trends in regional labor market conditions.[19] For instance, a sustained economic upturn may lead to both budget surpluses and higher labor demand in one region, while a persistent downturn could cause the opposite situation.

During the period of our analysis, this does not seem to be a problem. For one, these trends are likely to influence both the treated and the untreated in our sample, in a way that they should not influence the results. Additionally, there were no consistently different trends in the regional unemployment rates across the various REOs between 1982 and 1993, as shown in figure 6 (upper panel), rather a slight convergence to an average unemployment rate of about 7.5%. Besides, as long as the development of regional labor market conditions was expected, this would have led to a corresponding adjustment in the allocated budget for training programs for that region.

This reasoning is supported by the development of the surplus variable per region over time, as depicted in figure 6 (lower panel). It demonstrates that no region remained at either side of the deficit-surplus measure consistently, which suggests that there is strong random variation in the instrument by region and time. Finally, we also run a regression of the surplus on the average unemployment rate in a region three years down the road when program participation should be finished. The estimated coefficient is very small and insignificant (the results are available upon request), thus suggesting that indeed our surplus variable should affect the employment status of an individual only through its effect on the probability to enter a training program.

5 Econometric approach

We estimate treatment effects accounting for the dynamic sorting of participants with respect to elapsed unemployment duration following Sianesi (2004, 2008).[20] Effectively, we estimate the effect of training versus waiting, where non-treatment up to a certain

[19] Unobserved personal characteristics of the unemployed should not play a role here, although they may strongly influence employment. This is due to the construction of the surplus variable at a more aggregate level, where there is no obvious link between the instrument and unobserved individual characteristics.

[20] See Fredriksson and Johansson (2008) for a critical assessment of this approach, and Biewen et al. (2014), Fitzenberger et al. (2013), or Lechner et al. (2011) for recent applications.

elapsed unemployment duration involves the possibility to be treated later during the course of the unemployment spell. To implement our instrumental variable (IV) approach, we use the flexible control function method for a random-coefficients model with both a binary endogenous treatment and a binary outcome, see Wooldridge (2014, section 6.4).

Consider a random sample of eligible individuals, who can enter a training program $p = \{RT, PF, SPST, STT, WS\}$ at any possible elapsed duration $el = \{1, ..., 12\}$ within the first year of unemployment. We define treatment as the first participation in one of the aforementioned training programs during the respective unemployment spell. To reduce the dimensionality of estimated treatment effects, we distinguish between four possible strata, $s = \{1, 2, 3, 4\}$, corresponding to treatment starting in the first, second, third, and fourth quarter of unemployment, respectively. Further, we ensure that treatment and comparison groups have the same elapsed unemployment duration up to the start of the treatment by restricting the group of eligible non-participants to those individuals that are still unemployed in the month prior to the beginning of the respective stratum and that will not started any of the aforementioned training program before the end of the stratum. In addition, to account for the fact that the composition of treated and non-treated individuals depends upon the monthly elapsed duration, all individuals in our panel data set used for estimation are replicated for each month they remain unemployed and are eligible for treatment. An example, if an individual starts a treatment in month 7 of the unemployment spell and does not find a job before month 8, the individual is used as control observation for treatment starts during months 1 to 6 (first two strata, $s = 1, 2$) and as treatment observation for month 7 (third stratum $s = 3$). For simplicity, we suppress the indices p and s in the following, and we refer to the references in the first paragraph for a detailed presentation and discussion of the dynamic treatment approach taken here. The average effect estimates reported below by stratum are the weighted averages of the three corresponding month-specific estimates (see e.g. Biewen et al. (2014) for a formal description). The average effect estimates for participation during the first twelve months of unemployment (all four strata) reported below are again the weighted averages of the twelve month- (and strata-) specific estimates. The weights are the monthly (strata-specific) number of entries into treatment.

In the remainder of this section, we discuss the estimation of the average treatment effect for the treated, who start treatment in month el of their unemployment spell. We define N_1 as the number of treated in el and N_0 the number of eligible nonparticipants in month el. Then, the total sample of eligible unemployed, indexed by $i = 1, ..., N$, involves

$N = N_1 + N_0$ observations and N varies by el. We omit the index i when describing the data generating process.

The outcome variable of interest, y_t, is a dummy variable for employment (without receiving a wage subsidy with a training component) at time t, where $t = 0, 1, ..., T$ is time since treatment start. Following the potential outcome framework (Roy, 1951; Rubin, 1974), we assume that for each individual there are two potential outcomes, $\{y_t^1, y_t^0\}$ at time t associated with a time invariant binary treatment indicator d determined at time $t = 0$ (month el). $d = 1$ indicates participation in training and $d = 0$ non-participation, respectively, as explained above. We assume further that the observed outcome variable is expressed in terms of potential outcomes as $y_t = y_t^0 + (y_t^1 - y_t^0)d$. We impose the following latent index structure for the two binary indicators y_t and d:

$$\text{(2a)} \qquad y_t = \mathbb{1}\left[y_t^* \geq 0\right] = \mathbb{1}\left[a_{t0} + z_1 b_0 + (b_{t0} + z_1(b_1 - b_0))d + b_t^d d + u_t \geq 0\right]$$

$$\text{(2b)} \qquad d = \mathbb{1}\left[d^* \geq 0\right] = \mathbb{1}\left[\gamma_0 + z_1\gamma_1 + z_2\gamma_2 + \nu \geq 0\right] = \mathbb{1}\left[z\gamma + \nu \geq 0\right],$$

where y_t^* and d^* are latent indices, z_1 involves the observed exogenous covariates, z_2 is the set of excluded instruments, $z \equiv (z_1, z_2)$, and b_t^d, u_t and ν are unobserved random variables. $\mathbb{1}\left[\mathbb{A}\right]$ denotes the indicator function with a value of one if $\mathbb{A}$ is true and of zero otherwise. We allow for separate time effects (a_{t0}, b_{t0}) and for separate effects (b_0, b_1) of the covariates z_1 by treatment status. We assume a probit model for the treatment dummy, i.e. $\nu \mid z \sim \mathcal{N}(0, 1)$. The potential outcome representation in equation (2a) accounts for selection into treatment based on observable characteristics z_1, z_2, unobservables u_t, and unobservable random gains from treatment b_t^d. b_t^d has expectation zero ($E(b_t^d \mid z) = 0$). Selection on unobservables is reflected in the statistical dependency between ν and (b_t^d, u_t) (Wooldridge (2014), section 6.1, Blundell et al. (2005), section 3.4.1). Furthermore, we assume that b_t^d and u_t each follow a univariate normal distribution and that (b_t^d, u_t, ν) follow a joint continuous distribution which is independent of z.

The variables z_1 involve information on (i) individual characteristics like gender, age, education, family status, nationality; (ii) occupation- and job-related variables from previous employment like employment status, earnings, firm size, and industry structure; (iii) individual work history and indicators of former participation; (iv) regional information on the state level as well as time-specific variables. The instrument z_2 involves the budget surplus as described in the previous section.

To clarify the effect heterogeneity accounted for by equation (2a), the effect of treatment (the partial effect of d) on the employment index y_t^* for an individual i with covariates z_{i1} and unobservable b_{it}^d can be written as $b_{t0}+z_{i1}(b_1-b_0)+b_{it}^d$. Therefore, we account in a flexible way for effect heterogeneity both with respect to observables z_1 and unobservable b_t^d. However, because y_t is a nonlinear function of y_t^*, the individual specific treatment effects for y_t^* (or averages thereof) do not translate directly into the treatment effects (partial effects of d) for y_t (Wooldridge, 2005; Terza, 2009).

We estimate the average effect of treatment on the treated (ATT) for employment at time t given by

$$\tau_{ATT,t} \equiv E\left[y_t^1 - y_t^0 \mid d=1\right] = Pr(y_t^1 = 1 \mid d=1) - Pr(y_t^0 = 1 \mid d=1) \tag{3}$$

for which we first have to take the discrete difference between the average potential outcomes for the treatment and the nonparticipation outcome. The average potential outcome for $\tilde{d}=0,1$ is given by the conditional expectation integrating out the distribution of (z_1, b_t^d, u_t) among the treated $d=1$. By the law of iterated expectations, we can do this in two steps based on

$$Pr(y_t^{\tilde{d}} = 1 \mid d=1) = E_{z_1|d=1}\left\{ASF(\tilde{d}, z_1, d=1)\right\}$$

where the average structural function (ASF), which was originally introduced by Blundell and Powell (2003, 2004),[21] is defined as

$$ASF(\tilde{d}, z_1, d=1) = \tag{4}$$

$$E_{b_t^d,u_t|d=1,z_1}\left\{\mathbb{1}\left[a_{t0} + z_1 b_0 + (b_{t0} + z_1(b_1 - b_0))\tilde{d} + b_t^d\tilde{d} + u_t \geq 0\right]\right\}.$$

Note that $Pr(y_t^{\tilde{d}} = 1 \mid d=1)$ does not depend upon z_2 because of the exclusion restriction for z_2 in equation (2a).

Following Wooldridge (2014, section 6.3), we assume that we have a set of control functions $e_2 = k_2(d, z_i, \theta)$, which are functions of the treatment dummy d, the exogenous covariates z, and some unknown parameters θ, which are redundant in equation (2a) by construction and which act as a sufficient statistic for capturing the endogeneity of d. e_2 is a sufficient statistic, if the distribution of (b_t^d, u_t) conditional upon d and z depends only

[21] Blundell and Powell (2003, 2004) and Wooldridge (2014) defines the ASF for the entire sample. Because of our interest in the ATT, we restrict attention to the treated, and we define the ASF for the two potential outcomes conditional on $d=1$.

upon e_2. A formal discussion of these conditions and the implied estimation approach is given in the additional appendix on page 47 and the following pages.

