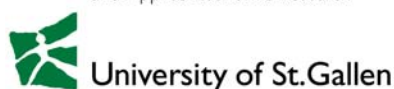


ARE TRAINING PROGRAMS MORE EFFECTIVE WHEN UNEMPLOYMENT IS HIGH?

Michael Lechner and Conny Wunsch
SIAW

Swiss Institute for International Economics
and Applied Economic Research



First version: August, 2006

Date this version has been printed: 20 September 2006

Comments are welcome

Abstract

We estimate short, medium, and long-run individual labor market effects of training programs for unemployed by following program participation on a monthly basis over a ten-year period. Since analyzing the effectiveness of training over such a long period is impossible with experimental data, we use an administrative database compiled for evaluating German training programs. Based on matching estimation adapted to the various issues that arise in this particular context, we find a clear positive relation between the effectiveness of the programs and the unemployment rate over time.

Keywords: Active labor market policy, long-run effects, matching estimation, causal effects, program evaluation, panel data

JEL classification: J 68

Addresses for correspondence

Michael Lechner, Conny Wunsch

Swiss Institute for International Economics and Applied Economic Research (SIAW), University of St. Gallen
Bodanstr. 8, CH-9000 St. Gallen, Switzerland

Michael.Lechner@unisg.ch, Conny.Wunsch@unisg.ch, www.siaw.unisg.ch/lechner

1 Introduction^{*}

Although the body of knowledge about the effectiveness of training programs for the unemployed is rapidly growing, there is not much convincing evidence on the relation of the effectiveness of the programs and the state of the economy. Such information is, however, important. If, for example, changes in the effectiveness of the policy or its different instruments are related to the business cycle, then policymakers can react by adjusting the policy accordingly. Thus, the policymaker should be interested in knowing under which macroeconomic circumstances the programs are more or less beneficial. It is the goal of this paper to provide first insights on this issue.

The empirical literature on the effects of active labour market policies (ALMPs) suggests that almost all programs reduce (unsubsidized) employment and earnings in the short run. This so-called lock-in effect is well documented in many studies and typically attributed to reduced search intensity of program participants or fewer job offers by caseworkers while participating in the program (e.g. van Ours, 2004). If this lock-in effect, which can be interpreted as one component of the cost of ALMPs, varies with labour market conditions, this would be an important argument for varying the composition of programs and program size over time.

With respect to the medium to long-run effects, some wage subsidies and training programs increase employability and earnings (e.g. Couch, 1992; Hotz, Imbens, and Klerman, 2000;

^{*} The first author has further affiliations with ZEW, Mannheim, CEPR, London, IZA, Bonn and PSI, London. Financial support from the Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nuremberg, (project 6-531a) is gratefully acknowledged. The data originated from a joint effort with Annette Bergemann, Bernd Fitzenberger, Ruth Miquel, and Stefan Speckesser to make the administrative data accessible for research. The paper has been presented at workshops at the University of St. Gallen and at the IRP conference in Madison, 2006. We thank participants, in particular John Ham and Jeff Smith, for helpful comments and suggestions. The interested reader will find additional background material for this paper in an internet appendix on our website www.siaw.unisg.ch/lechner/lw_cycles.

Winter-Ebmer, 2001; Jacobson, LaLonde, and Sullivan, 2004; Jespersen, Munch, and Skipper, 2004; Fitzenberger and Speckesser, 2005; Lechner, Miquel, and Wunsch, 2005). Most of this particular literature, which is more optimistic about the effectiveness of ALMPs than most of the older experimental literature, is based on large administrative data sources with long follow-up periods. Understanding the differences between short-run lock-in effects and medium- to long-run effects that may capture more accurately the effects of the human capital added by the programs was an important step towards understanding how these programs work.¹ In fact, this difference will turn out to be crucial for the interpretation of our findings in this paper as well. However, none of these studies systematically investigates if and eventually why effects of different types of programs change over time.

