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: October 2015

Abstract: This study re-estimates the employment effects of training programs for the unemployed using exogenous variation in participation caused by budget rules in West Germany in the 1980s and early 1990s. As funds could not be transferred to the next year or to other labor market instruments, a budget surplus (deficit) after the first half of the year increases (decreases) overall 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 programs improve employment in the medium and long run, our findings imply long-lasting negative effects of short-term training and practice firms.

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

JEL-Classification: J64, J68, H43

* Humboldt University Berlin, IFS, CESifo, IZA, ROA, and ZEW.
** Humboldt University Berlin.
*** University of Freiburg.
**** University of Freiburg and Walter Eucken Institute.

Corresponding Author: Bernd Fitzenberger, Humboldt University Berlin, School of Business and Economics, Spandauer Strasse 1, 10099 Berlin, Germany. E-mail: fitzenbb@hu-berlin.de.

1   Introduction

The effectiveness of training programs for the unemployed is a long-debated issue Card et al. (2010). Proponents argue that they are a good investment, since enhancing the abilities and skills of unemployed individuals would lead to a quicker reintegration into the labor market, thus resulting in a win-win situation for the unemployed, the government, and the employers. Critics claim that the resources spent on training programs are mostly wasted, arguing that the programs themselves do not have any positive impact on later employment and that good actual outcomes may only reflect a positive selection of training participants. Based on a selection on observables identification strategy, the empirical evidence on the employment effect of participation in training is quite ambiguous.[1] Furthermore, any empirical examination has to face the possibility that the assignment to a training program may depend upon characteristics of the unemployed which are not observed by the researcher and which influence the individual's employment chances (selection on unobservables). Effect estimates based on a selection on unobservables identification strategy are rare and the results of such studies are also ambiguous.[2] To investigate the importance of unobservables, Caliendo et al. (2014) use survey data on some variables typically not available in administrative data. The study finds that the variables considered are significant determinants of participation in training but accounting for them leaves the effects estimates basically unchanged. However, it remains an open question as to whether unobservables different from those considered by Caliendo et al. (2014) could play a role. There is little evidence on this issue so far based on instrumental variables (IV) methods (see Frölich and Lechner (2010) for a notable exception).

---

[1]See e.g. the survey in Card et al. (2010) as well as the literature reviews in the recent studies by Biewen et al. (2014), Heinrich et al. (2013), Osikominu (2013), or Richardson and van den Berg (2013). Considering a selection on observables identification strategy, positive employment effects in the medium run and long run are found for instance by Heinrich et al. (2013) for the case of the U.S. and by various studies regarding longer training programs for West Germany in the early 1990's and late 1980's, see Fitzenberger and Speckesser (2007); Fitzenberger et al. (2008); Lechner and Wunsch (2009); Lechner et al. (2011). Insignificant medium-run and long-run effects are found by Wunsch and Lechner (2008) for training programs in Germany during the early 2000's. For the same time period, slightly better employment effects for West Germany are found by Biewen et al. (2014), however the effect estimates are smaller than for the earlier decades.

[2]Again, see (Card et al., 2010). While Richardson and van den Berg (2013) find negative effects for long training programs in Sweden, Osikominu (2013) finds positive long-run effects. Both studies are based on the timing-of-events approach of Abbring and van den Berg (2004) in continuous time, which assumes time-invariant random effects independent of the covariates governing 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. Based on a discrete time model with selection on unobservables, Fitzenberger et al. (2010) find positive employment effects.

This paper contributes to the analysis of the employment effects of training by using an IV approach to control for selection on unobservables. To instrument training, we exploit 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. Specifically, 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 labor market instruments 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 surplus early on than they otherwise would have done. Thus, we use the variation in the remaining funds during the second semester of a calendar year to instrument the individual participation in a training program. Doing so allows us to reduce the potential influence of selection of participants and program and to come closer towards estimating the causal effect of further training on the future employment prospects of participants.

For the empirical analysis, we implement our IV strategy using the control function method for a random-coefficients model proposed by Wooldridge (2014), which accounts in a flexible way for the discrete nature of the outcome variables and which allows us to estimate the average effect of treatment on the treated. Our instrumental variation proves highly significant for participation in a number of important programs. Our main source of data involves administrative records from the German Federal Employment Office (FEO) 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 0.3% 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 reports of the FEO, our rich data allow us both to calculate the size of first-semester deficits or surpluses for each regional employment office (REO) and to follow individual employment histories up to ten years after treatment start.

Our findings imply that a budget surplus in a REO in the first half of the year results in an increase in the probability of an unemployed individual to enter a training program after the summer holidays. Additionally, we observe a shift in the composition of chosen programs, away from the short and cheap measures towards the long and more expensive

2

ones. The reverse holds for people living in regions in which the employment office runs a deficit in the first half of the year. The evidence suggests that spending the entire assigned budget and signaling sustained need for funds was a goal of the employment offices. With respect to the long-run impact of training programs on the employment of participants, we find quite heterogeneous effects. On the one hand, long programs with a strong focus on acquiring new skills increase the chances of participants to find a job afterwards. On the other hand, we find negative effects of shorter programs in which the unemployed revise basic skills for job applications or work in simulated firms in order to maintain their general work skills. These results confirm that the different program types should not be pooled together in empirical analyses as this may disguise their different individual effects.

Our paper proceeds as follows: In section 2, we provide detailed information on active labor market policy in Germany during the time of our study and relate our work to the existing literature on the effectiveness of 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 motivate 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. Further details on the data and on the econometric approach can be found in the additional appendix.

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 public-sponsored training programs (*Fortbildung und Umschulung*) as the most important component at that time.[3] 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.

---

[3]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*).

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 1.7 billion Euros), representing about 11.4% of the annual budget of the FEO at the time.[4]

A detailed description of the types of FuU training programs can be found in Bender et al. (2005) and Fitzenberger et al. (2008). In this paper, we concentrate our analysis on the five most important of these programs: Short-term training (*Kurzzeitmaßnahmen*, STT), Practice Firms (*Übungsfirmen*, PF), Wage Subsidies (*Einarbeitungszuschüsse*, WS), Provision of Specific Professional Skills and Techniques (*Bereitstellung von spezifischen Kenntnissen und beruflichen Fähigkeiten*, SPST), and Retraining (*Umschulung*, RT).[5] In the following, we briefly describe each of these programs in turn, sorted by their average planned duration.

Short-Term Training (STT) programs with an intended duration of no more than six weeks were created in 1979 and focus on hard-to-place and low-skilled individuals. Short-term training courses were intended to inform job seekers about employment options as well as possibilities for participation in more comprehensive programs. Furthermore, STT participants were taught some general labor market relevant skills, including job search assistance, counseling, and communication training. The program may also provide employer contacts. In general, participants did not have to take an exam at the end of the course and did not obtain any official certificate at the end (Schneider, 1981). Due to tight budgets after German unification, 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).

Practice Firms (PF) involve a simulated firm environment where participants practice 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 the participants' ability for

---

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

[5]For the classification of programs, we follow the definitions developed by Fitzenberger and Speckesser (2007).

particular professions. Similar to STT, participants do not receive a certificate, since the program concentrates on exercising existing skills rather than learning new ones.

Wage Subsidies (WS) aim to overcome the disadvantages of the unemployed by reducing the cost of hiring them and preparing them for the labor market through on-the-job training and work experience. The program includes initial skill adaptation training by the employer and continues for around six months, during which the employer is partially compensated through the subsidy.

Specific Professional Skills and Techniques (SPST) programs focus on providing more specific skills like computer or accounting courses. The goal of SPST is to facilitate the reintegration of unemployed individuals into the labor market by improving their skills and providing signals to potential employees. A completed vocational training degree is usually required to take part in this type of training. The courses focus on classroom training, 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. Due to the wide variety of courses with durations from several months to up to two years, SPST is the most flexible program and represents the largest share among all public-sponsored training programs.

Finally, the longest and most expensive 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 participants in RT already hold a different vocational training degree for a specific occupation but the prospects to find a job in their previous profession is small. RT is also an option for individuals without any vocational degree, provided they meet additional eligibility criteria. RT 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.

To qualify for these training programs, unemployed 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 (Bender et al., 2005; Lechner et al., 2011). With the exception of WS participants, full-time enrolled unemployed receive income maintenance payments throughout the duration of their training.

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- and long-run effects.[6] For training programs starting up to the early 1990s in West Germany, corroborating evidence based on large administrative data can be found e.g. in Fitzenberger et al. (2008) and Lechner et al. (2011).[7]

Following the dynamic evaluation approach proposed by Sianesi (2004, 2008), Fitzenberger et al. (2008) estimate the long-run employment effects of further training programs in a dynamic context conditional on the starting date of the treatment (treatment vs. waiting) and confirm both the negative lock-in effect after program start and significantly positive employment effects in the medium and long run. Lechner et al. (2011) estimate 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 run and positive employment effects in the long run, with retraining exhibiting the largest effect on later employment with about 20 percentage points after eight years.

The aforementioned studies 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 afterwards anyway, comparing the later employment status of participants and non-participants does not allow to estimate 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 endogene-

---

[6]See e.g. Heckman et al. (1999) for a comprehensive survey of the early international literature on the evaluation of various active labor market instruments. Empirical evidence on the effectiveness of training programs from the European countries is extensively reviewed in Kluve (2010).

[7]See e.g. Wunsch and Lechner (2008) and Biewen et al. (2014) for an empirical evaluation of further training programs during the early 2000s and Fitzenberger et al. (2013) for the evaluation of short-term training programs in two different time periods, 1980-1992 and 2000-2003.

ity 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). 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 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 and use a combination of conditional IV and matching methods in order to estimate the average treatment effect of participation for compilers. For Germany, Fitzenberger et al. (2010) use Bayesian techniques in a dynamic framework to 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. In a recent evaluation study, Caliendo et al. (2014) investigate the sensitivity of estimators that only control for observed characteristics to the particular consideration of usually unobserved variables such as personality traits, attitudes, and job search behaviour. Using a unique data set for Germany which combines survey information with large administrative records, the authors find that compared to the inclusion of detailed labor market histories, usually unobserved characteristics are not of primary importance for the estimation of treatment effects, even though such variables prove to be strong predictors for selection into training and the future labor market prospects of participants.

The present paper makes two important contributions to this literature. First, we try to examine the causal effect of training participation using a novel identification strategy. In particular, we exploit the "end-of-year spending" effect in the budget policy of the FEO in the 1980s and early 1990s in West Germany as source of conditional exogenous variation for the treatment probability and use this to instrument participation. Second, we implement our IV strategy using a two-step control function approach that allows us to recover the average treatment effect on the treated and to account for selection issues with respect to both observed and unobserved characteristics in a very flexible way. Furthermore, by extending the period of examination after program participation to up to ten years, we provide evidence on the long-run effects of training programs and

the relative importance of selection based on unobservables. Considering training as an investment, 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 should be recouped by positive effects on employment for several time after the treatment. For the analysis of the overall effectiveness of training, it is interesting to analyze as to whether the potentially positive effects of training programs after the lock-in-period are sustained over time.