Under these assumptions, the average structural function can be expressed by integrating out the control functions e_2 as

$$ASF(\tilde{d}, z_1, d = 1) = \tag{5}$$

$$E_{e_2|d=1,z_1}\left\{Pr\left(a_{t0} + z_1 b_0 + (b_{t0} + z_1(b_1 - b_0))\tilde{d} + b_t^d \tilde{d} + u_t > 0 \mid d = 1, z_1, e_2\right)\right\} .$$

For estimation purposes, we assume that $b_t^d \tilde{d} + u_t$ is normally distributed conditional on e_2 and the expectation is a linear function of e_2. To operationalize the estimator, we first estimate the probit equation (2b) and calculate the estimate of the inverse Mills ratios $\lambda(z\gamma)$ and $\lambda(-z\gamma)$. Following Wooldridge (2014, section 6), e_2 then involves the generalized residual $gr = d\lambda(z\gamma) - (1-d)\lambda(-z\gamma)$ (the standard Heckman (1978) selection correction), its squares, interactions with z_1, and the interaction of gr with the treatment dummy d. We add estimated versions of these control functions to a second stage probit regression of employment, where we regress y_t on d, z_1, interactions between d and z_1, time effects m_t and interactions between m_t and d, and e_2. In its most general specification, the estimated regression for observation i ($i = 1, ..., N$) is

$$\widehat{Pr}(y_{it} = 1 \mid d_i, z_{i1}, e_{i2}) = \tag{6}$$

$$\Phi\left(m_t\hat{\delta}_{0t} + m_t\hat{\delta}_{1t}d_i + z_1\hat{b}_0 + z_1\hat{\delta}_1 d_i + \hat{\omega}_0\widehat{gr}_i + \hat{\omega}_1\widehat{gr}_i d_i + \hat{\omega}_2\widehat{gr}_i^2 + z_{i1}\widehat{gr}_i\hat{\psi}\right) ,$$

where m_t represents a full set of time dummies, $\Phi(.)$ is the standard normal distribution function, and $\hat{\delta}_{0t}$, $\hat{\delta}_{1t}$, $\hat{b}_0$, $\hat{\delta}_1$, $\hat{\omega}_j$ ($j = 0, 1, 2$), $\hat{\psi}$ are coefficient estimates. As part of the specification search in our application, we routinely test for significance of $\widehat{gr}_i^2$ and particular interaction terms in $z_{i1}\widehat{gr}_i$ and we drop insignificant terms for the control function. Equation (6) is estimated as a pooled probit over time t. In order to keep the estimation approach tractable, we impose the restriction that the coefficients for the selection correction terms $\hat{\omega}_j$ and $\hat{\psi}$ are time-invariant.

Based on the estimated equation (6), we can estimate the ATT for period t by integrating out the distribution of $z_{i1}, \hat{e}_{i2}$ among the treated $d_i = 1$ as

(7)

$$\widehat{\tau}_{ATT,t} = \frac{1}{N_1}\sum_{d_i=1}\left\{\Phi\left(m_t\hat{\delta}_{0t} + m_t\hat{\delta}_{1t} + z_1\hat{b}_0 + z_1\hat{\delta}_1 + \hat{\omega}_0\widehat{gr}_i + \hat{\omega}_1\widehat{gr}_i + \hat{\omega}_2\widehat{gr}_i^2 + z_{i1}\widehat{gr}_i\hat{\psi}\right)\right.$$

$$- \Phi\left(m_t\hat{\delta}_{0t} + z_1\hat{b}_0 + \hat{\omega}_0\widehat{gr}_i + \hat{\omega}_2\widehat{gr}_i^2 + z_{i1}\widehat{gr}_i\hat{\psi}\right)\Big\} \ .$$

Note that the same control function terms for treated individuals apply to the potential treatment and nontreatment state except for the coefficient $\hat{\omega}_1$ being 'switched off' in the nontreatment state (second term). This is because the $\hat{\omega}_1$-component of the individual selection effect does not have an impact in the nontreatment state (see Blundell et al. (2005), section 3.4.1).

We do inference on $\widehat{\tau}_{ATT,t}$ by applying a cluster bootstrap based on 250 replications, where we re-estimate the entire estimation procedure for each resample. The unemployed individual is used as the cluster unit accounting for the fact that an individual can be both a control observation and a treatment observation for different months el of treatment start.

To assess the influence of modelling assumptions and the importance of accounting for selection based on unobservables, we contrast the estimated ATT's based on equation (7) with ATT estimates based on pooled OLS regressions with the same specification for observables as in equation (6) but without the control functions for selection on unobservables.

6 Effects of training on subsequent employment

6.1 Descriptive statistics

Table 4 provides an impression of the relevant sample sizes in our evaluation sample with the reduced control group. It reports the size of the treatment group for each of the three evaluated types of training programs and by length of unemployment at program start. Additionally, we also show the size of the respective control groups, which consist of all eligible unemployed who do not take part in a training program at the time when participants start theirs. This also includes everybody who enters a training program later on. From the numbers in the table, we can see that retraining is the dominant type among the evaluated courses. Furthermore, the absolute number of entries into training programs declines with higher duration of unemployment, which is intuitive since the group of eligible gets smaller as more and more individuals find a job. At the same time, the ratio of individuals who start a program to those who do not tends to increase over the length of unemployment, e.g., for long SPST programs it increases from 14% in the first stratum to 18% in the fourth. This may reflect an increasing awareness of case workers in the employment offices that something has to be done for this group of people in order to prevent long-term unemployment. By contrast, for retraining we do not observe a clear tendency.

Table 5 presents the comparison of average characteristics between individuals in the treatment and control group of the respective evaluated program in our analysis.[22,23] They show that in many dimensions, program participants are not very different from non-participants or between different programs. Some characteristics differ, however.

For instance, while budget surplus is almost identical across control groups (with exception of short-term training), we observe some variation across the treatment groups of different programs. For retraining and practice firms, surplus is higher in the treatment compared to the control group, meaning that these programs were offered more in the regions with budget reserves. On the other hand, the reverse is true for long SPST programs. Furthermore, the share of foreigners is smaller in the treatment group than in the control group in all programs (5-9% vs. 12%, respectively). Additionally, we observe more females in long SPST training programs than in other programs and the control group (53% vs. 48% in the control group and 32 and 42% in retraining and practice firms, respectively).

Next, we note that a majority of over 50% of program participants is between 25 and 34 years old in all program types, which mirrors the age composition of the unemployed without participation. As expected, we observe a decreasing participation incidence with the older age. The only difference is that there seems to be a higher concentration of young people in retraining compared to the other training programs and the control group. While in all programs the share of individuals aged 25-29 amounts to approx. 30%, it is nearly 42% for retraining. Likewise, the share of older individuals (aged 45-50) in retraining amounts to only 4.5%, whereas it is 13-14% in all other programs. This supports the notion of training program participation as an investment in your own human capital.

With respect to education, most program participants possess some kind of vocational training degree (more than 70% in each program type). Apart from that, however, we see a larger fraction of individuals without vocational training in practice firms, and more people with higher-education degrees in long SPST programs and retraining. The same

[22] A detailed description of the variables and their definition is reported in table 2.

[23] Descriptive statistics for short-term training and wage subsidies are available from the authors upon request.

is true compared with the control group.

6.2 Estimated employment effects

We estimate the effect of training on subsequent employment of eligible unemployed individuals using the CF approach described in section 5. The empirical findings from the first stage probit regressions show that our instrument, the regional budget surplus, reliably affects program participation for retraining (RT), practice firms (PF), and specific professional skills and techniques with a planned duration of more than six months (long SPST). In each of these cases, the F-value from the respective first stage regression is greater than 10. In order to avoid the problem of weak instruments, we therefore concentrate our analysis of employment effects on these three types of training programs.

The results of these specifications are displayed in figures 7 to 9. The darker lines depict the IV-control function estimates for the employment effect of the respective training program for each month from its start until 9 years afterwards. The corresponding bootstrapped confidence intervals are depicted in weaker dashed lines.[24] Every point estimate represents the weighted average of the separate estimates for the four different quarters of unemployment duration in which an individual could enter the training program.[25] To be able to compare our results with those of a standard OLS regression of employment on training participation and the same set of controls, we additionally report the estimated coefficients for a simple OLS regression as a continuous gray line.

The findings confirm the already well-established pattern of "lock-in" effects in the short run, i.e., participation in any kind of training program leads to negative employment outcomes in the first months, as the unemployed reduce their search efforts during that time. The duration and strength of these initial negative effects depends on the type of program. On the one hand, the longer the program duration, the longer the lock-in effects seem to last. Thus, while it takes about 12 months to get back to the initial difference in employment rates compared to non-participants in the case of practice firms and SPST, we observe a lock-in period of around 24 months for retraining. On the other hand, the shortest program, practice firms, seems to have the strongest retention effect, with a negative effect of around -35 ppoints at the peak, compared to maximum negative effects of -20 ppoints for the longer-running SPST and retraining.

[24]We ran 250 repetitions to obtain the correct bootstrapped standard errors in each case.

[25]The weights used are the fractions of participants who were in the respective quarter of their unemployment spell when they started the program.

Towards the end of each program's duration, the estimates increase again and then turn positive for retraining and long SPST programs. Thus, we find stable medium- and long-run effects for these programs with a magnitude of around 20 ppoints for retraining and 5 to 10 ppoints for long SPST programs, although the latter result is not statistically significant due to the large standard errors of the estimation. It is striking, however, that taking part in a practice firm seems to cause a sizable negative effect even in the long-run. The respective CF estimates are statistically significant and fluctuate around -20 ppoints up to the end of our examination period. Compared to the results of OLS specifications with the same covariates, these findings indicate higher returns to taking part in retraining and SPST programs, but much worse outcomes for working in a practice firm than what we would expect from the respective OLS estimates.

These results lead us to two important insights. The first is that the effectiveness of the different further training programs varies stronger than previously thought. Programs which focus on the acquisition of new skills fare exceedingly better than programs with an emphasis on exercising and practicing already existing skills. In fact, based on the above findings of negative effects in the long run, the existence and format of the latter should be closely examined and reconsidered in order to achieve better results for the unemployed with the allocated funds. The second insight relates to the empirical question of whether selection into further training programs takes place and which way it goes. Here again, the results suggest that the answer is not straightforward. While the CF estimates show more negative effects of participating in a practice firm than the corresponding OLS ones, indicating a positive selection of participants with respect to important unobservable characteristics, we find the reverse for the longer programs retraining and SPST. This means that we cannot confirm general claims or assumptions of positive selection into further training programs, but rather have to examine each program and its characteristics individually.

In the case of the examined programs in the present study, we have a situation with negative selection on unobservables into the longer programs, while at the same time observing more highly educated individuals there than in the short program and the control groups. A possible reason could be that caseworkers may expect a greater payoff for higher educated individuals from longer and more education intensive programs, but send those unemployed whom they believe to lack the necessary ambition and drive to succeed in the labor market anyway to these programs. Similarly, caseworkers may think that practice firms are not a good way to improve the chances of higher educated individuals, but that they can be used to keep the most motivated busy or to reward them for their good intentions, as they may gain additional months of welfare benefit receipt that way.[26]

6.3 Tests for robustness

The next step is to test the robustness of these results. We do this in two ways: First, we want to see how our first stage estimates react to certain changes in the specification. That is, we take a closer look at what drives the result that the budget situation after the first half of the year affects the probability to enter a training program in the months after the summer holidays. To do this, we run five different regressions of the probability to enter a specific program on our surplus variable, each time varying either the set of covariates or the computation of budget surplus. The results for each of the programs under consideration, as well as for short and long SPST programs separately, are displayed in table 6. In the first column, we start off by simply regressing the probability to enter the respective program on our surplus variable. Then, we sequentially add time-, region-, and person-specific controls to the econometric model in columns 2 to 4. And finally, we use all controls, but change the computation of budget surplus in column 5. In this case, we drop the correction of predicted entries by changes in the aggregate spending for job training programs, as the necessary information for this is only available up until 1989,[27] which may cause a structural break in the quality of our data. The results show that adding time and regional specific information influences the estimates, but we cannot find a general pattern of this effect, neither on the coefficients and their signs and significances, nor on the resulting F-test statistics. Besides, our approach relies on the variation in budget surplus that is not explained by year or regional fixed effects, so including them should matter for the relevance of our instrument. More important for the robustness of the first stage is therefore whether adding controls that are not strictly necessary changes size and significance of the effect of budget surplus. When comparing the coefficients reported in columns 3 to 5, we see that this is not the case. This shows that our result in the first stage is not affected by the particular choice of covariates or the correction factor in the computation of surplus.[28]

[26]Throughout the time of our examination, individuals younger than 43 could only receive a maximum of 12 months of unemployment benefits.