Although direct evidence based on individual data is missing on this issue, there is some evidence based on analyzing regional data over time. For example, Johansson (2001) uses variation of Swedish active labor market programs over municipalities. She shows that the effect of these programs is to prevent unemployed from leaving the labor force during a downturn. She concludes that ALMP programs are most effective during a downturn.²

An alternative to macro studies that come with the usual caveats of aggregation bias and policy endogeneity is to exploit the fact that different micro studies are conducted under different economic conditions. Meta studies are based on this idea. For example, Kluve (2006) combines more than 100 studies, and each study (or specification within a study) constitutes one

¹ The recent increase in evaluation studies is documented for example by the surveys of Fay (1996), Heckman, LaLonde, and Smith (1999), Martin and Grubb (2001), Kluve and Schmidt (2002), and Kluve (2006). For examples of studies based on a selection on observables strategy, see Gerfin and Lechner (2002), van Ours (2004), or Sianesi (2004). A recent example of papers using instrumental variable types of assumptions is Frölich and Lechner (2006). The experimental literature is well documented in the survey by Heckman, LaLonde, and Smith (1999). Boone and van Ours (2004) provide and survey empirical evidence based on aggregated time series data.

data point. In a regression type approach he controls for different aspects of methods and data used, features of the program, as well as the economic environment. Although the analysis of the latter is not the main thrust of his study, he finds the program effects to be somewhat larger when unemployment rates are higher. Thus, his results seem to be roughly in line with Johansson (2001). Although meta studies provide interesting summaries of the literature, there are problems as well. The different individual studies that are treated as the data of the meta analysis are based on heterogeneous programs that are run in different institutional environments and economic conditions, and with different types of participants. It is obviously very challenging to control for all these background factors within a regression framework using only a few control variables and tight functional forms dictated by the limited degrees of freedom available.

In this paper, we retain the advantage of the classical micro evaluation studies, like nonparametric identification and heterogeneity of the program effects, but adjust the standard methodology to learn important aspects about the evolution of the effects over time. Since there are no experiments running for a sufficiently long period to be interesting for such an investigation, any such endeavor has to rely on observational data. Survey data, however, are typically problematic because of insufficient sample sizes, insufficient covariate and program information, short time windows to observe outcomes, as well as misreporting. Newly available high-quality administrative data can overcome these problems. Europe, where experiments are rare because of strong political resistance, has gained competitive advantage in providing large and informative administrative data bases that allow much richer analyses than experimental data which are usually used in the U.S.³

² This mechanism of the programs leading to a redirection of the flows from unemployment to out-of-labor-force towards unemployment and then towards employment appears in the cross-sectional study by Lechner, Miquel, and Wunsch (2005), as well.

³ There are only few observational studies using U.S. data and non of the data bases used is sufficiently informative in terms of covariates and the time horizon covered to study time variation of the effects of

We exploit a particularly informative administrative micro data set for Germany that became available only recently. These data contain reliable information on participants (and nonparticipants) in different types of training programs on a monthly basis from 1986 to 1995. Information on labor market outcomes is available monthly from 1980 to 2003. Thus, the data allow to investigate whether changes in labor market conditions influence the lock-in effects in a different way than the medium- or long-run effects.

The data have been used recently in classical evaluation studies by Fitzenberger and Speckesser (2005), and Lechner, Wunsch, and Miquel (2005), among others. These studies argue that the data are informative enough to control for selective participation and thus allow identification of program effects by matching methods. Based on this identification strategy, we analyze the effects of training programs on short- to long-run labor market outcomes for unemployed entering programs over 10 years on a monthly basis. Another advantage of using Germany for analyzing potential time variation in the effects of training is that no major changes occurred within the broad types of training programs considered in this paper or in the institutional setup.