3   Data description

In this study, we use a unique new data set combining information from different administrative sources on further training program participation in Germany.[8] The two main components are the *Integrated Employment Biographies* (IEB)[9] and the *FuU* data on program participation. The IEB data are based on administrative *daily* spells reported by employers or the FEO. The data contain employment register data for all employees subject to social insurance contributions for the years from 1975 to the end of 2004. Thus, it provides very long panel data on employment and unemployment periods, on benefit recicpiency from the FEO and on a wide range of personal and job-specific characteristics. Our second data source, the FuU data set consists of *monthly* information about participation in public-sponsored training programs between 1980 and 1997, collected by the FEO for controlling and statistical purposes. While the FuU data set provides further details on participation in training programs and thus enables a precise and detailed identification of training. After merging the two data sources, the identification 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.

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 0.3% subsample from the IEB

---

[8]The data was generated as part 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 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.

[9]In contrast to the standard version of the IEB, our data only contain information from BeH (*Beschäftigten-Historik*) and LeH (*Leistungsempfänger-Historik* of the IAB).

data without any program participation as control group, together with individuals who only entered a program at a later point in time.[10] In each data source, we only consider information starting in 1980 for reasons of data reliability.[11]

The resulting data set is quite informative for two reasons: First, the data involve panel information on employment and unemployment spells from 1980 to the end of 2004. Second, we have data on 50% of all individuals who participated at least once in a training program during the observation period between 1980 and 1997.

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, as we need the earlier cohorts for the construction of the instrument. 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 to issues of data quality.[12] Finally, we only consider individuals aged 25 to 50 at the beginning of the unemployment spell, who live in West Germany.[13]

---

[10]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.

[11]More information on the construction of the data set, subsample, and weighting is provided in the additional appendix on page 53.

[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.

4 Budget surplus as instrument

To estimate the impact of training participation on subsequent employment, we need to address various potential sources for selection bias. First, it may be the case that those unemployed individuals are more likely to 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 a training opportunity to an unemployed, 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 would lead to a positive selection of participants, the opposite would happen if the third source dominated. In all three cases, we may obtain biased estimates of the effect of training participation on employment for any evaluation relying on a selection on observables identifications strategy, if we do not observe all information available to the caseworker.

This paper attempts to come closer to estimate causal effects by exploiting budget rules for active labor market policies in Germany which create a source of exogenous variation in training probabilities. We implement this IV approach using a version of the control function method for a random-coefficients model proposed by Wooldridge (2014). In this section, we first present the institutional background that motivates our instrument. The econometric approach will then 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. Up until 1994, the most important aspects for the purpose of this study were the following: (a) The FEO was organized in three levels, with the central office in Nuremberg, nine regional employment offices (REOs) at the intermediate level, and 142 local employment offices at the lowest organizational level.[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 possessed limited discretion in the use of their

---

[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.

allocated funds for training programs. (c) The budget for training programs was planned and allocated separately from other programs like job creation schemes and could not be transferred to other purposes. (d) The allocation of funds top-down to the regional and to the local offices was based primarily on past levels of program participation, but adjusted for anticipated changes in need for the next year. (e) Unused funds from one fiscal year could not be transferred to the following year.[15]

These budget rules had an influence on the management of training programs for the unemployed by the REOs. 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, a REO could not simply overuse its budget every year to secure a continuous rise of its funding. 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 to assign an additional unemployed into training at the local level. If budgets were almost exhausted, caseworkers may have hesitated more in assigning more participants. If resources were abundant and needed to be deployed, the chance to be assigned to a training program increased.

There is a lot of anecdotal and suggestive evidence for this phenomenon of "end-of-year spending" 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 evidence on its effects due to a lack of reliable data (General Accounting Office, 1998). To our knowledge, the only study is Liebman and 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.

For our analysis, we use a proxy for the difference between planned and actual spending during the first half of the year to instrument program participation in the final months of the year. Planned spending is based on previous experience and on projections for the current year. The timing of events in the budgeting process of the REOs is displayed in

---

[15]This institutional framework remained stable up until 1994, when new rules granted a modest level of budget autonomy to the local employment offices. In 1998, a global budgeting system was introduced leading to a further flexibilization of the budget system at the FEO.

figure 2. The first semester budget review took place in July after all the information regarding the first six months was available. The accumulated budget surplus or deficit during the first semester then influenced the participation decisions in the remaining months of the budget year after the summer holidays. Due to a rotation system for the summer holidays between the German states, the "end-of-year" period in our examination therefore 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 count for the January budget.[16]

This strategy is supported by the following two arguments, both originating from personal interviews and correspondence with FEO experts and practitioners. First, the non-transferability of funds between different years caused caseworkers to set program starts as early as possible in the year in order to ensure that available funds were spent during 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. This is because the programs required a stable group size and an economically viable number of participants. Therefore, the usual point in time for readjustments in program assignments in response to a gap between planned and actual spending during the first semester was after the summer holidays, which ended 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.

The evidence in figure 3 on monthly shares of the total entries into training programs over the year supports this reasoning. There is a pronounced seasonal pattern with many entries during the first semester and a large spike after the summer holidays. In contrast, entry is lowest in June, July, and December. This suggests that caseworkers took both summer holidays as well as the end of the budget year into account. The next section describes the computation of the instrument in detail.

4.2 The derivation of the budget surplus

We follow four 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 number of program entries in each region for every year and month between January 1980 and

---

[16]We treat August as the first month of our examination period for a particular region and year if at most half of the workdays in August belonged to the school holidays, and September otherwise.

December 1993. (2) We predict the intended number of new participants at the regional level for each of the first six months in a year based on past entry patterns. (3) We compute the "budget surplus" for each of the first six months m in year j and region r as follows:

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

The resulting monthly surplus is positive if the planned number of new participants exceeds the actual entries in the respective month and negative if it is the other way round. (4) To obtain the cumulative budget surplus, the monthly numbers for each REO are aggregated over the first six months of each year. 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. One is our use of program entries to proxy the budget. To our knowledge, no information is available on actual *and* planned expenditures for the nine REOs. However, the budget rules at the time in fact involved a direct link between program entries and available funds based on head counts (Bach et al., 1993). Since funds allocated to training programs could not be transferred to other ALMP programs (and vice versa), there was no way to spend these funds for a program different from training. 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.

The second issue is how to get the exact numbers for the respective quantities. Based on our 50% sample of all program participants, we can compute the number of new participants by region and month 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

13

training programs at the federal level. For this purpose, we calculate the ratio between intended spending in a certain year and actual spending in the previous year.[17] A ratio less than one implies a planned 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. Based 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 during the first semester, i.e., the executives on average ran surpluses which is plausible in light of the budget rules discussed above. 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 the program inflow of about one week. There is a large variation in first-semester surpluses across region and year, as indicated by the distribution of surpluses in figure 5. While the vast 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 programs after the summer holidays in order to spend their whole budget, whereas others needed to cut back substantially if they did not want to exceed their funds and get reproached for that.

Apart from general prediction errors regarding actual program entries, the major driving force for the variation in the usage of available funds should be unexpected changes in regional labor market conditions. A sudden improvement in regional business conditions may leave some unplanned open slots in training programs as more potential participants find a job more easily. We expect the reverse effect for an unexpected downturn, when increasing unemployment rates result in higher assignments to training programs. Such unforeseen developments do not have a long-lasting effect on the surplus, however, as the mechanism to determine the budget for the next year would take into account the changes in regional labor market conditions. Thus, we do not observe extended periods of large surpluses or deficits for any regional employment office in our data, but fairly unsystematic fluctuations around the yearly average surplus. This is illustrated in the

---

[17]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.

lower panel of figure 6, which depicts the deviations of surplus for each region and year relative to the average surplus across regions in the corresponding year.

4.3 Plausibility of the identification strategy

Relevance condition

To investigate the reasoning presented above, we first check 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.[18,19] We estimate pooled monthly probit regressions conditional on still being unemployed and not having entered a training program before the month of interest.[20] The resulting average partial effects (APE) based on the probit regressions are reported in table 1.

The estimates show that for three out of the five main types of training programs (short-term training, practice firms, and retraining) the probability to participate in the final months of the year reacts strongly (on the 1% significance level) to our measure of budget surplus during the first semester. In particular, the chances of entering a practice firm or retraining program increase with the size of the surplus, while they decrease for short-term training. This pattern allows for two conclusions: First, there is a positive overall effect of running a first-semester surplus on the probability to enter a training program after the summer holidays, as the positive coefficients for practice firms and retraining overcompensate the negative one for short-term training. And second, surpluses additionally change the composition of program participations towards the longer and more expensive measures, at the cost of the shorter and cheaper one. These findings indicate that officials at the employment offices reacted to the budgetary environment with the aim of spending unused funds by the end of the year (or to meet the budget in the case of deficits).

Contrary to our prior expectations, we do not observe any significant impact of sur-

---

[18]Table A.3 in the appendix provides a description of all explanatory variables used in our empirical analysis.

[19]As we discuss in the following section, we use our instrument to estimate the causal effect of various training programs compared to non-participation in any training program.

[20]See the subsequent section 5 on how this fits into the approach taken to estimate dynamic treatment effects.

15

pluses on the probability to take part in the wage subsidy and SPST programs, although both can have a longer duration and the later is the most common type of further training. In the case of wage subsidies, however, this is plausible, because the assignment into this program depends largely on labor demand for subsidized work on the firm side and cannot be controlled unilaterally by the employment office. Likewise, it is possible that the relevant adjustment dimension in the case of SPST courses was at the intensive rather than at the extensive margin, i.e., that budget surpluses had an effect on the length of the programs instead of the number of entrants.[21] We examine whether this happened in the case of budget surpluses by looking at entries into long and short SPST programs separately, where we define 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 by significantly increasing entries into a long SPST program, while reducing participation in short SPST courses (the latter effect is statistically insignificant). This finding matches the results for the other programs, namely zero or negative impact for the cheaper short programs and higher numbers of entries into the more expensive long programs. Interestingly, the increase in entries into longer programs like long SPST courses and retraining does not only serve to spend the remaining funds during the current year, but also absorbs part of the budget in subsequent years. This could possibly reflect a commitment device used by local employment offices to signal spending needs in the future, providing an additional safeguard against future budget cuts.

To illustrate the magnitude of the individual effects, a one standard-deviation increase in the surplus (i.e., additional 8.5 available program slots that have not been filled during the first semester) leads to a decline of the participation probability STT by -2.2 percentage points (ppoints) after the summer holidays. The corresponding effects for practice firms, long SPST courses, and retraining are +1.9, +1.7, and +1.5 ppoints, respectively. While the direction of the effects on the training probability can be inferred based on the stratified sample used, where training participants are overrepresented, the quantitative effect estimates cannot be taken literally. For this reason, we also calculate the ratio of these ppoints changes to the average participation probability in our sample, which approximates the relative change in the absolute number of training participants because the vast majority of the unemployed in the full sample to not participate in training. The resulting ratios are -6.1% for STT, +7.1% for practice firms, +4.4% for long SPST courses,

---

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

and +3.0% for retraining. These numbers point to economically sizeable variations in program participation in response to changes in the surplus instrument.