[27]Due to the reunification of Germany in 1990, the statistical yearbooks of the Federal Employment Office do not report the budget for training programs for West Germany separately anymore.

[28]We also test whether we get different results in the second stage when we use the instrument without correction factor. The outcomes are very similar there as well, both in terms of sizes and significances.

Second, we examine how the choice of evaluation method influences our main findings in the second stage, i.e., the effect of program participation on later employability. We therefore rerun the analysis using both OLS and IV (using predicted probability of participating in training as instrument) and compare the results with those obtained by our control function approach. The estimates are displayed in table 7. For brevity, we abstain from reporting them for every single month after treatment start, but pool them for three distinct periods over the evaluation time: the first year to compare the measures for the lock-in effect, the second year to capture the transition from program participation to work, and years three to nine for the long-run effects on employability. Three aspects of table 7 are noteworthy. First, there is no difference between the three methods with respect to the pattern of results they produce. That is, all three of them exhibit significantly negative effects of program participation in the first year, slightly better ones in the second, and the comparatively best results in the long-run. Second, the estimates obtained by OLS differ in magnitude from those of the two types of IV approaches, indicating again the strong presence of selection into program participation. Third, and most importantly for us, the results seem to be pretty robust between IV and our control function approach. While they differ somewhat in the point estimates, we generally find the same sign, significance, and order of magnitude over the three evaluation periods. Only in the case of long SPST programs do we get slightly more positive results with the CF approach compared to the IV estimation. All in all, these findings make us more confident in the results obtained in the second stage.

6.4 Heterogeneous effects by elapsed duration

In a last step, we evaluate the employment effects of participating in one of the examined training programs separately by the elapsed time in the unemployment spell. The main reason for doing so is a practical one: If the effects differ by the length of time already in unemployment, this could provide valuable information for the design and conduct for policymakers and officials at the job centers. Additionally, individuals with different durations of unemployment may not be the same with respect to their unobservable characteristics and the incentive structure they are facing, so we want to restrict our comparison to those individuals best comparable.

The respective results for retraining, long SPST programs, and practice firms are presented in figures 10 to 12, respectively. The four graphs per figure display the outcomes for the different quarters of unemployment duration. Looking at each examined program in turn, we can see in figure 10 that retraining has a positive and significant effect on later employment independent of the time the individual has already spent in unemployment. The only (small) difference across strata lies in the magnitude of the point estimates, which are a bit bigger in the first two strata than in the later two (more than 20 ppoints vs. less than 20 ppoints, respectively).

The effects of SPST programs behave in a similar way, as can be seen from figure 11. Although the estimates are significantly different from zero and the OLS results only in the case of starting the program in the second quarter of unemployment, we observe that the point estimates for the medium- and long-run effects are, on average, higher in the first two strata than in the later ones. In numbers, the estimated coefficients 9 years after the start of the program are +6 ppoints and +20 ppoints in the first and second stratum, respectively, while they are close to zero in both the third and fourth. Thus, both types of training programs seem to have a larger impact when their participants did not experience a long time in unemployment before. This could be the case either because the capacity to properly learn new skills deteriorates over time in unemployment, or because the signal of starting a course late is worse than beginning it shortly after losing the job.

For practice firms, on the other hand, we see a different pattern in figure 12. Here, the point estimates for the long-run effect after 9 years get better (less negative) the longer the participating individuals have been unemployed before entering the program. In particular, the estimated coefficient changes from -25 ppoints in the first stratum to around -20 ppoints, -18 ppoints, and -15 ppoints in the second, third, and fourth, respectively. For participants entering a practice firm in their 10th to 12th month of unemployment, the estimated effect is even insignificantly different from zero, although this is partly caused by the more imprecise estimation in the fourth strata due to the smaller number of treated observations. This suggests that practice firms are more useful for longer-term unemployed and that their use should be concentrated on this group of eligible, if at all. A possible reason for this finding may lie in the nature of the practice firm program which focuses on exercising already existing skills rather than teaching new ones. This may be more helpful for individuals who are about to lose their attachment to the labor market than for those who only recently lost their jobs and do not need a refreshment of their skills.

To summarize, the analysis by duration of unemployment provides an interesting and relevant result. Programs which focus on learning new skills and techniques should start early in the unemployment spell to have the largest positive impact, whereas programs concentrating on practicing and revising existing skills should be used exclusively with longer-term unemployed.

7 Conclusions

This paper studies the long-run effects of public-sponsored training programs for the unemployed. In order to come closer to estimating the causal effect, we take advantage of strict budget rules in the 1980s and early 90s to implement an IV strategy that instruments program participation by whether and how much the respective regional employment office spent less than planned in the first half of the year. Since transferring funds from one instrument of active labor market policy to another or into the next year was not allowed, employment offices running a surplus faced incentives to increase their spending in the months following the summer holidays in order to preserve their budget in the next year. We show that this need for end-of-year spending led to an increase in the probability to enter certain types of training programs which is not related to the observed or unobserved characteristics of an unemployed individual.

Based on this result, our empirical analysis leads to the following main findings: a) The long-run effects of further training on employment are not homogenous but differ by the type of program. Measures focusing on teaching new skills increase the job chances of participants, whereas programs which emphasize the practice of existing skills reduce the later success on the labor market. b) Selection based on unobservable characteristics like motivation, ambition, and strive is not uniformly positive as often assumed. While this is true for participants in practice firms, the opposite seems to be the case for both retraining and SPST programs. c) Examining the program effects separately by the length of previous unemployment reveals that retraining and long SPST programs should start very early after the loss of a job, whereas practice firms should be used only in the case of longer-term unemployment, if at all.

These findings suggest that the public debate about the usefulness of public-sponsored job training programs is often misled. Instead of talking about keeping or abolishing these instruments, the focus should lie on the details and merits of each individual program and how it can be improved.

References

Aakvik, A., J. Heckman, and E. Vytlacil (2005). Estimating treatment effects for discrete outcomes when responses to treatment vary: an application to Norwegian vocational rehabilitation programs. *Journal of Econometrics 125*(1–2), 15–51.

Abbring, J. and G. van den Berg (2004). The Nonparametric Identification of Treatment Effects in Duration Models. *Econometrica 71*(5), 1491–1517.

Bach, H.-U., H. Kohler, H. Leikeb, E. Magvas, and E. Spitznagel (1993). Der Arbeitsmarkt 1993 und 1994 in der Bundesrepublik Deutschland. *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung 26*, 445–466.

Bender, S., A. Bergemann, B. Fitzenberger, M. Lechner, R. Miquel, S. Speckesser, and C. Wunsch (2005). Über die Wirksamkeit von FuU-Maßnahmen. *IAB: Beiträge zur Arbeitsmarkt- und Berufsforschung 289*, 410.

Biewen, M., B. Fitzenberger, M. Paul, and A. Osikominu (2014). The Effectiveness of Public Sponsored Training Revisited: The Importance of Data and Methodological Choices. *Journal of Labor Economics 32*(4), 837–897.

Blundell, R., L. Dearden, and B. Sianesi (2005). Evaluating the effect of education on earnings: models, methods and results from the National Child Development Survey. *Journal of the Royal Statistical Society: Series A 168*(3), 473–512.

Blundell, R. and J. Powell (2003). Endogeneity in Nonparametric and Semiparametric Regression Models. *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress II*, ed. by M. Dewatripont, L.P. Hansen, and S.J. Turnovsky. Cambridge, U.K.: Cambridge University Press, 312–357.

Blundell, R. and J. Powell (2004). Endogeneity in Semiparametric Binary Response Models. *Review of Economic Studies 71*(3), 655–679.

Bundesanstalt für Arbeit (1993). Geschäftsbereicht der Bundesanstalt für Arbeit (ANBA).

Card, D., J. Kluve, and A. Weber (2010). Active labor market policy evaluations: A meta-analysis. *The Economic Journal 120*(4), 742–784.

Comptroller General (1980). Federal Year-End Spending - Symptom of a larger problem. *US House of Representatives*.

Douglas, J. W. and A. L. Franklin (2006). Putting the Brakes on the Rush to Spend Down End-of-Year Balances: Carryover Money in Oklahoma State Agencies. *Public Budgeting & Finance 26*(3), 46–64.

Fertig, M. and C. Schmidt (2000). Discretionary measures of active labor market policy: the German employment promotion reform in perspective. *IAB Discussion Paper 182*.

Fitzenberger, B., O. Orlanski, A. Osikominu, and M. Paul (2013). Déjà Vu? Short-Term Training in Germany 1980-1992 and 2000-2003. *Empirical Economics 44*(1), 289–328.

Fitzenberger, B., A. Osikominu, and M. Paul (2010). The heterogeneous effects of training incidence and duration on labor market transitions. *IZA Discussion Paper 5269*.

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 Statisque 91–92*, 321–355.

Fitzenberger, B. and S. Speckesser (2007). Employment Effects of the Provision of Specific Professional Skills and Techniques in Germany. *Empirical Economics 32*(2/3), 529–573.

Fitzenberger, B. and R. Völter (2007). Long-Run Effects of Training Programs for the Unemployed in East Germany. *Labour Economics 14*(4), 730–755.

Fredriksson, P. and P. Johansson (2008). Dynamic Treatment Assignment – The consequences for evaluations using observational data. *Journal of Business and Economic Statistics 26*(4), 435–445.

Frölich, M. and M. Lechner (2010). Exploiting regional treatment intensity for the evaluation of labour market policies. *Journal of the American Statistical Association 105*(491), 1014–1029.

General Accounting Office (1998). Year-end spending: Reforms underway but better reporting and oversight needed. *Publication No. GAO/AIMD–98–185 Washington, D.C.: U.S. Government Printing Office*.

Heckman, J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica 46*(4), 931–959.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Kimhi, A. (1999). Estimation of an endogenous switching regression model with discrete dependent variables: Monte-Carlo analysis and empirical application of three estimators. *Empirical Economics 24*(2), 225–241.

Lechner, M., C. Wunsch, and R. Miquel (2011). Long-Run Effects of Public Sector Sponsored Training in West Germany. *Journal of the European Economic Association 9*(4), 742–784.

Lee, L.-F. (1982). Some Approaches to the correction of selectivity bias. *The Review of Economic Studies 49*(3), 355–377.