Our empirical strategy relies on different matching estimators. We begin with analyzing the evolution of the effects over time. Thus, in this specification the characteristics of participants and the use of different program types may vary over time. Any time pattern of the effects that we might isolate from this step may thus be due to changes in the composition of programs, of participants, and/or of economic conditions. Next, by modifying the matching estimator, we keep the characteristics of the program participants constant over time. Thus, the remaining dynamics in the effects reflect changes in program composition and economic conditions only. Then, keeping the shares of the various subprograms and planned program durations constant also allows us to isolate the effects of the economic environment. Finally,

ALMP in such detail as we do (see in particular the survey by Heckman, LaLonde and Smith, 1999, and

to improve our confidence in a causal interpretation of the strikingly clear pattern we obtain, the results are subjected to an intensive sensitivity analysis.

In line with the recent literature mentioned above, we consistently find negative lock-in effects as well as positive medium to long-run employment and earnings effects of the training programs in the 10-year period we consider. However, we detect considerable variation of those effects over time which remains even for a fixed population of participants and a fixed composition of the programs. This variation is clearly related to the unemployment rate prevailing at the start of the program: The negative lock-in effects are smaller and the positive long-run effects are larger in times of higher unemployment. The effects are related to the unemployment rate at the time when the outcome is measured as well. However, whereas the relation to the unemployment rate at program start has direct policy implications, it is harder to see the implications of the relation to the unemployment rate measured much later, because the latter is unknown at the time when decisions about program sizes are to be made.

The remainder of the paper is organized as follows: Section 2 provides background information on the economic conditions, the unemployment insurance system, and the use of active labor market policies in West Germany in the relevant period. In Section 3, the data and the sample are outlined. Section 4 details the econometric identification and estimation strategy. In Section 5, we discuss in detail the effects of training over time. In the following section, we analyze how changing characteristics of participants or the changing composition of programs over time may have influenced the effectiveness of training. Section 7 describes the result of our extensive sensitivity analyses. The last section concludes. An appendix contains further details on the data, the definition of our sample and the outcome variables as well as on the estimation procedure. A second appendix, that is available in the internet, contains detailed background material.

Jacobson, LaLonde and Sullivan, 2004, for example).

2 Economic conditions and institutions in West Germany

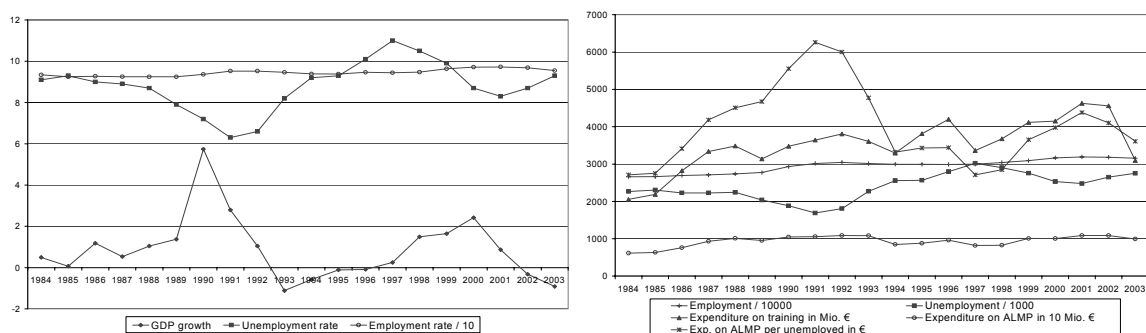
2.1 *The West German economy between 1984 and 2003*

During the economic slowdown following the second oil-price shock, unemployment in West Germany had risen to a quite persistent 9% in the mid 1980s.⁴ Economic activity kept declining until 1988 when a slow recovery started. Directly after unification in 1990, West Germany experienced a boom with substantial East German spending diverted away from domestic products to previously unavailable West German goods. Accordingly, production and labor demand increased in West Germany. GDP grew 5.7% in 1990 and 5% in 1991. Registered unemployment declined to a rate of 6.3% in 1991 despite a significant growth of the labor force due to migration from East Germany and Eastern Europe. At the same time, the world economy was experiencing a recession. In 1992, this recession hit West Germany as well. Economic growth slowed down to only 1.7%. One year later, the West German economy was deep in recession. GDP declined by 2.6% in 1993 and unemployment rose to 8%. With the recovery of the world economy in the late 1990s, the situation began to improve in West Germany as well. GDP growth increased from only 0.6% in 1996 to more than 3% in 2000. However, economic growth decelerated following the slowdown of the world economy after September 11, 2001, and registered unemployment returned to more than 9% in 2003.