Overall, the results of the first stage indicate a significant impact of budget surplus on the probability to participate in individual training programs after the summer holidays. For all four measures, the corresponding F-statistic is larger than 10, which suggests a sufficiently strong instrument.

Exclusion restriction

Apart from showing an impact on the treatment probability, it is crucial for the validity of our approach that our instrument does not directly influence the outcome variable, i.e., that surplus only affects later employment through its impact on the treatment probability. Put differently, it is necessary that no omitted variable simultaneously affects the budget surplus and employment. An obvious candidate could be persistent trends in regional labor market conditions.[22] For instance, a sustained economic upturn may lead to both persistent budget surpluses and sustained higher labor demand in one region, while a persistent downturn could cause the opposite situation.

For various reasons, this does not seem to be an issue in our case. First, the regional trends in labor demand are likely to influence both the treated and the untreated in the region in a similar way and, if anything, the impact on the training effects are likely to be restricted to the lock-in period (Lechner and Wunsch, 2009). Second, note that we control for differences in the local labor conditions and that our instrument varies only at the level of the REOs, i.e. at a more aggregated level. Thus, we only use the variation in the instrument which is not related with differences in local labor market conditions. Third, there were no persistent differences in the unemployment trends across the various REOs between 1982 and 1993, as shown in figure 6 (upper panel). In fact, one observes 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. The plausibility of our reasoning is supported by the development of the surplus variable by region over time, as depicted in figure 6 (lower panel). No region remains persistently on one side of the deficit-surplus measure, which suggests that there is strong

---

[22]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.

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 basically participation in all training programs considered should have ended. The estimated coefficient is very small and insignificant (the results are available upon request), thus suggesting that our surplus instrument should indeed 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).[23] Effectively, we estimate the effect of training versus waiting, where non-treatment up to a certain elapsed unemployment duration involves the possibility to be treated later during the course of the unemployment spell. To implement our IV approach, we use a flexible two-step control function method for a random-coefficients model with both a binary endogenous treatment and a binary outcome, see Wooldridge (2014). We estimate the employment effect of participation in a specific training program compared to non-participation in any training program making use of one instrument. Heckman et al. (2008) suggest IV estimation in settings with multiple unordered treatments, where the estimation of pairwise effects of one treatment versus another requires the availability of as many instruments as there are treatments.[24] In this section, we describe our sample and estimation approach first and then discuss how it relates to Heckman et al. (2008). In this discussion, keep in mind that we estimate the effect of participation in a specific training program vs. not-taking part in any, and not the pairwise effect of participation in one training program compared to taking part in another.

---

[23]See also Fredriksson and Johansson (2008) for a related approach and Biewen et al. (2014), Fitzenberger et al. (2013), or Lechner et al. (2011) for applications.

[24]Based on a choice-theoretic analysis of local IV estimation, Heckman et al. (2008) show that to estimate the causal effect of one treatment versus another treatment in general requires a covariate (instrument) for each treatment that changes the value of one treatment in the discrete choice model but does not affect both the value of the other treatments and the outcome variable (in our case employment). For the comparison of a treatment to the next best alternative, the necessary assumption is a bit weaker. In this case, only one instrument for the specific treatment investigated is required that satisfies this identification assumption.

Sample definition

Consider a random sample of eligible individuals, who can enter one of the training programs $p \in \{STT, PF, WS, SPST, RT\}$ at any possible elapsed duration $el = 1, ..., 12$ during 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 four 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 who are still unemployed in the month prior to the beginning of the respective stratum and who will not have started any of the aforementioned training programs before the end of the stratum. This defines the non-treatment group $\{d = 0\}$. Correspondingly, the treatment group $\{d = 1\}$ for each training program p involves those individuals who enter p during the stratum considered.

When we estimate the outcome equation for employment by month since treatment start, we account for the fact that the composition of treated and non-treated individuals depends upon the elapsed duration in months. To construct our sample by potential treatment starts, all individuals are replicated for each month they remain unemployed and they are eligible for treatment. For 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 by stratum are the weighted averages of the three corresponding month-specific estimates (see 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.

Estimating the ATT

In the following, we discuss the estimation of the average treatment effect for the treated (ATT), 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:

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

(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. We assume $E(b_t^d \mid z) = 0$. 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 at the state level as well as time-specific variables. The instrument z_2 involves the budget surplus 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 ATT for employment at time t given by

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

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 $d = 1$. Formally, this is

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

By the law of iterated expectations, this is obtained in two steps based on the average structural function (ASF), which was originally introduced by Blundell and Powell (2003, 2004). 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). A more detailed discussion of our estimation approach can be found in the additional appendix A.2.

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 sufficient statistics to capture 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 upon e_2. Under these assumptions, the ASF can be expressed by integrating out the control functions e_2 among the treated $\{d = 1\}$. 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 $gr = d\lambda(z\gamma) - (1-d)\lambda(-z\gamma)$ (Heckman 1978), its square, 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 pooled probit employment regression, 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. As part of the specification search, we routinely test for significance of $\widehat{gr}^2$ and interaction terms $z_{i1}\widehat{gr}$, and we drop insignificant terms for the control function. In order to keep the estimation approach tractable, we impose the restriction that the coefficients for the selection correction terms are time-invariant.

Based on the estimated employment equation, 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

(4)
$$\widehat{\tau}_{ATT,t} = \frac{1}{N_1} \sum_{d_i=1} \left\{ \Phi\left(m_t\hat{\delta}_{0t} + m_t\hat{\delta}_{1t} + z_1\hat{b}_0 + z_1\hat{\delta}_1 + \hat{\omega}_0\widehat{gr}_i + \hat{\omega}_1\widehat{gr}_i + \hat{\omega}_2\widehat{gr}_i^2 + z_{i1}\widehat{gr}_i\hat{\psi}\right) \right.$$

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

where $\Phi(.)$ is the standard normal distribution function. The first (second) term in the difference denote the employment probability for the treated (nontreated).

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 our two-step control function approach with ATT estimates based on pooled OLS employment regressions with the same specification for observables but without the control functions for selection on unobservables.

Finally, we discuss how our approach relates to IV estimation for multiple unordered treatments in Heckman et al. (2008), who require one treatment-specific instrument (see footnote 24). Note first that our analysis does not attempt to estimate an employment equation which simultaneously models the effect of all training programs, i.e. we do not estimate a model with multiple endogenous treatment variables in one outcome equation. We have only one instrument which potentially affects participation in all training

programs considered and therefore the nontreatment group considered may change its selection in response to changes of the instrument. Nevertheless, we think that our binary comparison of treatment in one training program versus nonparticipation in any training program is useful for two reasons. First, our first stage probit selection model is estimated separately for each training program based on those participating in the program considered and those not receiving any treatment. We take this as an approximation of the selection of the treated relative to the nontreated regarding the nontreatment outcome, which is what we need for the estimation of the ATT. Second, the treatment probability of all training programs by stratum is very small (note that the treated are overrepresented in our analysis sample) so that the composition of the nontreatment group is hardly affected by the effect of our budget instrument on the participation in the alternative training programs. Thus, the key issue is to control the selection bias among the treated. Note that it would be more difficult to justify our approach for the estimation of the pairwise effect between two different training programs.

6 Effects of training on subsequent employment

6.1 Descriptive statistics

Table 2 provides an impression of the relevant sample sizes of the treatment groups for different training programs, by length of unemployment at program start. Additionally, we also show the size of the respective control groups, consisting of all eligible unemployed who do not take part in a training program at that time, including those who enter later on. Retraining is the most frequently used type among the evaluated programs. Furthermore, the absolute number of entries into training programs declines for a longer elapsed duration of unemployment, reflecting the fact that the number of unemployed falls as more and more individuals find a job. At the same time, the fraction of individuals who start a program increases over the length of unemployment for each type. In particular, the share of starters in short-term training even increases from 8.8% in the first quarter of unemployment up to 18.2% in the fourth.[25]

Table 3 involves a comparison of mean characteristics between individuals in the treat-

---

[25]Note that these unweighted numbers reflect the situation in our evaluation sample in which treated individuals are over-represented. The fraction of program starts is much smaller in the population, but the relative changes are correctly represented in our sample.

ment and in the control group.[26] It turns out that participants differ significantly from non-participants regarding almost all characteristics considered. For instance, participants are more likely to be in the youngest age group (25-29 years) than non-participants, but less likely to be in the oldest (45-50 years), independent of the respective program type. This is consistent with training being a human capital investment, paying off the longer the younger the age at the time of training. Likewise, we observe smaller fractions of married individuals and foreign citizens among the treated in all programs compared to the control group. This fits to the investment logic, as married individuals are older on average than singles, and foreign citizens may be expected to remain less time in the country, i.e., have a shorter payoff period. In other cases, the differences between participants and nonparticipants are not uniform across programs, possibly reflecting different target groups. More practically oriented programs like short-term training and practice firms, for example, tend to have lower educated individuals than the control group, but higher educated ones in long SPST courses which focus more on acquiring new skills.

Altogether, these differences reveal strong selection effects in observed characteristics among participants in training. This highlight the challenge faced by an identification strategy based on a selection-on-observables assumption, and it may very well be the case that further systematic differences also exist regarding important unobserved characteristics.

6.2   Estimated employment effects

In order to account for a possible selection-on-unobservables in treatment participation, we estimate the average treatment effect on the treated (ATT) of training on subsequent employment of eligible unemployed individuals using the IV-control function approach described in section 5. As reported above, the first stage probit regressions show that our instrument, the regional budget surplus, strongly affects program participation for short-term training, practice firms, long SPST courses, and retraining. In each of these cases, the F-value from the respective first stage regression is greater than 10, suggesting there is no weak instrument problem.

For the estimation of the second stage, we first use the data to determine the best specification of our control function. For this purpose, we start with a flexible model that includes not only the generalized residuals obtained from the first stage, but also inter-

---

[26]Descriptive statistics for short SPST courses and wage subsidies are available from the authors upon request.

actions between the generalized residual, the treatment dummy, and selected important covariates, as well as the generalized residual squared. We then test separately whether the components of this general model are (jointly) significant and take out those that do not contribute much in explaining the variation in employment probabilities. The results of these tests are reported in the upper panel of table 4. They reveal a strong joint significance of all selection correction terms together (the Chi2-test has a p-value smaller than 0.015 for all programs), which indicates the high degree of endogeneity in the participation in training. Looking at the interaction terms between the generalized residuals, the treatment dummy, and the control variables considered yields the same result. This suggests that the selection into participation varies by characteristics of the individual. The squared generalized residual, however, is never significant with p-values between 0.2 and 0.3. The lower panel of table 4 therefore shows the outcomes of joint significance tests for a reduced second model without the squared generalized residual. Here, the results of the Chi^2-tests indicate a significant joint influence both for all correction terms together and for the interaction terms alone. Thus, we do not further reduce the model and use this flexible specification for the remainder of the analysis

The results of the second stage obtained from this specification are displayed in figures 7 to 10. The dark lines depict the IV-control function estimates for the employment effect of the respective training program for each month over 10 years after program start.[27] The corresponding bootstrapped confidence intervals obtained from 250 repetitions are depicted as dashed lines. As comparison, we also report the estimated coefficients and confidence intervals for a standard OLS regression with the same set of controls as continuous and dashed gray lines, respectively.