Liebman, J. and N. Mahoney (2013). Do expiring budgets lead to wasteful year-end spending? Evidence from federal procurement. *NBER Working Paper 19481.*

McPherson, M. (2007). An analysis of year-end spending and the feasibility of a carryover incentive for federal agencies. *Master dissertation. Naval Postgraduate School.*

Merchant, K. A. (1985). Budgeting and the propensity to create budgetary slack. *Accounting, Organizations and Society 10*(2), 201–210.

Osikominu, A. (2013). Quick job entry or long-tern human capital development? The dynamic effects of alternative training schemes. *Review of Economic Studies 80*(1), 313–342.

Rivers, D. and Q. Vuong (1988). Limited information estimators and exogeneity test for simultaneous probit models. *Journal of Econometrics 39*(3), 347–366.

Roy, A. (1951). Some thoughts on the distribution of earnings. *Oxford Economic Papers 3*(2), 135–146.

Rubin, D. (1974). Estimating Causal Effects of Treatment in Randomized and Nonrandomized Studies. *Journal of Educational Psychology 66*(5), 688–701.

Schneider (1981). Erfahrungen mit "41a". *Arbeit und Beruf 4*(1981), 97–99.

Sianesi, B. (2004). An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s. *Review of Economics and Statistics 86*(1), 133–155.

Sianesi, B. (2008). Differential effects of active labour market programs for the unemployed. *Labour Economics 15*(3), 370–399.

Terza, J. (2009). Parametric nonlinear regression with endogeneous switching. *Econometric Reviews 28*(6), 555–580.

Terza, J., A. Basu, and P. Rathouz (2008). Two-stage residual inclusion estimation: addressing endogeneity in health econometric modeling. *Journal of Health Economics 27*(3), 531–543.

Wooldridge, J. (2005). Unobserved heterogeneity and estimation of average partial effects. *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D.W.K. Andrews and J.H. Stock. Cambridge, U.K.: Cambridge University Press, 27–55.

Wooldridge, J. (2014). Quasi-maximum likelihood estimation and testing for nonlinear models with endogenous explanatory variables. *Journal of Econometrics 182*(1), 226–234.

Wunsch, C. and M. Lechner (2008). What did all the money do? On the general ineffectiveness of recent West German labour market programmes. *Kyklos 61*(1), 133–174.

Figures and tables

Figure 1: Regional employment offices

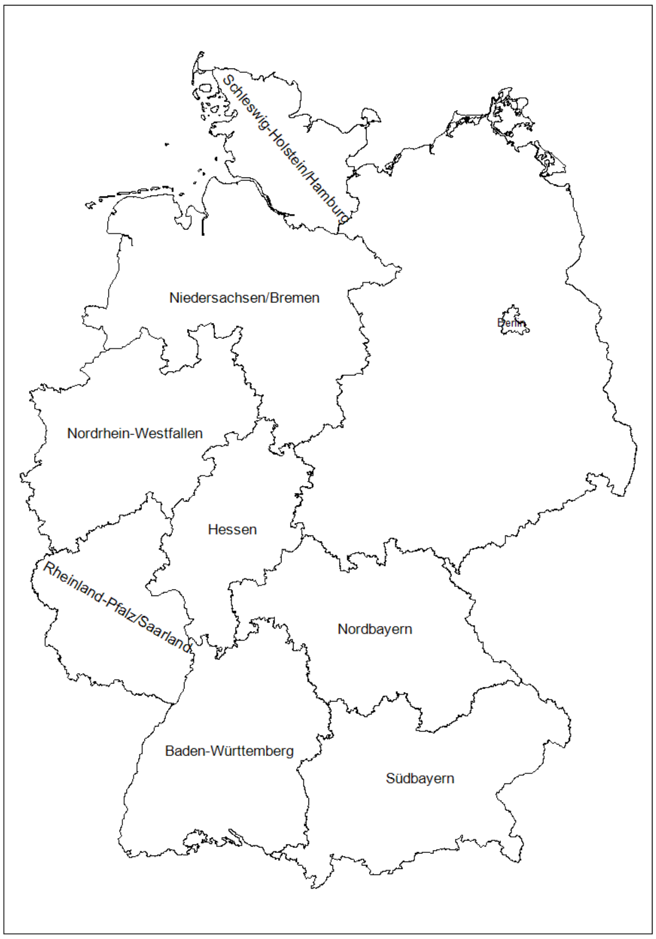


Figure 2: The employment office's budget year and the construction of budget surplus

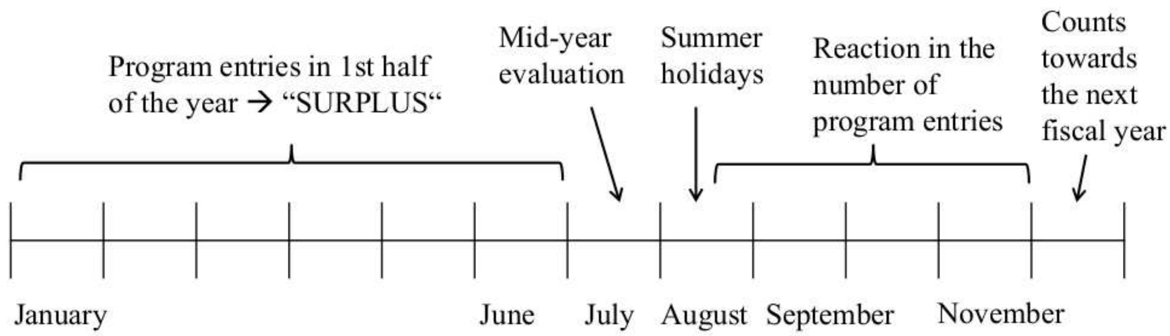
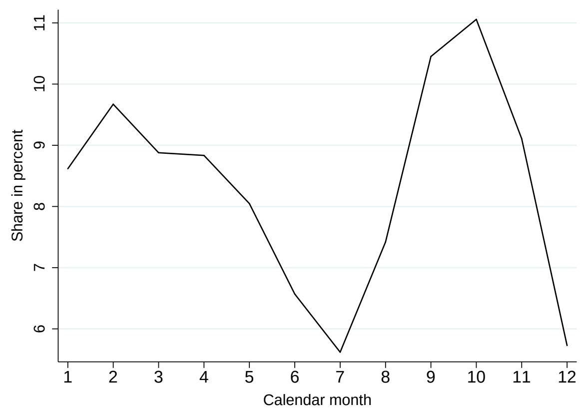
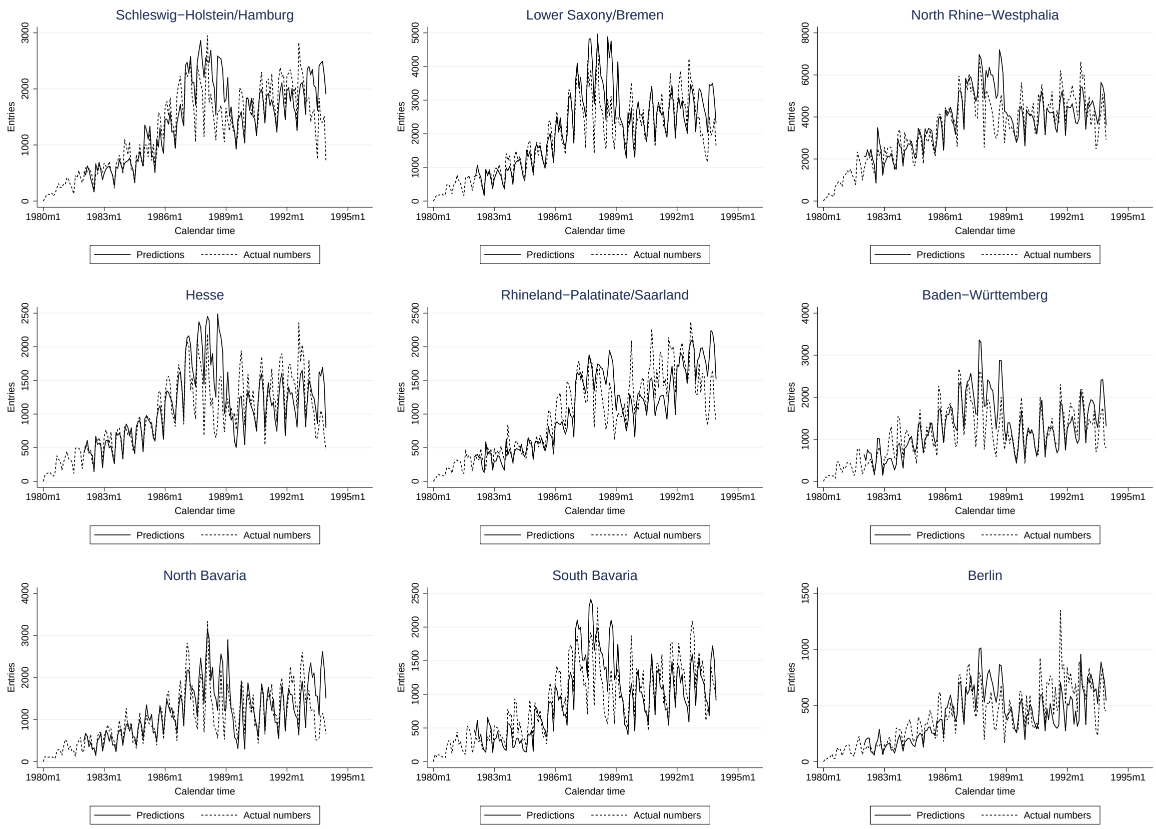


Figure 3: Monthly shares of total year entries into training programs – 1982 to 1993



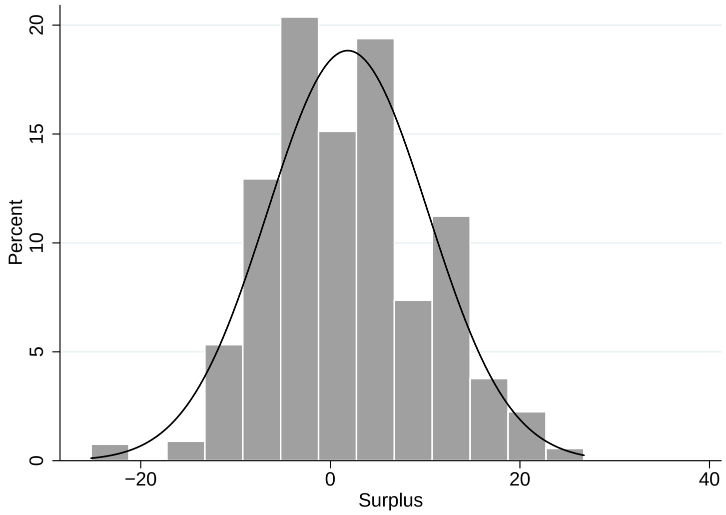
Note: Shares are calculated on the basis of entry numbers averaged over region and time.