During the period 1984-2003, economic activity shifted especially from the primary and secondary sector to the service sector. The structure of unemployment changed as well. The fraction of unemployed without any professional degree declined constantly from almost 50% in 1984 to 41% in 2003. The share of foreigners increased over time by about 4% to 17% in 2003 with a temporary dip during the post-unification boom. Long-term unemployment has largely moved with total unemployment varying between 26% and 38% in the period 1984-2003.

As shown by Figure 2.1, expenditures on ALMP (training) varied by up to 20% (30%) per year. However, they are only mildly correlated with GDP growth and unemployment (note the different scaling used for ALMP expenditures), because political considerations (e.g. upcoming elections in 1986, 1990, and 1998) and changes in the mix of ALMP instruments (1997, 2003) had strong impacts on ALMP expenditure. The fraction spent on labor market training almost continuously increased from 33% in 1984 to almost 45% in 1998. It dropped slightly afterwards. In 2003, there was a large decline to 30% resulting from a paradigm change in the use of training from longer, more intense programs to short courses with less substantial adjustment of skills. The changes that occurred after 1995 are of limited interest to our empirical study, because we analyze programs that start between 1986 and 1995, only.

Figure 2.1: Selected indicators for business cycle movements in West Germany



Sources: Official statistics published by the Federal Employment Agency, the Institute for Employment Research and the Federal Statistical Office.

Note: The employment rate is calculated as employment plus unemployment as a percentage of the labor force (potential employment). Expenditure on active labor market policies (ALMP) and training are at 1995 prices.

⁴ All numbers presented in this section are taken from official statistics published by the Federal Employment Agency, the Institute for Employment Research and the Federal Statistical Office 1984-2004.

2.2 Unemployment insurance in Germany 1986 to 1995

In Germany, unemployment insurance (UI) is compulsory for all employees with more than a minor employment including apprentices in vocational training.⁵ German UI does not cover self-employed. Persons who have contributed to the UI for at least 12 months within the three years preceding an unemployment spell are eligible for unemployment benefits (UB). The minimum UB entitlement is six months. The maximum claim increases stepwise with the total duration of the contributions in the seven years before becoming unemployed, and age, up to a maximum of 32 months at age 54 or above with previous contributions of at least 64 months. Participation in government-sponsored training counts towards the contribution period for both the acquisition and the duration of UB claims. Actual payment of UB for eligible unemployed is conditional on active job search, regular show-up at the public employment service (PES), and participation in ALMP measures. Since 1994, the replacement rate is 67% of previous average net earnings from insured employment with dependent children and 60% without. Before, replacement rates were 68% and 63%, respectively.

Until 2005, unemployed became eligible for unemployment assistance (UA) after exhaustion of UB. In contrast to UB, UA was means tested and potentially indefinite. However, like UB, UA was proportional to previous earnings but with lower replacement rates than UB (before 1994 58%/56%, thereafter 57%/53% with and without dependent children, respectively).

Unemployed who were ineligible for UB and UA could receive social assistance, which was a fixed monthly payment unrelated to previous earnings, means tested and administered by local authorities.

Note that except for the change in the UB/UA replacement rate, UI institutions have been stable in the period 1986-1995.

⁵ However, civil servants (Beamte), judges, professional soldiers, clergymen and some other groups of persons are exempted from contributions. For further details on the German UI and ALMP, see the comprehensive survey by Wunsch (2005).