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, the longer the lock-in effects last. Thus, while it takes around seven to eight months to pass the dip in employment during participation compared to the non-participants in the case of short-term training and practice firms, we observe lock-in periods of around 10 and 22 months for long SPST

---

[27]Recall that a point estimate represents the weighted average of the separate estimates by month of program entry during the first year of the unemployment spell. Further, recall that we estimate pooled probit regressions for program starts in a month during one of the four three-months strata.

courses and retraining, respectively. On the other hand, the "shorter" programs, short-term training, practice firms, and long SPST, seem to have stronger retention effects, with peak values between -30 and -35 ppoints, compared to only -23 ppoints for retraining.

Towards the end of the duration of each program, the estimates increase again and then remain remarkably stable in the medium- and long-run, from the third or forth year onwards. These long-run effects vary strongly across program type, however. While retraining increases the probability to be employed 10 years after program start by around 20 ppoints, the picture is much bleaker for the other programs in our study. Long SPST courses do not improve the long-run employment chances at all (the coefficient is around -3 ppoints and insignificant), whereas short-term training and working in practice firms even decrease the later chances on the labor market significantly, with estimates of around -15 ppoints in both cases. Compared with the OLS outcomes, these findings indicate lower returns to taking part in short-term training, practice firms, and long SPST programs (the latter is insignificant), but higher ones for retraining.

Table 5 summarizes the control function results from figures 7 to 10 numerically. For simplicity, we average the month-specific results for three time categories: The first year after the start for the immediate short-run impact, the second year capturing the transition period, and years three to ten for the long-run average effects. The patterns from the graphs directly translate into the numbers, showing improving effects over time for each of the four programs and substantial variation in the final long-run outcomes. It also displays nicely the trend towards better long-run results with increasing program duration.

The above results lead us to two important insights. The first is that the effectiveness of the different further training programs varies more strongly than previously thought. Longer programs which focus on the acquisition of new skills fare exceedingly better than shorter programs with an emphasis on exercising, practicing, and deepening the existing stock of 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.[28] The second insight relates to the empirical question of whether selection into training takes place and which way it goes. Here again, the results suggest that the answer is not straightforward. On the one hand, the CF estimates tend to show more negative effects of short-term training,

---

[28]Short-term training was in fact discontinued at the beginning of 1993 and reintroduced in a different format in 1997.

practice firms, and long SPST courses than the corresponding OLS ones, indicating a positive selection of participants with respect to important unobservable characteristics. On the other hand, we find the reverse for retraining. This means that we cannot confirm claims or assumptions of positive selection into training programs in general, but rather each program and its characteristics have to be examined separately.

6.3   Tests for robustness

The next step is to test the robustness of these results. We do this in two ways: First, we check how our first stage estimates react to certain changes in the the empirical specification and the derivation of our instrument. And second, we compare our main results with those of alternative econometric approaches to examine how the choice of evaluation method influences our main findings and their interpretation in the second stage.

Table 6 reports the average partial effects of surplus on the treatment probability in four different probit specifications with increasing numbers of control variables. 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. The results show that adding time and regional specific information influences the estimates as expected, since our approach relies on the variation in budget surplus that is not explained by year or regional fixed effects. We cannot find a general pattern of this effect, however, neither on the coefficients and their signs and significance, nor on the resulting F-statistics. Even more important is whether adding controls, that do not seem strictly necessary, changes size and significance of the effect of budget surplus. When comparing the coefficients reported in columns 3 and 4, we see that this is not the case. This demonstrates that our result in the first stage is not affected by the particular choice of covariates.

Furthermore, the first stage is also robust to different ways to compute the instrument. Table 7 displays the results from the benchmark derivation (column 1) next to the outcomes of two alternative ways of deriving the surplus variable in columns 2 and 3: First, we drop the "macro adjustment", i.e., the correction of predicted entries for changes in aggregate spending for job training programs. The reason for this check is that the necessary information only exists up until 1989, because after German Unification no separate numbers for West Germany have been reported in the statistical yearbooks of the Federal

Employment Office. And second, we include the December of the previous year in the computation of budget surplus in the first half of the respective year, because entries in December only affected the budget in January. In both cases, the magnitudes and signs of the average partial effects do not change much compared to our benchmark in column 1. The only relevant difference is that the F-test for the instrument decreases to under 10 for retraining in column 2.

Next, column 4 of table 7 reports the first stage results if the outcome variable, program participation, only included training starts taking place between August and October, but not November. In this case, we see that the average partial effect of surplus decreases somewhat for all programs, while the corresponding standard errors increase. Thus, we observe that the F-statistic for the relevance of our instrument is smaller in this case and only surpasses 10 for one of the four programs anymore. This finding is not surprising, however, because "end-of-year" spending behavior naturally occurs up to the end of the year. Thus, restricting the time period for the effect to unfold reduces what we see as response to unexpected changes in the budget.

With respect to the second stage, we want to see how using a different identification strategy with the same data and variables would change the results compared to our control function approach. Therefore, we repeat the analysis with four alternative methods: Simple differences between treatment and control group, OLS, matching (using inverse probability weighting), and standard IV using predicted probability of participating in training as instrument. The estimated coefficients for the treatment are reported in table 8. As before, we abstain from reporting them for every single month after treatment start, but pool them for the same three distinct periods over the evaluation time.

Three aspects of table 8 are noteworthy: First, there is no difference between the various methods with respect to the pattern of results they produce. That is, all of them exhibit significant lock-in effects in the first year, slightly better outcomes in the second, and the even better results in the long run. The only exception are practice firms, for which the IV estimates for the long-run impact are worse than for the second year. Second, the estimates obtained by selection-on-observables approaches (OLS and matching) differ in magnitude from those of the two types of IV approaches, although we control for a large number of relevant explanatory variables. This illustrates the presence of strong selection into program participation and the great challenge of taking it into account. Note also that the estimators that allow solely for selection on observables can only to a limited extent explain post-treatment employment differences between participants and

nonparticipants as the difference between column (1) and e.g. column (2) in table 8 is, although mostly significant, relatively small.

Third, the comparison between IV and control function estimates highlights the two special features of the control function approach. On the one hand, the flexible specification of the selection correction yields more precise estimates than the traditional IV specification. In table 8, the resulting standard errors are smaller in the control function approach in 11 out of 12 cases (by an average of 18-19% over all programs and time periods). On the other hand, we can see that the difference between the IV and CF estimates follows a certain pattern. For both short-term training and practice firms, the CF point estimates for the long-run effect are much smaller than the IV estimates, whereas they are the roughly the same for long SPST courses and retraining. This highlights the difference in meaning and interpretation of the coefficients in the two approaches. While our control function approach estimates the average treatment effect on the treated (ATT) in general, standard IV measures the local average treatment effect (LATE) on the population of compliers. In the present case, this means that we compute the effect of program participation for all participants in the final months of the year, while IV reflects the impact on those who only take part because the respective regional employment office needed to spend their remaining funds, independent of whether that makes any sense for the individual. For the programs with negative long-run effects, we would expect the LATE estimates to show worse effects than the ATT estimates.

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 during the first semester of a year. Since a direct transfer of funds from one instrument (e.g. training as a whole) of active labor market policy to another one 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.

Our empirical analysis of the employment effects of training leads to the following main findings: First, at a methodological level, we show that our flexible control function approach, which allows us to estimate the average effect of treatment on the treated, leads to more precise estimates of the program impact compared to standard IV, which estimates the local average treatment effect. Second, our estimates for the long-run employment effects of training are not homogeneous, rather they differ by the type of program. Programs focusing on the acquisition of a large, well defined body of new general and specific skills through retraining towards a new vocational training degree strongly increase employment of participants in the long run. In contrast, long programs providing some specific skills (SPST) do not improve employment in the long run, and programs which focus on the practice of the existing stock of skills (practice firms) or programs involving short-term training and job search assistance even reduce employment in the long run. Third, selection based on unobservable characteristics like motivation, ambition, unobserved ability, or strive seems to be strong, but not uniformly positive or negative. While we see evidence for a positive selection of participants in short-term training, practice firms, and long SPST programs, participants in retraining are negatively selected. The later suggests that retraining is targeted towards individuals with fairly low employment chances.

These findings are useful for the policy debate on the effectiveness of public-sponsored job training programs. One can not overemphasize the heterogeneity of findings across programs. On the one hand, the strong effectiveness of the long, intensive retraining programs leading to a new vocational training degree suggest that intensive human capital acquisition is more effective than short-term training involving job search assistance or work experience in the artificial work environment of practice firms. On the other hand, the lack of positive long-run employment effects for long SPST programs cautions against a generalization to all forms of intensive human capital investment.

References

Aakvik, A., J. 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).

Caliendo, M., R. Mahlstedt, and O. Mitnik (2014). Unobservable, but Unimportant? The Influence of Personality Traits (and Other Usually Unobserved Variables) for the Evaluation of Labor Market Policies. *IZA Discussion Paper 8337*.

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.

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

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

Heckman, J. J., R. LaLonde, and J. Smith (1999). The Economics and Econometrics of Active Labor Market Programs. *In: O. Aschenfelter and D. Card (eds.), Handbook of Labor Economics Vol. 3 A, Amsterdam: Elsevier Science*, 1865–2097.

Heckman, J. J., S. Urzua, and E. Vytlacil (2008). Instrumental variables in models with multiple outcomes: The general unordered case. *Annales d'Economie et de Statistique*, 151–174.

Heinrich, C., P. Mueser, K. Troske, K. Jeon, and D. Kahvecioglu (2013). Do Public Employment and Training Programs Work? *IZA Journal of Labor Economics 2:6*, 1–23.

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.

Kluve, J. (2010). The effectiveness of European active labor market programs. *Labour Economics 17*(6), 904–917.

Lechner, M. and C. Wunsch (2009). Are Training Programs More Effective When Unemployment Is High? *Journal of Labor Economics 27*(4), 653–692.

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.

Richardson, K. and G. J. van den Berg (2013). Duration dependence versus unobserved heterogeneity in treatment effects: Swedish labor market training and the transition rate to employment. *Journal of Applied Econometrics 28*(2), 325–351.