Figure 4: Actual and predicted number of entrants in training programs by West German regions – 1980 to 1993



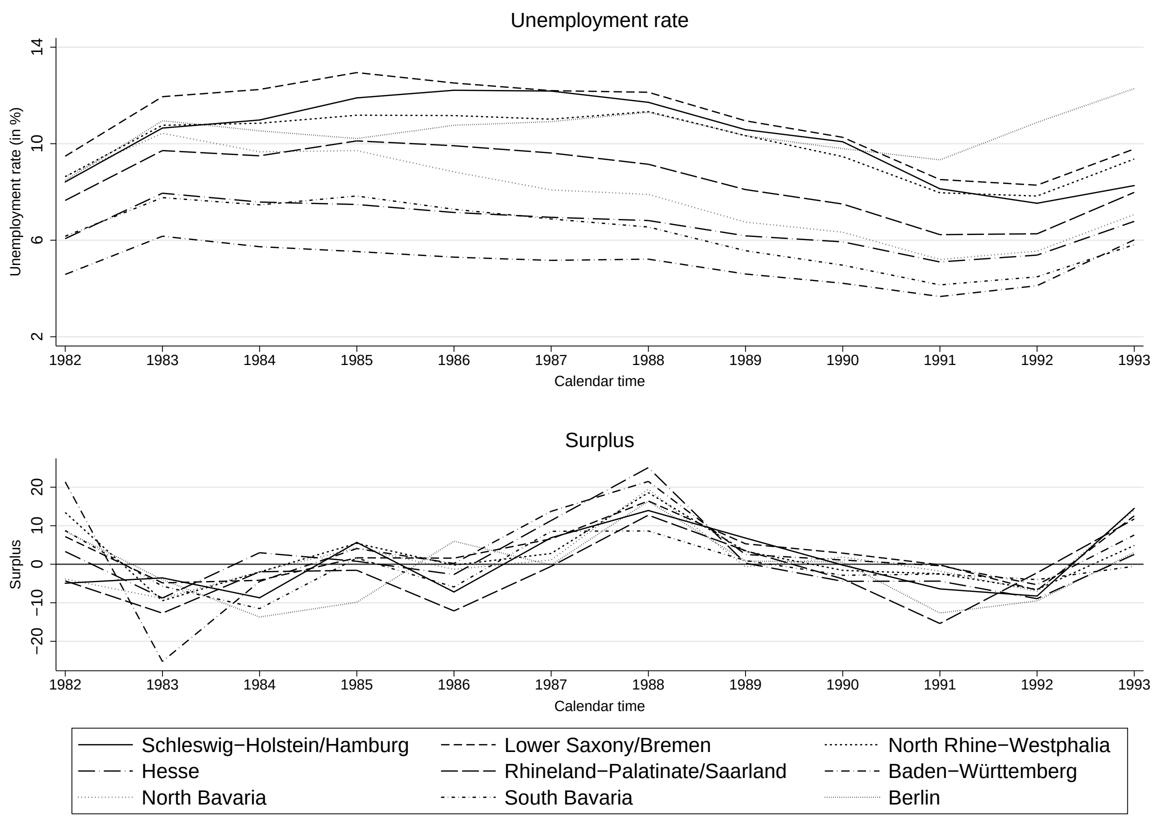
Note: Predicted numbers are out-of-sample predictions calculated by a regression of entry numbers on a set of the year and calendar month dummies, period number, and the monthly entries of the previous year.

Figure 5: Distribution of the budget surplus



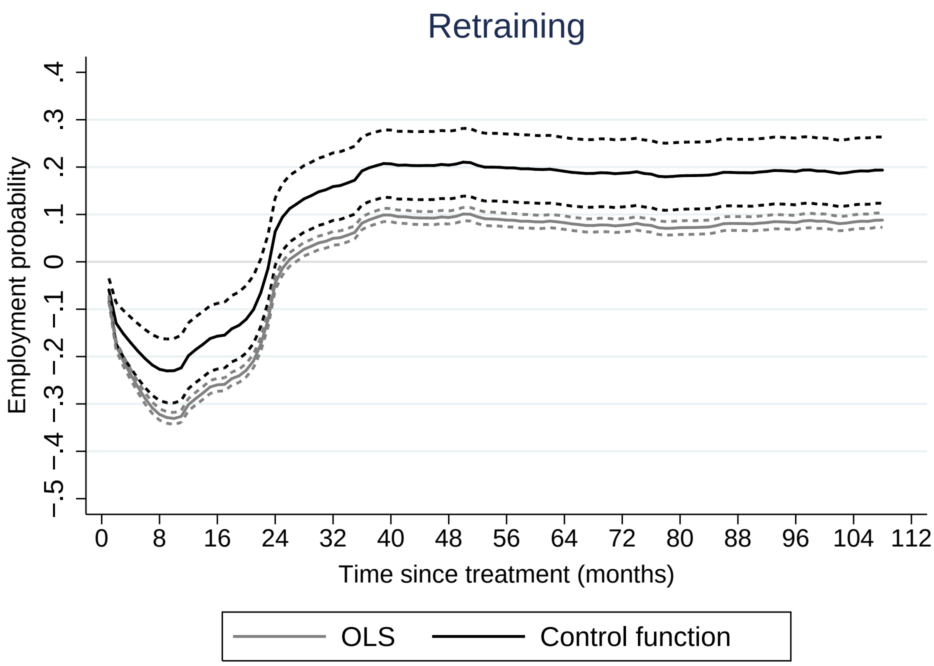
Note: Budget surplus is defined as the difference between planned and actual entries per 1000 eligible unemployed cumulated over the first half of the year.

Figure 6: Unemployment rate and budget surplus by region – 1982 to 1993



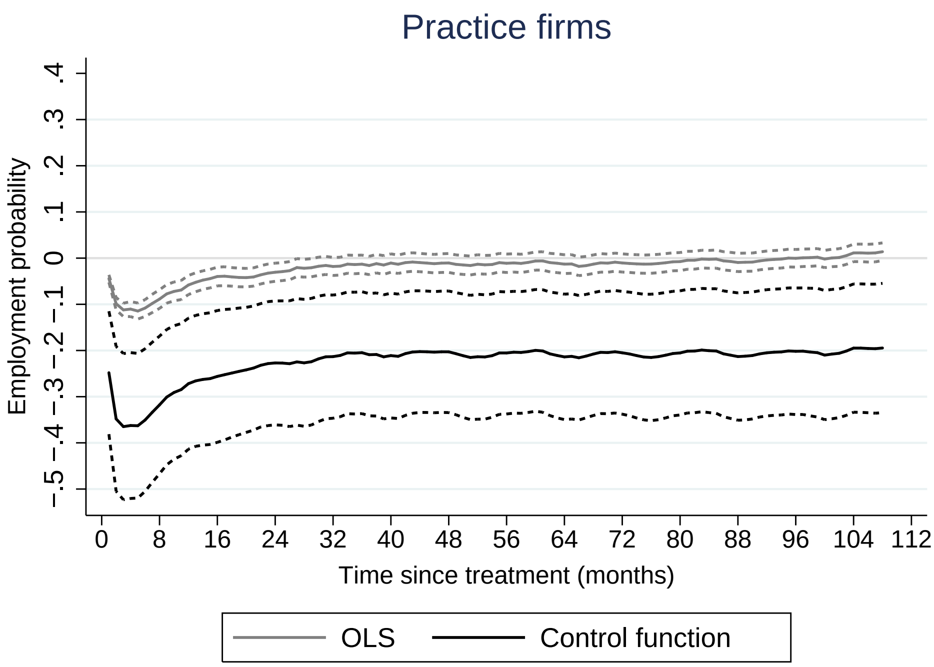
Note: The monthly unemployment rate is averaged over the first half of the year.

Figure 7: Estimated employment effects (ATTs) of retraining pooled over strata



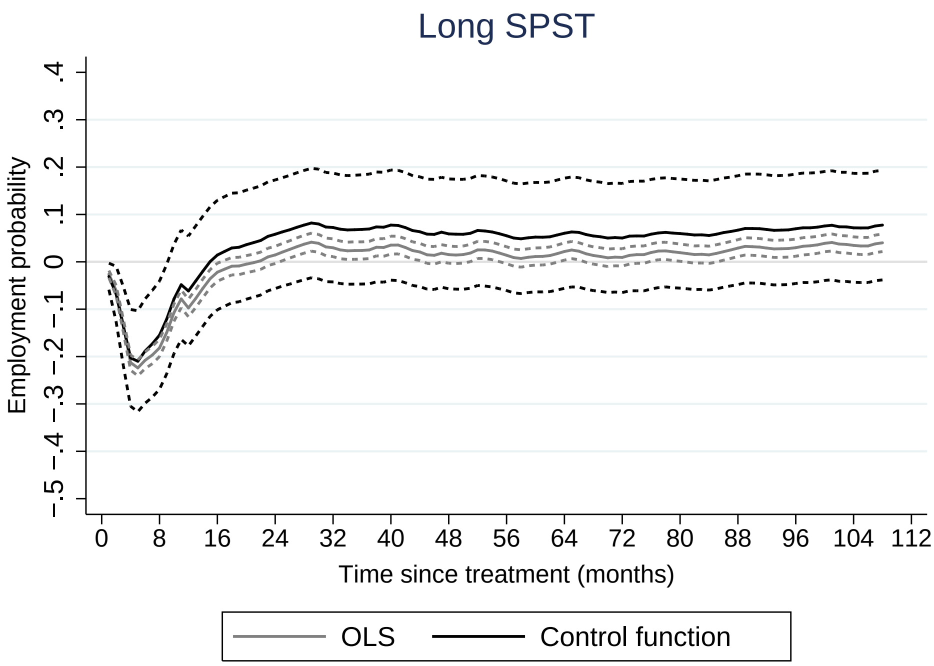
Note: Estimated effects are aggregated over the four quarters of elapsed unemployment duration weighted by the fraction of program participants in the respective stratum. 95%- confidence intervals (dashed lines) are obtained through cluster bootstrapping on 250 replications.

Figure 8: Estimated employment effects (ATTs) of practice firms pooled over strata



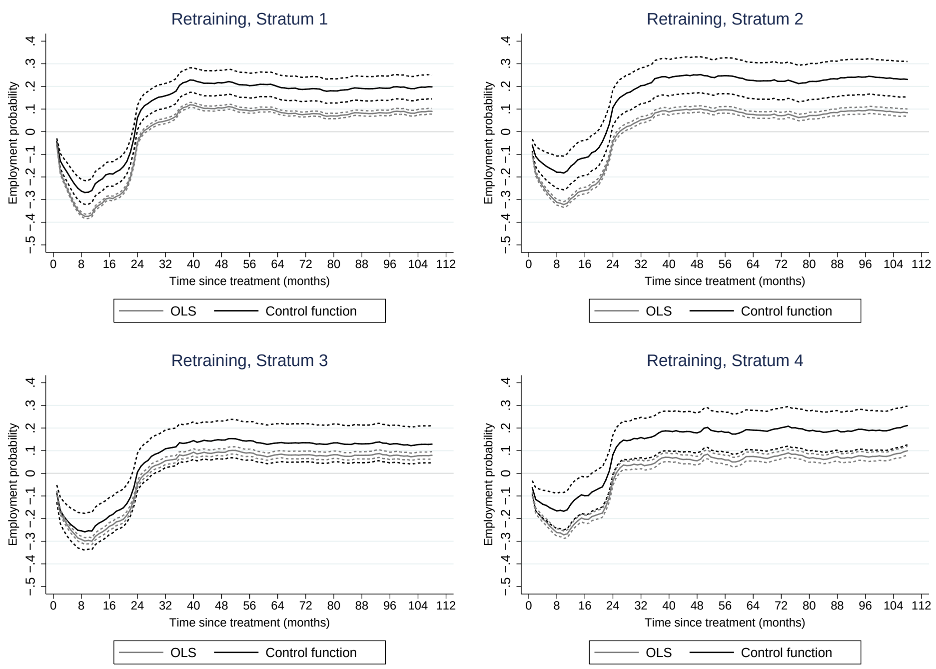
Note: Estimated effects are aggregated over the four quarters of elapsed unemployment duration weighted by the fraction of program participants in the respective stratum. 95%- confidence intervals (dashed lines) are obtained through cluster bootstrapping on 250 replications.