2.3 German ALMP 1986 to 1995

ALMP has a long tradition in Germany and among OECD countries expenditure on ALMP is one of the highest (OECD, 2004). With increasing unemployment in the 1980s, the main objective of German ALMP shifted from keeping employment high and fostering economic growth towards reducing unemployment by increasing the employability of jobseekers. The main instruments traditionally used in German ALMP are counseling and job placement services, labor market training, subsidized employment, and support of self-employment.

Training has always been the most important type of program in West Germany. It consists of heterogeneous instruments that differ in the form and intensity of the human capital investment as well as in their respective duration. Durations range from a few weeks to three years. Traditionally, German training courses have the aim of assessing, maintaining, or improving the occupational knowledge and skills of the participant, of adjusting skills to technological changes, of facilitating a career improvement, or of awarding a first professional degree. So called career improvement measures, for which also employed may be eligible, had played a major role before unemployment rose in the 1980s. Since then they became negligible as the focus shifted towards removal of skill deficits and skill mismatch of the unemployed.

In our analysis, we distinguish five types of training. Basic job-search assistance (JSA) existed only until 1992. So-called practice firms (PF) simulate - under realistic conditions - working in a specific field of profession. Short training (ST) with planned duration of up to six months, and long training (LT) with planned duration of more than six months provide a general update or adjustment of skills. Retraining (RT) provides a professional degree equivalent to a degree obtained in the German apprenticeship system. JSA and PF have always been a relatively small program. ST and LT were by far the most important programs with LT gaining importance relative to ST. ST more than doubled its share in the period we consider. RT was relatively small as well, but became more important from the early 1990s

on. However, given its long durations it is the most expensive program so its share in expenditure is substantially larger than its share among participants.

Access to training courses is largely limited to unemployed who are eligible for UB or UA. To underline the character of *further* job related training rather than primary occupational training, eligibility also required holding a first professional degree (before 1994, plus 3 years of work experience) or at least three years of work experience (before 1994, six years). Usually, participants receive a transfer payment, which is called maintenance allowance (MA). Since 1994, MA is of the same amount as UB. Before, MA had been somewhat higher than UB with a replacement rate of 73% with dependent children and 65% without. Moreover, the PES bears the direct cost of the program, and it may cover parts of additional expenses for childcare, transportation, and accommodation.

Note that with respect to eligibility and MA, replacement rates and training regulations have been relatively stable. Moreover, our data allow to control for the few changes that actually occurred, especially with respect to the shifting emphasis on specific types of programs.

3 Data and sample definition

We use the same administrative data sources as Lechner, Miquel, and Wunsch (2005) which combine information from social insurance records on employment, data on benefit receipt during unemployment and information on participation in training programs. The original data covers the period 1980-1997, but employment and unemployment records up to 2003 are added to allow construction of long-run outcome variables. The database is unique in several respects. In particular, it is much more informative than observational data that was available so far, e.g. for the US (e.g. Jacobson, LaLonde and Sullivan, 2004). It is the first micro database that allows analyzing program participation over a sufficiently long time (10 years) on a monthly basis to capture business cycle movements. Moreover, it allows reconstruction of up to 24 years of individual employment histories on a monthly basis, which includes be-

tween 6 and 15 years of pre-program history and 8 years after program start for observing outcome variables. Detailed personal, regional, employer and earnings information of good quality allow to control for all main factors that determine selection into programs (see the discussion in the next section) as well as a precise measurement of interesting outcome variables (e.g. employment status, earnings). Appendix A provides further details on the data.

For our analysis, we use a sample of participants in training and eligible nonparticipants. We focus on the prime-age part (age 20-55) of the West German labor force covered by social insurance (see Appendix A for all details on sample selection). All unemployed who start a program in a particular month in the period 1986-1995 (in total 120 months) are considered participants. In contrast, we define nonparticipants on a monthly basis as recipients of unemployment payments (UB/UA) but not starting a program in that and the following 11 months. We require the latter to ensure that nonparticipants are not too similar to participants with respect to program participation while keeping potential selection bias small. To ensure that we do not use unemployed who completed a program shortly before (potential) program start (are still in an earlier unemployment-participation-unemployment spell), we require that nobody participated in a program in the four years before the (potential) program start we consider. To obtain a sufficient number of participants we pool participants and nonparticipants over a six-month window in the estimation. Thus, we estimate effects for 115 different program starts in the period 1986-1995.