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 endogenous 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 the budget surplus

Program entries in 1st half of the year → "SURPLUS"

Mid-year evaluation

Summer holidays

Reaction in the number of program entries

Counts towards the next fiscal year

January   June   July   August   September   November
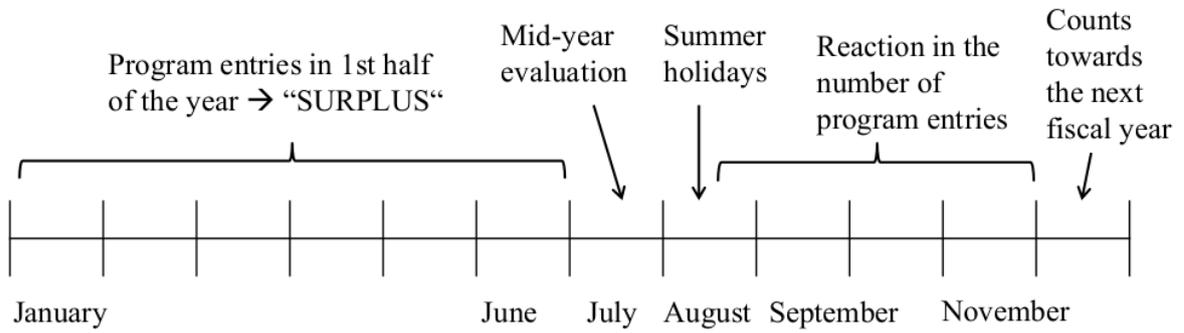
Figure 3: Monthly shares of total year entries into training programs, 1982 – 1993
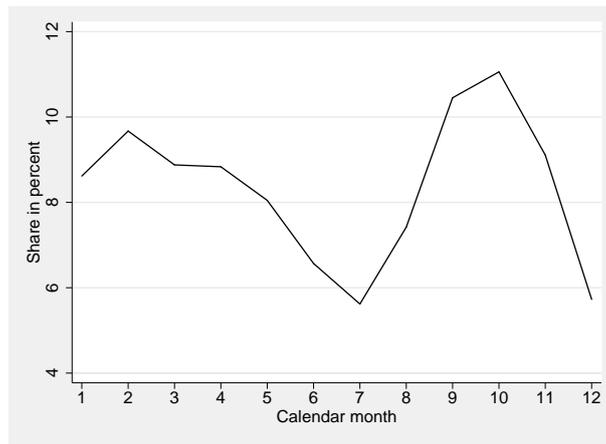
Note: Shares are calculated on the basis of entry numbers averaged over region and time.

Figure 4: Actual and predicted number of entrants in training programs by West German regions – 1980 to 1993
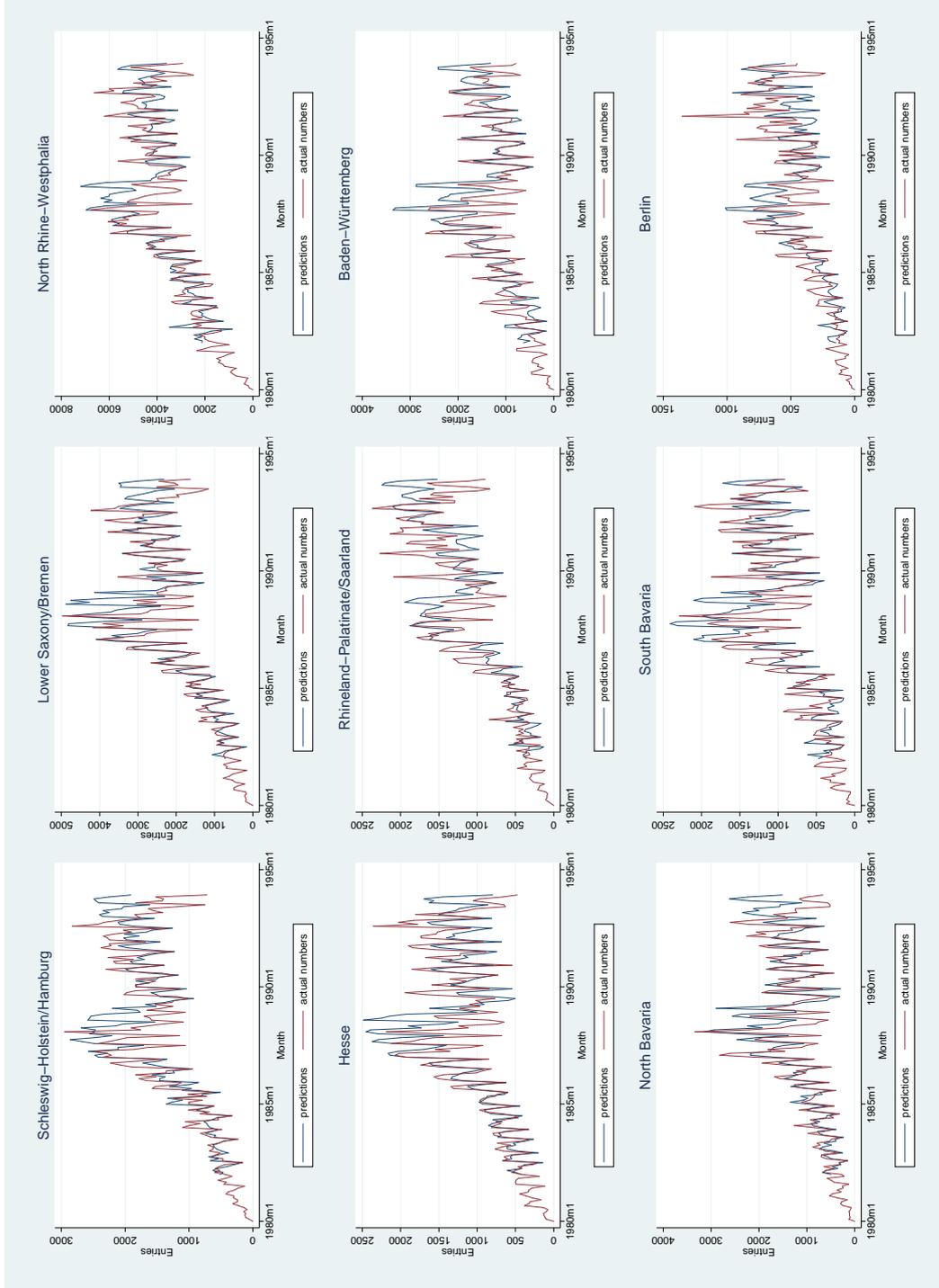
Note: Predicted numbers are out-of-sample predictions calculated by a regression of entry numbers on a set of the year and calendar month dummies, period number, and the monthly entries of the previous year.

39

Figure 5: Distribution of the budget surplus
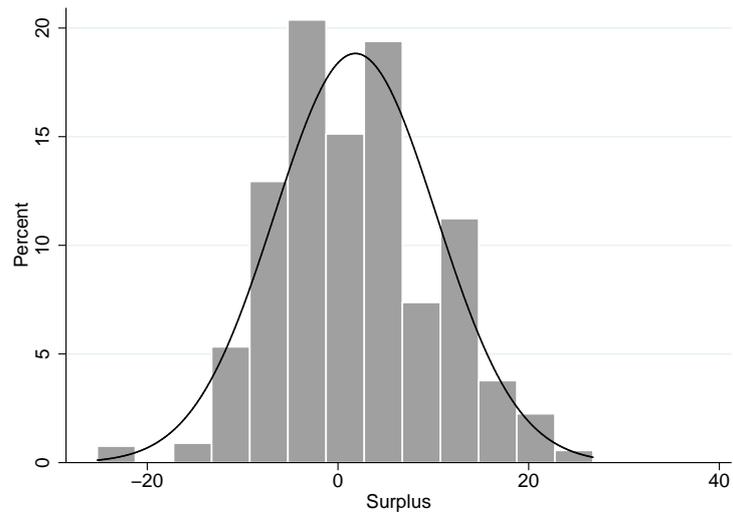
Note: Budget surplus is defined as the difference between planned and actual entries per 1000 eligible unemployed cumulated over the first half of the year.

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

Unemployment rate

Surplus (deviations from yearly averages)

Schleswig–Holstein/Hamburg	Lower Saxony/Bremen	North Rhine–Westphalia
Hesse	Rhineland–Palatinate/Saarland	Baden–Württemberg
North Bavaria	South Bavaria	Berlin
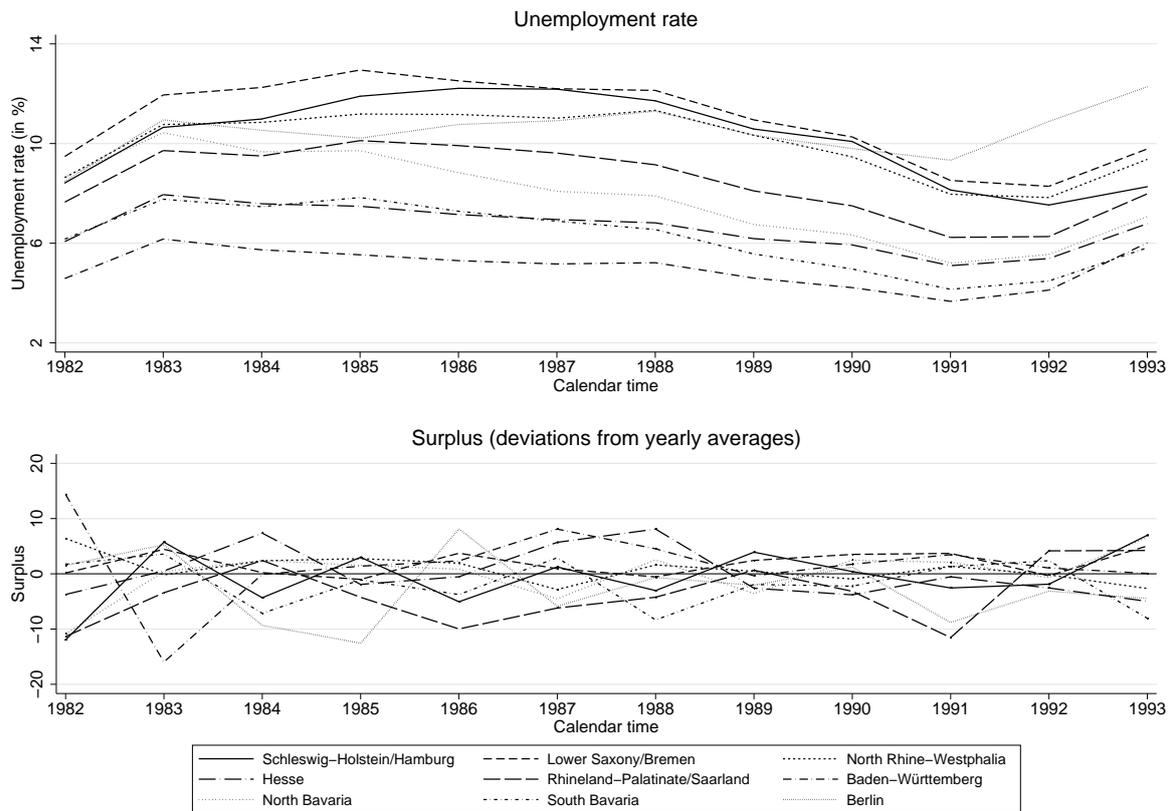
Note: The monthly unemployment rate is averaged over the first half of the year.