Figure 9: Estimated employment effects (ATTs) of long SPST pooled over strata



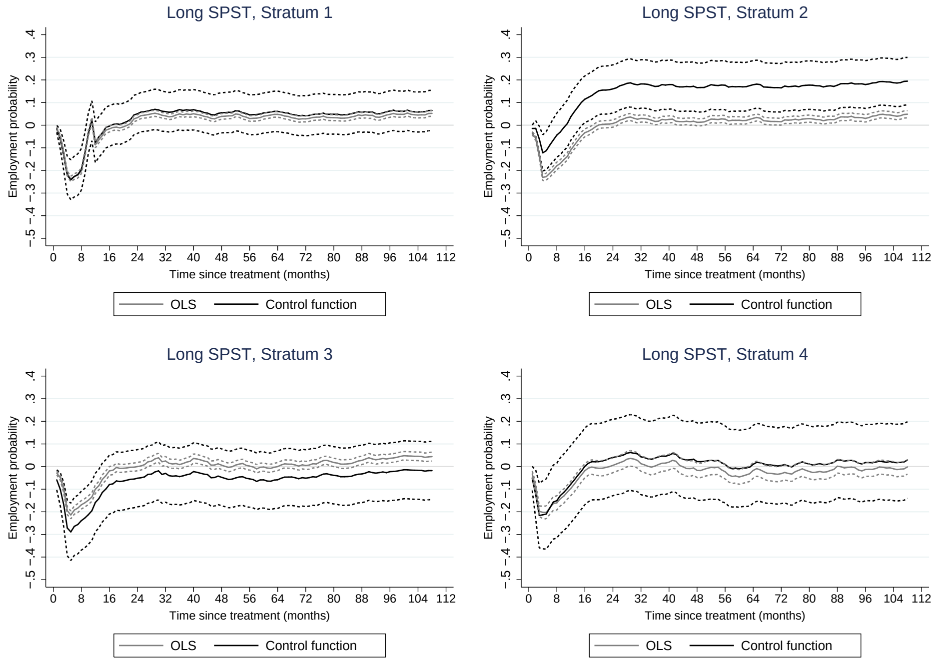
Note: Long SPST training includes programs providing specific professional skills and techniques with a planned duration of more than 6 months. Estimated effects are aggregated over the four quarters of elapsed unemployment duration weighted by the fraction of program participants in the respective stratum. 95%-confidence intervals (dashed lines) are obtained through cluster bootstrapping on 250 replications.

Figure 10: Estimated employment effects (ATTs) of retraining by strata



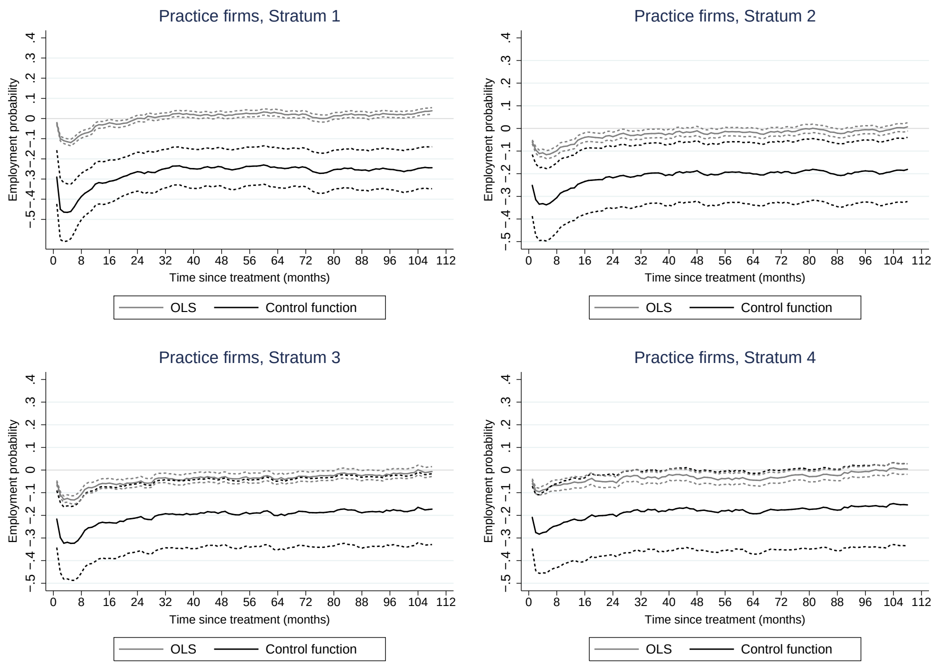
Note: 95%- confidence intervals (dashed lines) are obtained through cluster bootstrapping on 250 replications.

Figure 11: Estimated employment effects (ATTs) of long SPST by strata



Note: Long SPST includes programs providing specific professional skills and techniques with a planned duration of more than 6 months. 95%- confidence intervals (dashed lines) are obtained through cluster bootstrapping on 250 replications.

Figure 12: Estimated employment effects (ATTs) of practice firms by strata



Note: 95%- confidence intervals (dashed lines) are obtained through cluster bootstrapping on 250 replications.

Table 2: List of variables

Variable	Description
	Instrument
Budget surplus	difference between planned and actual participation over the first half of a calendar year per 1 000 unemployed at the level of the regional employment office (REO)
	Personal characteristics
Female	equal to one if female
Age	age years dummies: **25–29**, 30–34, 35–39, 40–44, 45–50
Education	education dummies: no vocational training degree, **vocational training degree**, uni/college degree, education unknown
Nationality	equal to one if foreigner
Marital status	equal to one if married
Children	equal to one if at least one child in household
	Previous employment and former treatment participation
Months employed	dummies for being employed in month M (M=6, 12, 24) before current unemployment
Firm size	dummies for < **10 employees**, ≥ 10 and <20 employees, ≥ 20 and < 50 employees, ≥ 50 and < 200 employees, ≥ 200 and < 500 employees, ≥ 500 employees/missing
Employment status	dummies for apprentice, **blue collar worker**, white collar worker, worker at home/missing, part-time worker
Industry	dummies for agriculture, basic materials, metal/vehicles/electronics, light industry, construction, **production oriented services/trade/banking**, consumer oriented, organizational and social services/missing
Occupation	dummies for farmer and fisher, **manufacturing occupations**, technicians, service occupation, miners/others/missing
Wage	log of daily earnings
Former participant	dummies for participation in any ALMP program reported in our data in year(s) Y (Y=1,2) before current unemployment
	Regional characteristics
Region	REO dummies: **Schleswig-Holstein and Hamburg (SHH)**, Lower Saxony and Bremen (NB), North Rhine-Westphalia (NW), Hesse (HE), Rhineland-Palatinate and Saarland (RPS), Baden-Württemberg (BW), North Bavaria (NBY), South Bavaria (SBY), Berlin (BE)
Summer vacation	Share of the vacation days in August and September at the regional level
	Time specific variables and unemployment duration
Year	calendar year dummies for the time period from 1982-**1993**
Month	calendar month dummies for the months August-**November**
Elapsed unemployment duration	dummies for 1–3 months (stratum 1), 4–6 months (stratum 2), 7–9 months (stratum 3), **10–12 months (stratum 4)**
Cohort	unemployment cohort (259, ..., 405)
	Interactions of variables
Vacation × region	Combination of the vacation share and region
Vacation × calendar month	Combination of the vacation share and the month dummy (August)

Note: Variables in bold are the omitted category in the empirical analysis.

Table 3: Estimates of the first stage treatment probability (evaluation sample)

Evaluated program	(1) RT	(2) PF	(3) SPST	(4) STT	(5) WS	(6) Short SPST	(7) Long SPST
Surplus	0.1724***	0.2151***	0.0645	-0.1081*	0.0807	-0.0708	0.2121***
	(0.0574)	(0.0630)	(0.0433)	(0.0637)	(0.0552)	(0.0568)	(0.0714)
F-statistic	10.17	11.67	2.26	2.97	2.14	1.62	12.98
No. of clusters	89 989	73 621	112 708	67 444	69 994	84 327	80 877

Note: Average partial effects per 1000 unemployed (in percentage points) obtained from a first stage probit regression of the probability to participate in retraining (RT), practice firms (PF), specific professional skills and techniques (SPST), short-term training (STT), wage subsidies (WS), short SPST, and long SPST. Standard errors are clustered at individual level and reported in parentheses. ***, **, and * indicate statistical significance at 1%, 5%, and 10% level, respectively.