Since these choices may affect our estimation results, we perform an extensive sensitivity analysis with respect to these issues which are detailed in Section 7.

4. Econometrics

We are interested in the mean effects of participating in training in period t (θ_t) for some population of participants (P_t). Varying the latter in an interesting way will be one of the key

issues in the following empirical sections. Based on the usual notation of the evaluation literature, we denote by Y_t^1 the potential outcome of participation in a program, and by Y_t^0 the potential outcome of not participating in a program. Thus, the mean of the effect of the policy for a member of the population of interest, P_t , is given by $\theta_t(P_t) = E(Y_t^1 | P_t) - E(Y_t^0 | P_t)$.

Typically, the population of interest is defined by a combination of the participation status ($D_t = 1$ indicates starting a program in month t) and a subset of the observed covariates (X_t).

P_t may or may not change over time. It includes only unemployed who are eligible for participation.

Since participation and non-participation are not observable for the same individual, the issue of the identification of the effects arises. Lechner, Miquel and Wunsch (2005) as well as Fitzenberger and Speckesser (2005) argue that given the institutional set-up, the newly created data are informative enough, such that a selection on observables strategy (the conditional independence assumption, CIA) identifies the effects conditional on treatment status and covariates. In particular, we obtain expressions for the mean potential outcomes conditional on covariates that are functions of participation status, observable outcomes (Y_t), and covariates only:

$$E(Y_t^{d_t} | D_t = d_t', X_t = x_t) = E(Y_t | D_t = d_t, X_t = x_t), \quad d_t, d_t' \in \{0, 1\}.$$

This equation holds for all values of x_t that are of interest.

As argued in Lechner, Miquel, and Wunsch (2005), selection into programs is determined by three main factors: (i) eligibility, (ii) selection by caseworkers and (iii) self-selection by potential participants. Eligibility is ensured by the construction of our sample (see Appendix A.2 for details). Caseworkers select participants based on individual employment prospects and corresponding skill deficits, chances for successful completion of a program and conditions

on the local labor market. For the unemployed a strong incentive to participate is the potential renewal or extension of unemployment benefit claims.

Our data allow to reconstruct between 6 and 15 years of individual pre-program employment histories on a monthly basis, and it contains detailed personal, regional, employer and earnings information of good quality (consult the internet appendix for a complete list of variables). Moreover, we are able to construct initial and remaining benefit claims from the data. Thus, it allows controlling for all main factors that determine selection into programs (see also the internet appendix for a detailed discussion of the validity of the CIA in our data).

In fact, since in most cases considered below we interpret the changes of the effects over time, any violation of the conditional independence assumption that leads to a bias that does not change over time would not hurt our main conclusions in this paper.

Given identification of the quantities mentioned above, under the usual assumptions a matching strategy identifies our parameters of interest, because

$$\theta_t(P_t) = \int E(Y_t \mid D_t = 1, X_t = x) f_{X_t|P_t}(x) dx - \int E(Y_t \mid D_t = 0, X_t = x) f_{X_t|P_t}(x) dx.$$

$f_{X_t|P_t}(x)$ denotes the distribution of X_t in the population P_t . In the next section, we call $f_{X_t|P_t}(x)$ the target population towards which the distributions of X_t for participants and nonparticipants are adjusted. An example of such a target population would be the participants in period t . In this case, we would estimate the average treatment effect on the treated (ATET). Alternatively, another popular choice would be the population of participants and nonparticipants in t , leading to θ_t being the average treatment effect (ATE).