Table 1: The effect of budget surplus on treatment probability

Evaluated program	(1) STT	(2) PF	(3) WS	(4) SPST	(5) RT	(6) Short SPST	(7) Long SPST
Surplus	-0.261***	0.224***	0.097*	0.014	0.176***	-0.078	0.198***
	(0.067)	(0.060)	(0.052)	(0.043)	(0.057)	(0.056)	(0.068)
F-statistic	17.55	14.11	3.48	0.10	10.56	2.04	11.54
No. of clusters	70 003	76 717	73 045	115 515	93 036	86 790	83 445

Note: Average partial effects per 1000 unemployed (in percentage points) obtained from a first stage probit regression of the probability to participate in short-term training (STT), practice firms (PF), wage subsidies (WS), specific professional skills and techniques (SPST), retraining (RT), 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 2: Size of treatment and control groups by elapsed unemployment duration and evaluated program

Elapsed duration	1–3 months	4–6 months	7–9 months	10–12 months
Short-term training (STT)				
Treatment group	4 040	3 442	2 970	2 651
Control group	45 710	27 020	20 406	14 558
Practice firms (PF)				
Treatment group	3 183	2 486	2 198	1 809
Control group	44 880	26 448	20 099	14 665
Long specific professional skills and techniques (long SPST)				
Treatment group	6 192	4 562	3 446	2 538
Control group	45 929	27 163	20 384	14 324
Retraining (RT)				
Treatment group	11 105	6 177	5 389	3 978
Control group	46 835	27 310	20 453	14 342

Table 3: Summary statistics for selected explanatory variables

| | Short-term training (STT) | | | | Practice firms (PF) | | | |
| | Treated | | Control | | Treated | | Control | |
Variable	Mean	St.dev	Mean	St.dev	Mean	St.dev	Mean	St.dev
Female and age								
Female	**0.441**	(0.496)	**0.395**	(0.489)	**0.323**	(0.468)	**0.396**	(0.489)
25-29 years old	**0.302**	(0.459)	**0.268**	(0.443)	**0.300**	(0.458)	**0.271**	(0.444)
30-34 years old	0.226	(0.418)	0.219	(0.413)	0.225	(0.418)	0.223	(0.416)
35-39 years old	0.177	(0.382)	0.176	(0.381)	**0.184**	(0.387)	**0.174**	(0.379)
40-44 years old	**0.151**	(0.358)	**0.159**	(0.366)	0.153	(0.360)	0.160	(0.367)
45-50 years old	**0.144**	(0.351)	**0.178**	(0.382)	**0.138**	(0.344)	**0.172**	(0.378)
Education								
No voc. training degree	**0.215**	(0.411)	**0.201**	(0.400)	**0.211**	(0.408)	**0.201**	(0.401)
Voc. training degree	**0.726**	(0.446)	**0.715**	(0.452)	**0.763**	(0.425)	**0.718**	(0.450)
Uni/college degree	**0.054**	(0.226)	**0.079**	(0.270)	**0.023**	(0.148)	**0.077**	(0.267)
Education unknown	0.005	(0.069)	0.005	(0.073)	**0.003**	(0.057)	**0.005**	(0.069)
Marital status, children in household, foreigner								
Married	**0.447**	(0.497)	**0.480**	(0.500)	**0.447**	(0.497)	**0.479**	(0.500)
Kids	**0.280**	(0.449)	**0.296**	(0.456)	**0.330**	(0.470)	**0.307**	(0.461)
Foreigner	**0.074**	(0.262)	**0.123**	(0.328)	**0.095**	(0.293)	**0.117**	(0.321)
Previous employment and employment status								
# months employed	**18.7**	(6.4)	**18.9**	(6.3)	**18.3**	(6.6)	**18.9**	(6.2)
Log wage	**3.815**	(0.663)	**3.791**	(0.772)	**3.857**	(0.603)	**3.801**	(0.777)
Apprentice	**0.010**	(0.099)	**0.014**	(0.119)	**0.010**	(0.101)	**0.014**	(0.115)
Blue collar worker	**0.554**	(0.497)	**0.578**	(0.494)	**0.636**	(0.481)	**0.577**	(0.494)
White collar worker	**0.350**	(0.477)	**0.313**	(0.464)	**0.287**	(0.452)	**0.307**	(0.461)
Worker at home	**0.001**	(0.036)	**0.005**	(0.068)	**0.001**	(0.030)	**0.004**	(0.066)
Part-time worker	0.085	(0.279)	0.090	(0.286)	**0.066**	(0.248)	**0.098**	(0.297)
Occupation from previous employment								
Farmer/Fisher	0.028	(0.165)	0.026	(0.160)	**0.038**	(0.190)	**0.029**	(0.167)
Manufacturing	**0.354**	(0.478)	**0.372**	(0.483)	**0.417**	(0.493)	**0.380**	(0.485)
Technicians	**0.031**	(0.174)	**0.035**	(0.184)	**0.018**	(0.133)	**0.036**	(0.186)
Service	**0.558**	(0.497)	**0.537**	(0.499)	**0.475**	(0.499)	**0.527**	(0.499)
Miners/Others/Missing	0.029	(0.169)	0.029	(0.169)	**0.052**	(0.222)	**0.028**	(0.166)
Former treatment participation before current unemployment								
1 year before	**0.040**	(0.197)	**0.034**	(0.182)	**0.063**	(0.243)	**0.034**	(0.180)
2 years before	**0.064**	(0.245)	**0.055**	(0.227)	**0.099**	(0.299)	**0.056**	(0.230)

<continued on next page>

Table 3 – <continued from previous page>

| | Long SPST | | | | Retraining (RT) | | | |
| | Treated | | Control | | Treated | | Control | |
Variable	Mean	St.dev	Mean	St.dev	Mean	St.dev	Mean	St.dev
Female and age								
Female	**0.531**	(0.499)	**0.400**	(0.490)	**0.425**	(0.494)	**0.394**	(0.489)
25-29 years old	**0.261**	(0.439)	**0.274**	(0.446)	**0.416**	(0.493)	**0.280**	(0.449)
30-34 years old	**0.253**	(0.435)	**0.219**	(0.414)	**0.283**	(0.451)	**0.219**	(0.414)
35-39 years old	**0.202**	(0.401)	**0.175**	(0.380)	**0.161**	(0.368)	**0.178**	(0.383)
40-44 years old	0.159	(0.366)	0.158	(0.365)	**0.094**	(0.292)	**0.160**	(0.367)
45-50 years old	**0.125**	(0.331)	**0.174**	(0.379)	**0.045**	(0.208)	**0.163**	(0.369)
Education								
No voc. training degree	**0.113**	(0.316)	**0.194**	(0.395)	0.199	(0.400)	0.202	(0.401)
Voc. training degree	**0.741**	(0.438)	**0.721**	(0.448)	0.714	(0.452)	0.713	(0.452)
Uni/College degree	**0.144**	(0.351)	**0.080**	(0.271)	0.084	(0.277)	0.081	(0.272)
Education unknown	**0.003**	(0.051)	**0.005**	(0.072)	**0.003**	(0.052)	**0.005**	(0.067)
Marital status, children in household, foreigner								
Married	**0.472**	(0.499)	**0.488**	(0.500)	**0.408**	(0.491)	**0.473**	(0.499)
Kids	**0.357**	(0.479)	**0.309**	(0.462)	**0.298**	(0.458)	**0.306**	(0.461)
Foreigner	**0.052**	(0.223)	**0.121**	(0.326)	**0.074**	(0.262)	**0.124**	(0.330)
Previous employment and employment status								
# months employed	19.0	(6.4)	19.1	(6.1)	18.9	(6.2)	18.9	(6.2)
Log wage	**3.838**	(0.783)	**3.792**	(0.804)	**3.835**	(0.718)	**3.793**	(0.792)
Apprentice	**0.012**	(0.109)	**0.015**	(0.121)	**0.019**	(0.136)	**0.015**	(0.120)
Blue collar worker	**0.321**	(0.467)	**0.567**	(0.496)	**0.564**	(0.496)	**0.579**	(0.494)
White collar worker	**0.541**	(0.498)	**0.319**	(0.466)	**0.327**	(0.469)	**0.309**	(0.462)
Worker at home	**0.001**	(0.036)	**0.004**	(0.066)	**0.002**	(0.040)	**0.003**	(0.054)
Part-time worker	**0.124**	(0.330)	**0.095**	(0.293)	**0.088**	(0.284)	**0.094**	(0.292)
Occupation from previous employment								
Farmer/Fisher	**0.018**	(0.133)	**0.029**	(0.168)	**0.021**	(0.145)	**0.027**	(0.161)
Manufacturing	**0.222**	(0.415)	**0.340**	(0.483)	**0.343**	(0.475)	**0.383**	(0.486)
Technicians	**0.074**	(0.262)	**0.040**	(0.195)	**0.030**	(0.171)	**0.037**	(0.189)
Service	**0.671**	(0.470)	**0.534**	(0.499)	**0.584**	(0.493)	**0.526**	(0.499)
Miners/Others/Missing	**0.015**	(0.123)	**0.028**	(0.165)	**0.021**	(0.143)	**0.027**	(0.163)
Former treatment participation before current unemployment								
1 year before	**0.058**	(0.234)	**0.034**	(0.180)	**0.037**	(0.189)	**0.033**	(0.179)
2 years before	**0.086**	(0.281)	**0.056**	(0.230)	0.055	(0.228)	0.0559	(0.230)

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. Long SPST training includes programs providing specific professional skills and techniques with a planned duration of more than 6 months.

Table 4: Test on the joint significance of the correction terms

|	Short-term training	Practice firms	Long SPST	Retraining
Specification 1				
Test for all correction terms				
Chi²-statistic	62.385	66.292	64.175	76.366
P-value	0.013	0.006	0.009	0.001
Degrees of freedom	39	40	40	40
Test for interactions between generalized residuals and covariates				
Chi²-statistic	54.388	60.748	57.995	67.268
P-value	0.046	0.012	0.015	0.003
Degrees of freedom	37	38	38	38
Test for generalized residuals squared				
Chi²-statistic	1.146	1.632	1.330	1.648
P-value	0.284	0.201	0.249	0.199
Degrees of freedom	1	1	1	1
Specification 2				
Test for all correction terms				
Chi²-statistic	61.039	64.586	62.746	74.780
P-value	0.014	0.006	0.009	0.001
Degrees of freedom	38	39	39	39
Test for interactions between generalized residuals and covariates				
Chi²-statistic	53.827	60.287	59.336	66.745
P-value	0.036	0.009	0.011	0.002
Degrees of freedom	36	37	37	37

Note: [a]Specification 1 contains the following correction terms: generalized residuals, interactions between generalized residuals and the treatment dummy, interactions between generalized residuals and selected covariables, and the squared generalized residuals. Specification 2 is identical with specification 1 but omits the squared generalized residuals.