Table 4: Size of treatment and control groups in the evaluation sample by elapsed unemployment duration and evaluated program

Elapsed duration	1–3 months	4–6 months	7–9 months	10–12 months
Retraining (RT)				
Treatment group	11 505	6 177	5 389	3 978
Control group	44 299	25 891	19 569	14 110
Practice firms (PF)				
Treatment group	3 183	2 486	2 198	1 809
Control group	42 356	25 357	19 209	14 027
Long specific professional skills and techniques (long SPST)				
Treatment group	6 192	4 562	3 446	2 538
Control group	43 686	26 042	19 462	14 060

Table 5: Summary statistics for selected explanatory variables – evaluation sample

	Retraining (RT)				Practice firms (PF)			
	Treated		Control		Treated		Control	
Variable	Mean	St.dev	Mean	St.dev	Mean	St.dev	Mean	St.dev
Instrument								
Budget surplus	1.9043	8.6118	1.8096	8.6513	2.5167	8.3959	1.8334	8.6803
Female and age								
Female	0.4247	0.4943	0.3935	0.4885	0.3228	0.4676	0.3937	0.4886
25-29 years old	0.4156	0.4928	0.2714	0.4447	0.3000	0.4583	0.2714	0.4447
30-34 years old	0.2832	0.4506	0.2217	0.4154	0.2252	0.4177	0.2216	0.4153
35-39 years old	0.1614	0.3679	0.1764	0.3811	0.1840	0.3875	0.1797	0.3839
40-44 years old	0.0945	0.2925	0.1578	0.3645	0.1533	0.3603	0.1548	0.3617
45-50 years old	0.0453	0.2079	0.1727	0.3780	0.1376	0.3445	0.1725	0.3779
Education								
No voc. training degree	0.1994	0.3995	0.2102	0.4075	0.2111	0.4081	0.1997	0.3998
Voc. training degree	0.7142	0.4518	0.7077	0.4548	0.7631	0.4252	0.7151	0.4514
Uni/college degree	0.0837	0.2770	0.0772	0.2670	0.0225	0.1484	0.0793	0.2702
Education unknown	0.0027	0.0523	0.0048	0.0694	0.0032	0.0565	0.0059	0.0766
Marital status, children in household, foreigner								
Married	0.4075	0.4914	0.4689	0.4990	0.4471	0.4972	0.4780	0.4995
Kids	0.2984	0.4576	0.3012	0.4588	0.3295	0.4700	0.3089	0.4620
Foreigner	0.0740	0.2618	0.1236	0.3291	0.0949	0.2931	0.1207	0.3257
Previous employment and employment status								
# months employed	18.867	6.2413	18.925	6.2325	18.271	6.5721	18.946	6.2198
Log wage	3.8375	0.7028	3.8023	0.7706	3.8572	0.5884	3.8024	0.7601
Apprentice	0.0190	0.1364	0.0127	0.1121	0.0103	0.1011	0.0123	0.1103
Blue collar worker	0.5640	0.4959	0.5808	0.4934	0.6359	0.4812	0.5825	0.4932
White collar worker	0.3271	0.4692	0.3088	0.4620	0.2868	0.4523	0.3117	0.4632
Worker at home	0.0016	0.0401	0.0033	0.0575	0.0009	0.0305	0.0033	0.0572
Part-time worker	0.0883	0.2837	0.0944	0.2923	0.0660	0.2484	0.0903	0.2866
Occupation from previous employment								
Farmer/Fisher	0.0214	0.1447	0.0278	0.1645	0.0375	0.1900	0.0285	0.1665
Manufacturing	0.3434	0.4748	0.3764	0.4845	0.4174	0.4932	0.3799	0.4854
Technicians	0.0740	0.2618	0.0373	0.1895	0.0179	0.1325	0.0381	0.1914
Service	0.5843	0.4929	0.5282	0.4992	0.4753	0.4994	0.5237	0.4994
Miners/Others/Missing	0.0208	0.1427	0.0305	0.1720	0.0519	0.2218	0.0298	0.1701
Former treatment participation before current unemployment								
1 year before	0.0372	0.1893	0.0329	0.1783	0.0632	0.2434	0.0328	0.1782
2 years before	0.0549	0.2278	0.0556	0.2291	0.0993	0.2991	0.0562	0.2302

<continued on next page>

Table 5 – <continued from previous page>

	Short SPST				Long SPST			
	Treated		Control		Treated		Control	
Variable	Mean	St.dev	Mean	St.dev	Mean	St.dev	Mean	St.dev
Instrument								
Budget surplus	1.3305	8.6692	1.7916	8.5973	1.5897	8.3985	1.8274	8.7300
Female and age								
Female	0.4681	0.4990	0.3952	0.4889	0.5311	0.4990	0.3885	0.4874
25-29 years old	0.2839	0.4509	0.2741	0.4461	0.2632	0.4404	0.2693	0.4436
30-34 years old	0.2289	0.4201	0.2246	0.4173	0.2535	0.4350	0.2231	0.4163
35-39 years old	0.1838	0.3873	0.1713	0.3768	0.2016	0.4012	0.1804	0.3845
40-44 years old	0.1617	0.3682	0.1538	0.3607	0.1590	0.3657	0.1580	0.3647
45-50 years old	0.1426	0.3497	0.1762	0.3810	0.1253	0.3311	0.1693	0.3750
Education								
No voc. training degree	0.1138	0.3176	0.2026	0.4020	0.1128	0.3164	0.2004	0.4003
Voc. training degree	0.8009	0.3993	0.7116	0.4530	0.7423	0.4374	0.7153	0.4513
Uni/College degree	0.0806	0.2723	0.0785	0.2689	0.1439	0.3510	0.0793	0.2702
Education unknown	0.0032	0.0561	0.0073	0.0853	0.0026	0.0512	0.0050	0.0706
Marital status, children in household, foreigner								
Married	0.4751	0.4994	0.4751	0.4994	0.4715	0.4992	0.4700	0.4991
Kids	0.3181	0.4657	0.3079	0.4616	0.3566	0.4790	0.3011	0.4587
Foreigner	0.0721	0.2586	0.1210	0.3262	0.0523	0.2227	0.1252	0.3309
Previous employment and employment status								
# months employed	19.423	6.1011	19.034	6.1757	19.006	6.3366	18.993	6.2434
Log wage	3.8351	0.8163	3.7983	0.7751	3.8457	0.7617	3.7922	0.7932
Apprentice	0.0122	0.1096	0.0143	0.1187	0.0121	0.1095	0.0134	0.1150
Blue collar worker	0.4004	0.4900	0.5780	0.4939	0.3253	0.4685	0.5786	0.4938
White collar worker	0.4893	0.4999	0.3129	0.4637	0.5413	0.4983	0.3142	0.4642
Worker at home	0.0007	0.0272	0.0030	0.0545	0.0013	0.0362	0.0040	0.0634
Part-time worker	0.0980	0.2973	0.0918	0.2888	0.1240	0.3296	0.0898	0.2859
Occupation from previous employment								
Farmer/Fisher	0.0137	0.1162	0.0281	0.1652	0.0180	0.1329	0.0298	0.1700
Manufacturing	0.2722	0.4451	0.3794	0.4852	0.2217	0.4154	0.3812	0.4857
Technicians	0.0727	0.2596	0.0384	0.1921	0.0740	0.2618	0.0373	0.1895
Service	0.6275	0.4835	0.5250	0.4994	0.6708	0.4699	0.5225	0.4995
Miners/Others/Missing	0.0133	0.1148	0.0291	0.1681	0.0154	0.1232	0.0292	0.1684
Former treatment participation before current unemployment								
1 year before	0.0518	0.2216	0.0334	0.1798	0.0583	0.2343	0.0337	0.1805
2 years before	0.0795	0.2705	0.0568	0.2315	0.0863	0.2209	0.0560	0.2300

Note: Mean and standard deviation for selected explanatory variables are reported over the first 12 months of unemployment. The calculation is based on weighted individual observations where each individual is replicated based on the number of months he/she remains unemployed and is eligible for treatment. Short SPST training includes programs providing specific professional skills and techniques with a maximum planned duration of 6 months. Long SPST training includes programs providing specific professional skills and techniques with a planned duration of more than 6 months.

Table 6: First stage probit results of the effect of surplus on treatment probability

Specification	(1)	(2)	(3)	**(4)**	(5)[a]
			Retraining (RT)		
APE	.0078	.0370***	.0432***	**.0473*** **	.0461***
	(.0069)	(.0132)	(.0139)	**(.0139)**	(.0137)
F-Statistic	1.29	7.84	9.65	**11.91**	11.53
			Practice firms (PF)		
APE	.0340***	.0585***	.0395***	**.0415*** **	.0389***
	(.0044)	(.0089)	(.0103)	**(.0144)**	(.0138)
F-Statistic	59.58	43.17	14.90	**16.52**	14.96
			Specific professional skills and techniques (SPST)		
APE	-.1494***	.0600***	.0276	**.0253**	.0265
	(.0085)	(.0168)	(.0177)	**(.0173)**	(.0172)
F-Statistic	305.20	12.78	2.42	**2.14**	2.37
			Short-term training (STT)		
APE	-.0763***	-.0897***	-.0419***	**-.0342*** **	-.0219*
	(.0056)	(.0105)	(.0126)	**(.0112)**	(.0112)
F-Statistic	184.59	69.15	10.87	**9.29**	3.84
			Wage subsidies (WS)		
APE	.0210***	.0103	.0108	**.0110**	.0129
	(.0034)	(.0068)	(.0077)	**(.0083)**	(.0085)
F-Statistic	38.65	2.27	1.95	**2.03**	2.84
			Short specific professional skills and techniques (short SPST)		
APE	-.0482***	-.0182	-.0191	**-.0211* **	-.0142
	(.0062)	(.0122)	(.0128)	**(.0127)**	(.0126)
F-Statistic	59.82	2.24	2.23	**2.77**	1.27
			Long specific professional skills and techniques (long SPST)		
APE	-.0201***	.0289**	.0437***	**.0445*** **	.0426***
	(.0057)	(.0117)	(.0132)	**(.0134)**	(.0132)
F-Statistic	12.57	6.09	11.07	**11.74**	11.07
Time specific covariates[b]	no	yes	yes	**yes**	yes
Regional information[c]	no	no	yes	**yes**	yes
Individual characteristics[d]	no	no	no	**yes**	yes

Note: Reported results are obtained on the entire sample. The benchmark specification used in the empirical analysis is in bold. Average partial effects (APE) are reported per 1000 unemployed in percentage points. [a]Surplus without correction for changes in intended spending. [b]Time variables include the year and calendar month dummies. [c]Regional information contains region dummies, share of summer holidays at the regional level, and interactions between these variables. [d]We control for personal characteristics, information on previous employment, former treatment participation, and elapsed unemployment duration. ***, **, and * indicate statistical significance at 1%, 5%, and 10% level, respectively.

Table 7: Employment effects estimated by different methods

Year since treatment start	OLS	IV	Control function
Retraining (RT)			
First year	-.2499 (.0029)***	-.2567 (.0177)***	-.1761 (.0172)***
Second year	-.2180 (.0038)***	-.2636 (.0278)***	-.1122 (.0223)***
Years 3-9	.0777 (.0037)***	.1941 (.0252)***	.1864 (.0222)***
Practice firms (PF)			
First year	-.0843 (.0044)***	-.3114 (.0499)***	-.3142 (.0488)***
Second year	-.0407 (.0056)***	-.1695 (.0627)***	-.2466 (.0455)***
Years 3-9	-.0088 (.0046) *	-.1686 (.0562)***	-.2077 (.0443)***
Long specific professional skills and techniques (long SPST)			
First year	-.1310 (.0041)***	-.0392 (.0120)***	-.1157 (.0293)***
Second year	-.0170 (.0055)***	-.0310 (.0164)**	.0225 (.0361)
Years 3-9	.0238 (.0050)***	.0318 (.0146)***	.0643 (.0363)*

Note: ***, **, and * indicate statistical significance at 1%, 5%, and 10% level, respectively. Standard errors are obtained through cluster bootstrapping based on 250 replications. Results in column (3) are obtained from a 2SLS procedure, which uses the fitted participation probabilities as the instrument for treatment. The IV parameter is estimated on average over the respective months, whereas results in the remaining columns refer to averages of the month-specific effects.