Having established identification of the effects, the question of the appropriate estimator arises. All possible parametric, semi- and nonparametric estimators are (implicitly or explicitly) built on the principle that for every comparison of two programs and for every participant

in one of those programs we need a comparison observation from the other program with the same characteristics regarding all factors that jointly influence selection and outcomes. Here, we use propensity score matching estimators to produce such comparisons. An advantage of these estimators is that they are essentially nonparametric and that they allow arbitrary individual effect heterogeneity (see Heckman, LaLonde, and Smith, 1999; Imbens, 2004, provides an excellent survey of the recent advances in this field). All details of the estimator are relegated to Appendix B.

5 The program effects over time

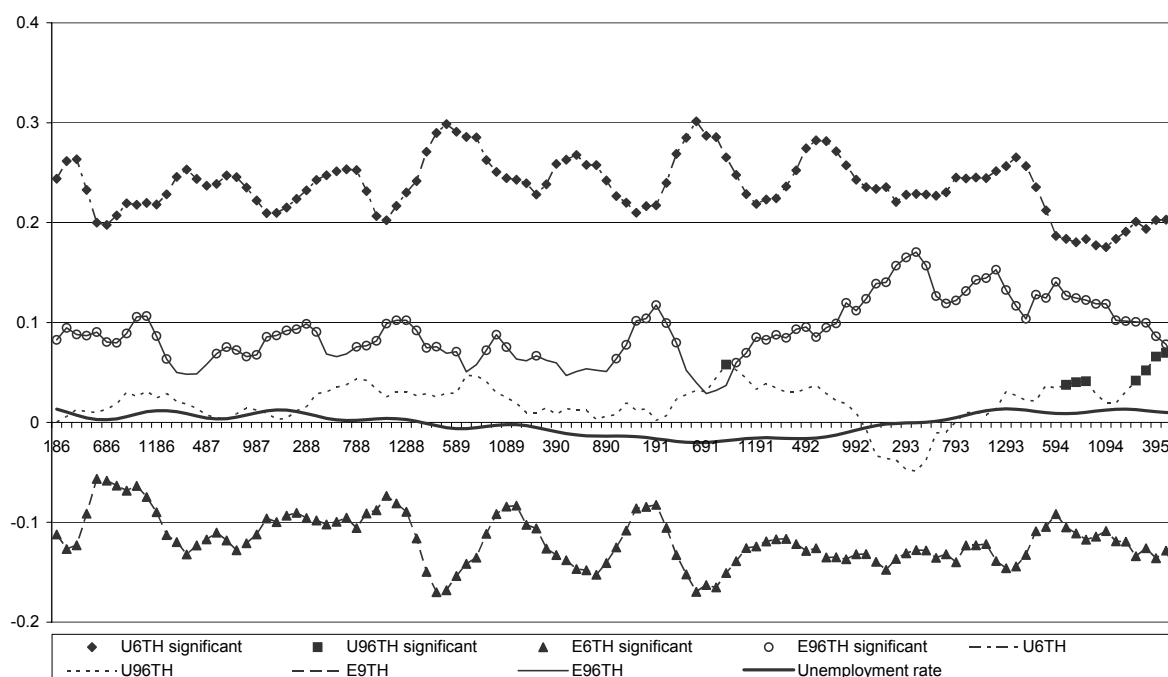
According to German legislation, the most important objectives of active labor market policy are to increase reemployment chances and to reduce the probability to remain unemployed. Therefore, we use outcome variables related to the employment status, in particular registered unemployment and employment subject to social insurance.⁶ We also consider gross earnings as a crude measure for individual productivity.

All effects are measured from the month of the (potential) program start on. Focusing on the beginning instead of the end of the programs rules out that programs appear to be successful, just because they keep their participants busy by making them stay in the program. We consider a program most successful if everybody would leave for 'good' employment immediately after starting participation. Whenever a person participates in a program, he is considered as registered unemployed (and not employed). We also consider a total effect, i.e. the cumulated effects of the program from its beginning to the respective point of measurement. Appendix A.3 contains further details on how the outcome variables are constructed.