Figure 7: Estimated employment effects (ATTs) of short-term training pooled over strata
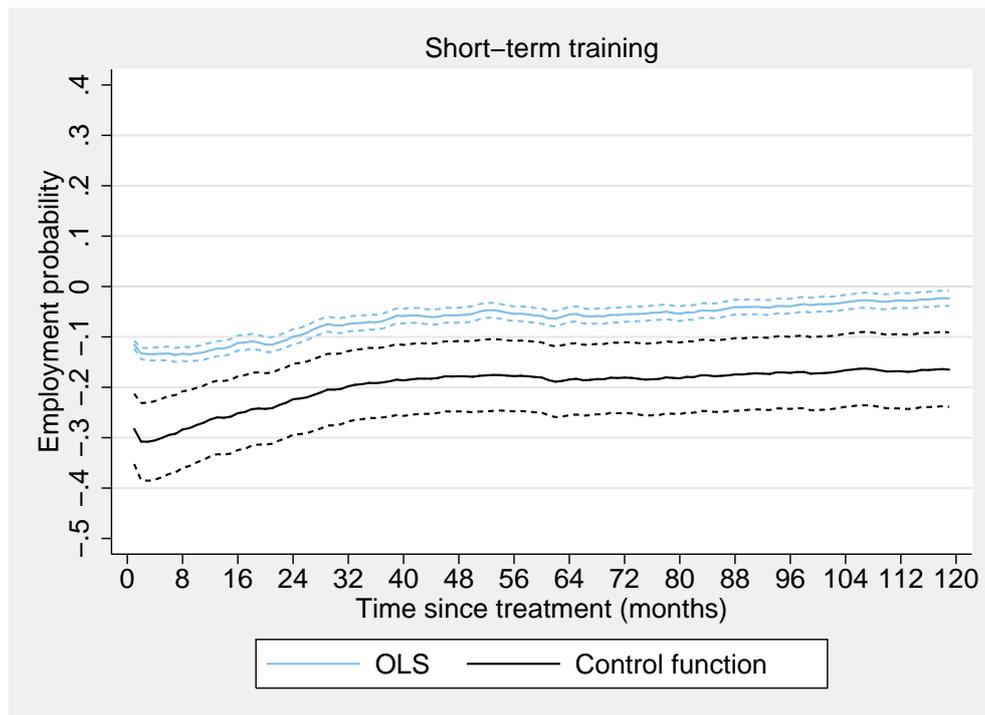
Short–term training

Note: Estimated effects are aggregated over the four quarters of elapsed unemployment duration weighted by the fraction of program participants in the respective stratum. 90% confidence intervals (dashed lines) are obtained through cluster bootstrapping on 250 replications.

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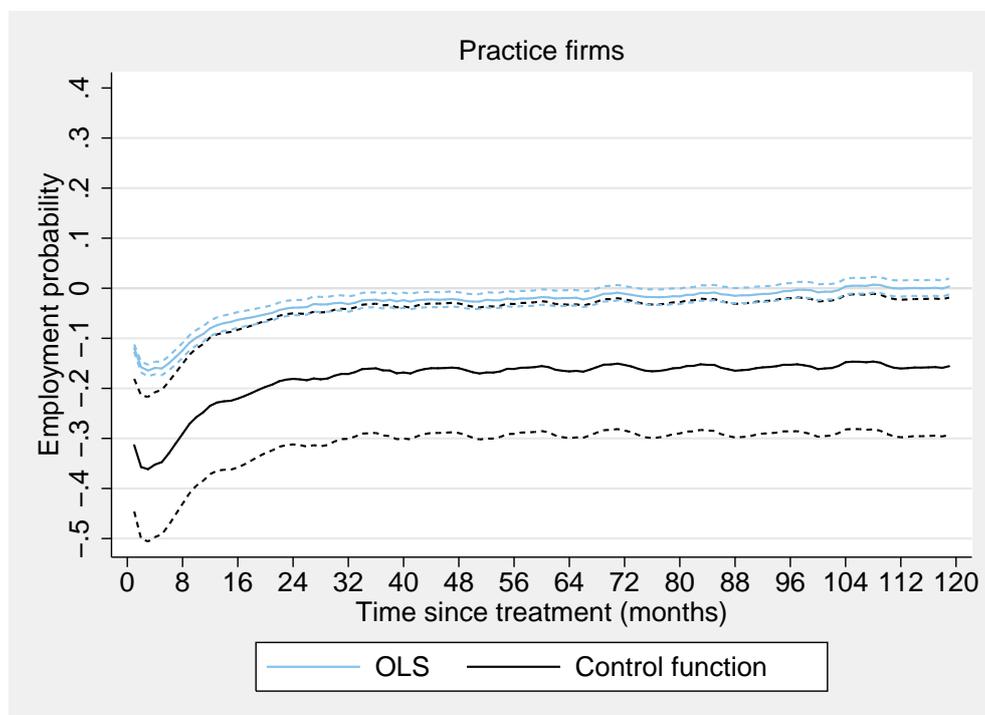
Practice firms

Note: Estimated effects are aggregated over the four quarters of elapsed unemployment duration weighted by the fraction of program participants in the respective stratum. 90% confidence intervals (dashed lines) are obtained through cluster bootstrapping on 250 replications.

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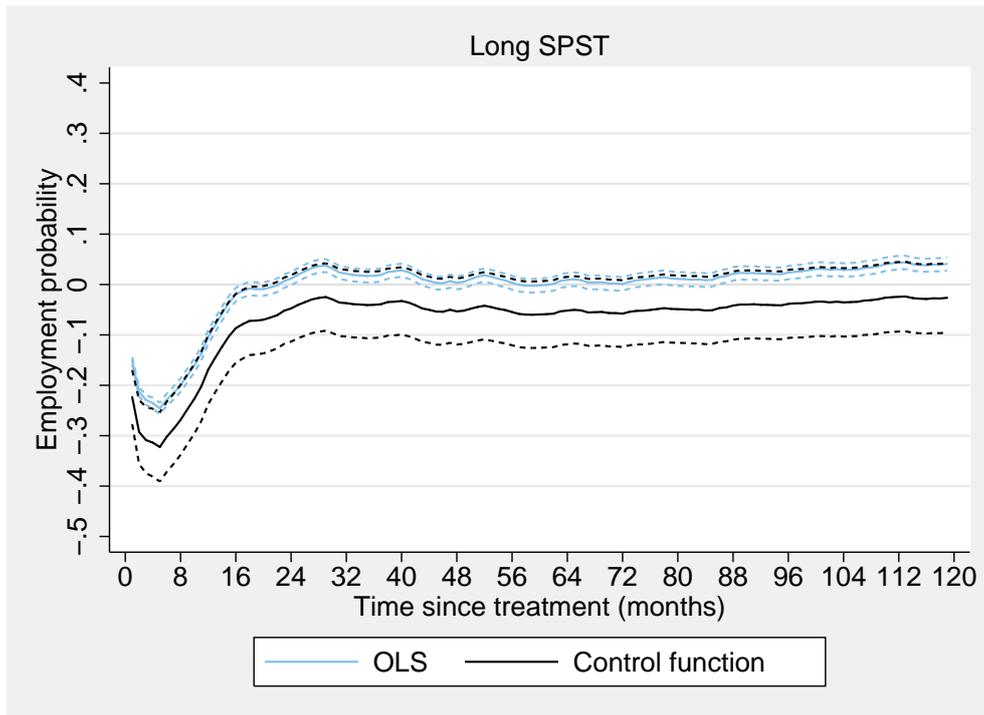

Note: Estimated effects are aggregated over the four quarters of elapsed unemployment duration weighted by the fraction of program participants in the respective stratum. 90% confidence intervals (dashed lines) are obtained through cluster bootstrapping on 250 replications.

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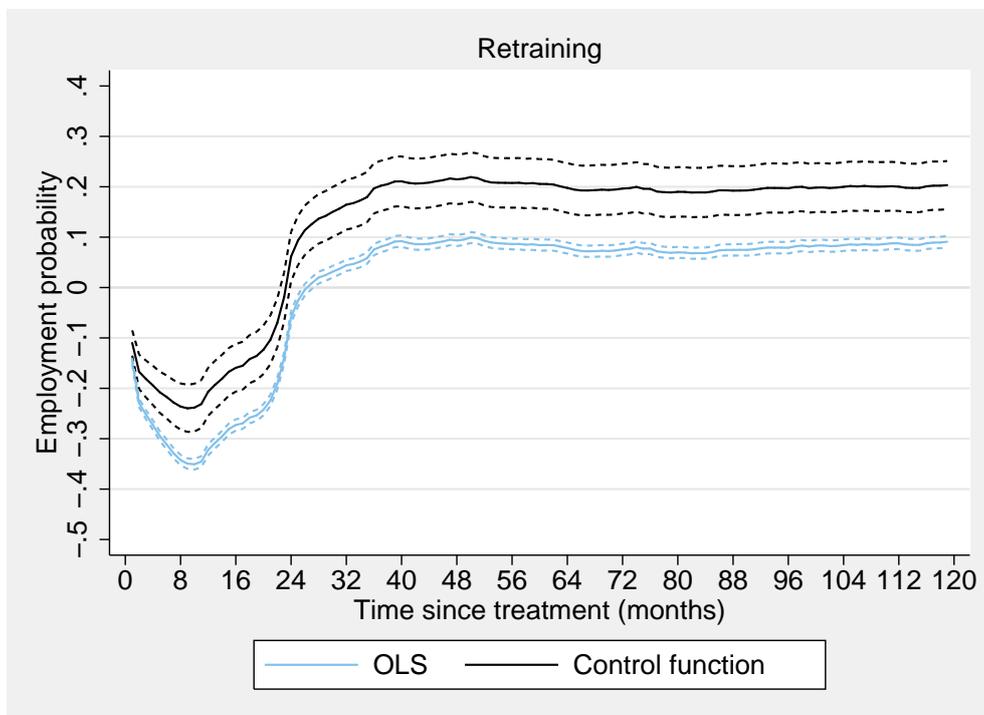

Note: Estimated effects are aggregated over the four quarters of elapsed unemployment duration weighted by the fraction of program participants in the respective stratum. 90% confidence intervals (dashed lines) are obtained through cluster bootstrapping on 250 replications.