A Additional appendix

A.1 Further information on the construction of the data set

All subsamples used in this study were drawn according to the so called "birthday concept". That is, in the subsamples both from the FuU data and from the IEB data, 50% of all possible birthdays starting with January, 2nd, are drawn and all observations with those 182 birthdays included. The 3% IEB subsample was obtained in the same way, just that here only 12 of the 182 birthdays chosen above are considered and all records that have already been drawn before are dropped.

The combined raw data had a spell form and contained a lot of temporal overlaps. We carried out a number of corrections, mostly based on Bender et al. (2005), in order to improve data quality and prepare the data for the empirical analysis. The most important data preparation steps involved extending the FuU data with information from IEB. The merge procedure was based on a personal identification number and additional criteria like consistency in time structure and contents of the corresponding spells. For all data sources, we adjusted the temporal overlaps between the different types of spells, corrected the education variable according to imputation rules developed by Fitzenberger et al. (2006), and generated the data on a monthly basis.

For the empirical analysis, we weight the observations depending on the data source and on the treatment status in the given evaluation window (stratum s) based on elapsed unemployment duration. Individuals in the control group who never participate in a training program (i.e., those from the 3% IEB subsample) receive a weight of 182/12, corresponding to the birthday concept as described above. The same applies for treated individuals if the program starts in stratum s. On the other hand, individuals from the 50% FuU or the 50% IEB sample receive a weight of 1 in the stratum s if participation takes place later.

For reasons of the computing power constraints, we reduce the control group by drawing a 10% random subsample conditional on the type of training we analyze. More specifically, the 10% subsample is drawn from a pool of control persons who do not participate in the evaluated program P within the 12 months after becoming unemployed, while individuals participating in program P within the first 12 months always stay in the control group for this program. We adjust the weights for these observations by multiplying the weights with the probability of 1/10.

A.2 Formal description of estimation of ATT

For the estimation of a random coefficients model with both a binary endogenous treatment and a binary outcome, we adopt a flexible CF approach for nonlinear models with discrete explanatory endogenous variables as described in Wooldridge (2014, section 6).[29] The idea is that a control function derived from a variable addition test for endogeneity can be used in a flexible way in a one-step or two-step quasi-maximum likelihood framework to identify and estimate treatment effects based on the estimation of the average structural function as introduced by Blundell and Powell (2003).[30] For computational simplicity, we opt for the two-step control function approach.

Following Wooldridge (2014), we maintain the following assumptions for identification:

(A1) $E\left[y_t \mid z, d, b_t^d, u_t\right] = E\left\{\mathbb{1}\left[a_{t0} + z_1 b_0 + (b_{t0} + z_1(b_1 - b_0))d + b_t^d d + u_t \geq 0\right] \mid z, d, b_t^d, u_t\right\}$,

(A2) $E\left[y_t \mid d, z_1, b_t^d, u_t, e_2\right] = E\left[y_t \mid d, z_1, b_t^d, u_t\right]$, and

(A3) $D(b_t^d, u_t \mid z, d) = D(b_t^d, u_t \mid e_2)$.

All variables are defined in section 5, and we maintain all further assumptions made there. $D(. \mid .)$ denotes the conditional probability law.

Assumption **(A1)** specifies the structural expectation as a probit response function with scaled coefficients (Wooldridge, 2005, 2014). Without further assumptions, treatment effects are not identified from the conditional expectation function since the outcome variabel y_t does not only depend upon observed characteristics but also on unobserved heterogeneity (b_t^d, u_t) which we allow to depend upon the treatment variable via ν.

Assumption **(A2)** is an ignorability condition on the control function in the structural conditional expectation and essentially holds by the definition of e_2. It means that once observed and unobserved factors are controlled for in the response function, proxies for observed and unobserved heterogeneity are redundant for y_t. Under the assumption that selection into treatment can be described by a probit model, a natural choice for the control function is using the generalized error gr of the probit model (gr involves the

[29]The approach builds on earlier work by Wooldridge (2005) and by Heckman (1978), Lee (1982), Rivers and Vuong (1988), Blundell and Powell (2003), and Terza et al. (2008).

[30]Similarly, Terza et al. (2008) suggest a computationally simple "two-stage residual inclusion" approach in a parametric nonlinear regression framework where the residuals from a first stage regression for an endogenous treatment dummy can be used as a control function. However, the first residual of the treatment dummy is not independent of the exogenous regressor. Terza (2009) suggests a computationally more expensive estimation approach which relies on correctly integrating out the control function, i.e. the distribution of the unobserved heterogeneity term given the endogenous treatment dummy, in a nonlinear regression specification.

standard Heckman (1978, 1979) selection correction term as defined in section 5).[31] Note that gr is determined at treatment start and does not change over time for each individual. However, the control function e_2 may vary over time because b_t^d, u_t can change over time.

Assumption **(A3)** imposes strong ignorability restrictions on the conditional distribution of unobserved heterogeneity. Thus, conditioning on e_2 in the structural expectation is sufficient to correct for selectivity bias arising from the endogeneity of treatment (Wooldridge, 2014).[32] Note that the impact of the selection correction term is not nonparametrically identified because the sign of the generalized residual gr is perfectly collinear with the treatment dummy (see Wooldridge, 2014, section 6.3). This is in contrast to the case of a continuous endogenous regressor as discussed in Blundell and Powell (2003). Wooldridge (2014) suggests adding the square of gr, interactions between gr and the treatment dummy and between gr and the observed characteristics z_1 to the vector of control functions.[33]

Under assumptions **(A1)**-**(A3)** and by the law of iterated expectations, the average structural function at time t among the treated $d = 1$ can be expressed as[34]

$$ASF(\tilde{d}, z_1, d = 1) = \tag{8}$$

$$E_{b_t^d, u_t | d=1, z_1} \left\{ \mathbb{1} \left[a_{t0} + z_1 b_0 + (b_{t0} + z_1(b_1 - b_0))\tilde{d} + b_t^d \tilde{d} + u_t \geq 0 \right] \right\} .$$

where $E_\xi [\cdot]$ indicates expectation with respect to the distribution of ξ.

As suggested by Wooldridge (2014, section 6.4), we use the following flexible set of control functions $\hat{e}_2(d_i, z_i) = (\widehat{gr}_i, \widehat{gr}_i^2, \widehat{gr}_i d_i, \widehat{gr}_i z_{i1})$, which we allow to enter the index function for the employment probit as additional linear regressors (see equation 8). The estimated generalized residual $\widehat{gr}_i$ is based on the estimates from the first stage probit for the treatment dummy d_i. As motivated by Lee (1982), the squared $\widehat{gr}_i^2$ and the interaction

[31] Wooldridge (2014) shows that under correct specification of the probit model for d, a variable addition test for treatment exogeneity based on the generalized errors is asymptotically optimal.

[32] Under independence of error terms and instruments ($(b_t^d, u_t, \nu) \perp\!\!\!\perp z$), assumption **(A3)** holds exactly for continuously distributed first-stage reduced form errors. Since the endogenous regressor in our application is a dummy variable, we should view the ignorability assumption about the conditional distribution of (b_t^d, u_t) only as an approximation for a given vector of control functions.

[33] Another possible extension is based on the assumption that (u_t, ν) are jointly normally distributed. In this case, the vector of proposed control functions consists of three components: gr, gr^2, and the interaction between gr and linear predictions from the first stage probit estimation. The last two terms are derived from the conditional variance of u_t, which in turn is heteroscedastic and varies across individuals (see Kimhi (1999) and Wooldridge (2014)).

[34] See Blundell and Powell (2003) and Wooldridge (2005) for a detailed derivation of the average structural function when the endogenous explanatory variable is continuous.

terms $\widehat{gr}_i z_{i1}$ account for deviations from the joint normality assumptions as imposed in Heckman (1978). The interaction term $\widehat{gr}_i d_i$ accounts for the random coefficient of the treatment dummy in the structural employment equation (2a), see e.g. Blundell et al. (2005, section 3.4.1) for the continuous outcome case.[35]

Under these assumptions, the average structural function can be expressed by integrating out the control functions e_2 as

$$\text{(9)} \qquad ASF(\tilde{d}, z_1, d = 1) =$$

$$E_{e_2|d=1,z_1}\left\{Prob\left(a_{t0} + z_1 b_0 + (b_{t0} + z_1(b_1 - b_0))\tilde{d} + b_t^d \tilde{d} + u_t > 0 \mid d = 1, z_1, e_2\right)\right\} .$$

Equation (9) makes explicit that once the observed conditional expectation of y_t given z, d, e_2 is consistently estimated, which in turn is implied by having sufficient variation in the instrumental variables z_2, identification of the average effect of treatment on the treated is feasible by integrating out the joint distribution of z, e_2 (Wooldridge, 2014) among the treated.

For estimation purposes, we add estimated versions of the control functions e_2 to a second stage probit regression of employment, where we regress y_t on d, z_1, interactions between d and z_1, time effects m_t and interactions between m_t and d, and $\hat{e}_2$. In its most general specification, the estimated regression for observation i ($i = 1, ..., N$) is

$$\text{(10)} \qquad \widehat{Pr}(y_{it} = 1 \mid d_i, z_{i1}, e_{i2}) =$$

$$\Phi\left(m_t\hat{\delta}_{0t} + m_t\hat{\delta}_{1t}d_i + z_1\hat{b}_0 + z_1\hat{\delta}_1 d_i + \hat{\omega}_0\widehat{gr}_i + \hat{\omega}_1\widehat{gr}_i d_i + \hat{\omega}_2\widehat{gr}_i^2 + z_{i1}\widehat{gr}_i\hat{\psi}\right) ,$$

where $\Phi(.)$ is the standard normal distribution function, m_t represents a full set of time dummies, and $\hat{\delta}_{0t}$, $\hat{\delta}_{1t}$, $\hat{b}_0$, $\hat{\delta}_1$, $\hat{\omega}_j$ ($j = 0, 1, 2$), $\hat{\psi}$ are coefficient estimates.

Based on the estimated equation (10), we can estimate the ATT at time t by integrating out the distribution of $z_{i1}, \hat{e}_{i2}$ among the treated $d_i = 1$ as

(11)

$$\widehat{\tau}_{ATT,t} = \frac{1}{N_1}\sum_{d_i=1}\left\{\Phi\left(m_t\hat{\delta}_{0t} + m_t\hat{\delta}_{1t} + z_1\hat{b}_0 + z_1\hat{\delta}_1 + \hat{\omega}_0\widehat{gr}_i + \hat{\omega}_1\widehat{gr}_i + \hat{\omega}_2\widehat{gr}_i^2 + z_{i1}\widehat{gr}_i\hat{\psi}\right)\right.$$

$$\left. - \Phi\left(m_t\hat{\delta}_{0t} + z_1\hat{b}_0 + \hat{\omega}_0\widehat{gr}_i + \hat{\omega}_2\widehat{gr}_i^2 + z_{i1}\widehat{gr}_i\hat{\psi}\right)\right\} .$$

[35]Under joint normality of ν, b_t^d, u_t, the coefficient of gr_i in the control function differs by the sign of d_i because the linear projection of the joint error term $b_t^d d_i + u_t$ on ν differs by d_i.