Table 5: Employment effects estimated by control function – Main specification

Evaluated program	(1) STT	(2) PF	(3) Long SPST	(4) RT
First year	-0.290***	-0.313***	-0.268***	-0.197***
	(0.030)	(0.061)	(0.026)	(0.016)
Second year	-0.248***	-0.210***	-0.091***	-0.138***
	(0.028)	(0.059)	(0.026)	(0.019)
Years 3-10	-0.180***	-0.161***	-0.043	0.193***
	(0.028)	(0.058)	(0.026)	(0.020)

Note: ***, **, and * indicate statistical significance at 1%, 5%, and 10% level, respectively. Standard errors are obtained through cluster bootstrapping based on 250 replications.

Table 6: Different specifications for the first stage

	(1)	(2)	(3)	(4)
Short-term training	-0.367***	-0.491***	-0.306***	**-0.261*****
	(0.034)	(0.059)	(0.067)	**(0.067)**
	[116.46]	[66.41]	[20.78]	**[17.55]**
Practice firms	0.190***	0.313***	0.243***	**0.224***
	(0.029)	(0.056)	(0.062)	**(0.060)**
	[41.69]	[30.86]	[15.63]	**[14.11]**
Wage subsidies	0.153***	0.110**	0.090*	**0.097***
	(0.026)	(0.050)	(0.054)	**(0.052)**
	[34.40]	[4.95]	[2.74]	**[3.48]**
SPST	-0.231***	0.070	0.011	**0.014**
	(0.024)	(0.044)	(0.045)	**(0.043)**
	[93.39]	[2.53]	[0.06]	**[0.10]**
Retraining	0.065**	0.176***	0.167***	**0.176***
	(0.029)	(0.053)	(0.057)	**(0.057)**
	[4.99]	[10.98]	[8.78]	**[10.56]**
Short SPST	-0.148***	-0.084	-0.064	**-0.078**
	(0.030)	(0.055)	(0.057)	**(0.056)**
	[23.96]	[2.36]	[1.23]	**[2.04]**
Long SPST	-0.053*	0.138**	0.204***	**0.198***
	(0.031)	(0.057)	(0.062)	**(0.068)**
	[3.03]	[5.78]	[11.02]	**[11.54]**
Time specific covariates[a]	no	yes	yes	**yes**
Regional information[b]	no	no	yes	**yes**
Individual characteristics[c]	no	no	no	**yes**

Note: The numbers report average partial effects (APE) per 1000 unemployed in percentage points for separate regressions of the probability to enter the respective program on budget surplus. The relevant F-statistics are reported in square brackets. The benchmark specification used in the empirical analysis is displayed in bold. [a]Time variables include the year and calendar month dummies. [b]Regional information contains region dummies, share of summer holidays at the regional level, and interactions between these variables. [c]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: Different derivations of surplus (1st stage)

	Benchmark	Without macro adjustment[a]	December entries included in surplus[b]	Entry probabilities August-October
	(1)	(2)	(3)	(4)
Short-term training	-0.261***	-0.195***	-0.258***	-0.238***
	(.067)	(0.064)	(0.063)	(0.077)
	[17.55]	[10.05]	[18.78]	[11.03]
Practice firms	0.224***	0.214***	0.215***	0.158**
	(0.060)	(0.059)	(0.057)	(0.069)
	[14.11]	[13.30]	[14.46]	[5.22]
Long SPST	0.198***	0.2045***	0.180***	0.190**
	(0.068)	(0.068)	(0.061)	(0.080)
	[11.54]	[12.65]	[10.59]	[8.17]
Retraining	0.176***	0.161***	0.165***	0.165***
	(0.057)	(0.056)	(0.053)	(0.061)
	[10.56]	[9.10]	[10.13]	[7.95]

Note: Average partial effects per 1000 unemployed (in percentage points) obtained from a first stage probit regression of the probability to participate in the respective program on budget surplus during the first semester. Standard errors are clustered at individual level and reported in parentheses. F-statistics are reported in square brackets. ***, **, and * indicate statistical significance at 1%, 5%, and 10% level, respectively. [a]Surplus without correction for changes in intended spending. [b]Surplus calculated for the time period December to June instead of January to June.

Table 8: Second stage employment effects estimated by different methods

Time since treatment start	(1) Descriptives	(2) OLS	(3) Matching (IPW)	(4) IV	(5) Control function
Short-term training					
First year	-0.146***	-0.131***	-0.132***	-0.235***	-0.290***
	(0.004)	(0.004)	(0.004)	(0.037)	(0.030)
Second year	-0.134***	-0.111***	-0.114***	-0.238***	-0.248***
	(0.005)	(0.005)	(0.005)	(0.047)	(0.028)
Years 3-10	-0.065***	-0.050***	-0.051***	-0.232***	-0.180***
	(0.005)	(0.005)	(0.005)	(0.042)	(0.028)
Practice firms					
First year	-0.155***	-0.143***	-0.144***	-0.160***	-0.313***
	(0.004)	(0.004	(0.004)	(0.056)	(0.061)
Second year	-0.085***	-0.059***	-0.062***	-0.093	-0.210***
	(0.006)	(0.005)	(0.005)	(0.070)	(0.059)
Years 3-10	-0.038***	-0.014***	-0.015***	-0.287***	-0.161***
	(0.005)	(0.005)	(0.005)	(0.063)	(0.058)
Long SPST					
First year	-0.191***	-0.202***	-0.207***	-0.266***	-0.268***
	(0.003)	(0.004)	(0.004)	(0.029)	(0.026)
Second year	-0.024***	-0.034***	-0.038***	-0.082**	-0.091***
	(0.005)	(0.006)	(0.006)	(0.039)	(0.026)
Years 3-10	0.025***	0.022***	0.020***	-0.027	-0.043
	(0.004)	(0.005)	(0.005)	(0.034)	(0.026)
Retraining					
First year	-0.282***	-0.293***	-0.294***	-0.303***	-0.197***
	(0.003)	(0.004)	(0.004)	(0.016)	(0.014)
Second year	-0.256***	-0.254***	-0.255***	-0.358***	-0.138***
	(0.004)	(0.005)	(0.005)	(0.025)	(0.019)
Years 3-10	0.074***	0.073***	0.072***	0.195***	0.193***
	(0.004)	(0.004)	(0.004)	(0.023)	(0.020)

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 (5) 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] Both for computational simplicity and for the possibility to use a vector of flexible control functions, we opt for the two-step control function approach to estimate the average structural function.[31] Wooldridge (2014) shows how the estimated average structural function can be used to estimate the average treatment effect (ATE). We extend the approach to estimate the average effect of treatment on the treated (ATT).

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, because the outcome variable y_t does not only depend upon observed characteristics but also on the unobserved

---

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

[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. 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.

[31]Wooldridge (2014, p. 233) points out that the two-step control function approach involves a different parametric approximation compared to one-step bivariate probit estimation, which tightly specifies the joint distribution of the error terms. Similarly, Terza's (2008) two-stage residual inclusion approach involves yet another parametric approximation.

heterogeneity effects, (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 functions e_2 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 residual gr of the probit model (gr involves the standard Heckman (1978, 1979) selection correction term as defined in section 5).[32] Note that we estimate the effect of treatment started at some point of time on future outcomes at subsequent times t, thus gr is determined at treatment start and does not change over time t. Nevertheless, the control function e_2 may vary over time because (b_t^d, u_t) can change over time.

Assumption **(A3)** imposes ignorability restrictions on the conditional distribution of unobserved heterogeneity, such that conditioning on e_2 in the structural expectation is sufficient to correct for selectivity bias arising from the endogeneity of treatment (Wooldridge, 2014).[33] 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. 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). To increase the flexibility and the robustness of the analysis, 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.[34]

Under assumptions **(A1)**-**(A3)** and by the law of iterated expectations, the average

---

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

[33] Note that a standard exogeneity assumption as used for IV estimation of linear regressions such that the implied error term in equation **(A1)** is independent of the exogenous covariates can not hold because y_t is a discrete outcome variable, see Wooldridge (2014, p. 232).

[34] An alternative extension builds 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 (Kimhi, 1999).

structural function at time t among the treated $d = 1$ can be expressed as

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

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

where $E_\xi [\cdot]$ indicates expectation with respect to the distribution of ξ. Blundell and Powell (2003, 2004) and Wooldridge (2014) define 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 $d = 1$.

As suggested by Wooldridge (2014, section 6.4), we use the following flexible set of control functions $\hat{e}_2(d_i, z_i) = (\hat{gr}_i, \hat{gr}_i^2, \hat{gr}_i d_i, \hat{gr}_i z_{i1})$, which we allow to enter the index function for the employment probit as additional linear regressors (see equation 5). The estimated generalized residual $\hat{gr}_i$ is based on the estimates from the first stage probit for the treatment dummy d_i. As motivated by Lee (1982), the squared $\hat{gr}_i^2$ and the interaction terms $\hat{gr}_i z_{i1}$ account for deviations from the joint normality assumptions as imposed in Heckman (1978). The interaction term $\hat{gr}_i d_i$ accounts (as an approximation) 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, using the approach suggested in Wooldridge (2014, section 6.4), the average structural function for the treated can be expressed by integrating out the control functions e_2 as

(6)
$$ASF(\tilde{d}, z_1, d = 1) =$$

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

Our suggested estimation approach is based on the following insight: Equation (6) makes explicit that once the observed conditional expectation of y_t given (z, d, e_2) is estimated consistently, 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) among the treated.

For estimation purposes, we add estimated versions of the control functions e_2 to a

---

[35]Under joint normality of (ν, b_t^d, u_t), the coefficient of gr_i in the control function differs by treated status d_i because the linear projection of the joint error term $b_t^d d_i + u_t$ on ν differs by d_i. In the absence of the random coefficient part, i.e. $b_t = 0$, and under joint normality of (ν, u_t), the generalized residual $\hat{gr}_i$ has the same coefficient irrespective of the value d_i takes and thus $\hat{gr}_i d_i$ should have a zero coefficient.

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, as well as interactions between m_t, d, and $\hat{e}_2$. In its most general specification, the estimated regression for observation i is

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

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

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 (7), we suggest to estimate the ATT at time t by integrating out the distribution of $z_{i1}, \hat{e}_{i2}$ among the treated $d_i = 1$ as

(8)
$$\hat{\tau}_{ATT,t} = \frac{1}{N_1} \sum_{d_i=1} \left\{ \Phi\left(m_t\hat{\delta}_{0t} + m_t\hat{\delta}_{1t} + z_1\hat{b}_0 + z_1\hat{\delta}_1 + \hat{\omega}_0\widehat{gr}_i + \hat{\omega}_1\widehat{gr}_i + \hat{\omega}_2\widehat{gr}_i^2 + z_{i1}\widehat{gr}_i\hat{\psi}\right) \right.$$

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

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).

A.3 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**, $\geq$ 10 and <20 employees, $\geq$ 20 and < 50 employees, $\geq$ 50 and < 200 employees, $\geq$ 200 and < 500 employees, $\geq$ 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 $\times$ region	Combination of the vacation share and region
Vacation $\times$ calendar month	Combination of the vacation share and the month dummy (August)

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