Since the effects measured for the different outcome variables appear to be in line with each other, the main body of the text presents results for the outcome variables *registered unemployment* and *employment* only. Detailed information for all other outcomes is relegated

to the internet appendix. We measure these outcomes at different distances to program start for a better understanding of the dynamic evolution of the effects over time. Usually, we expect the program to begin with a negative lock-in effect before the effect reaches its long-run level. The lock-in effect is approximated by the effect after 6 months. The long-run effect is approximated by the effect after 8 years. However, the effects appear not to change too much after about 3 years.

Figure 5.1: Effect of training on the employment and unemployment probabilities of participants



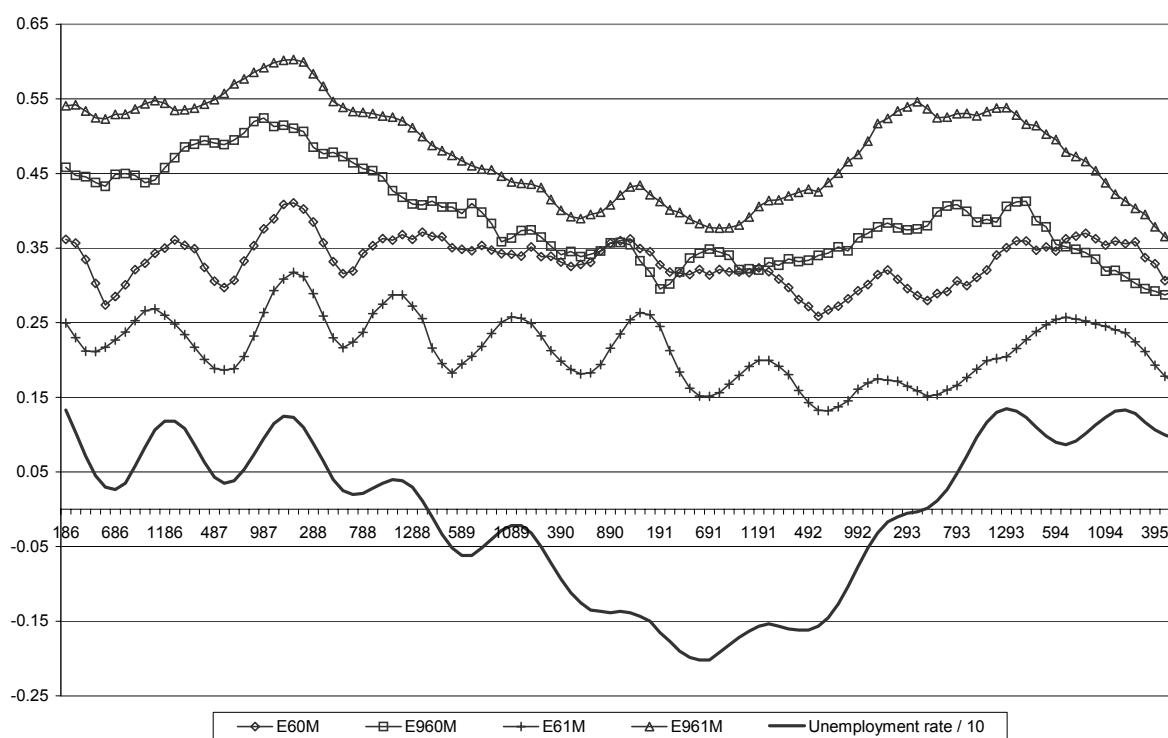
Note: The outcome variables are named as follows: U: unemployment, E: employment, 6 or 96: month after program start, TH: theta (average treatment effect on the treated). For each outcome variable, dots appear if the effect is significant at the 5% level in a particular month. The unemployment rate is presented net of its mean 1986-1995. All effects are smoothed using three-month moving averages.

Figure 5.1 shows the short-run and long-run effects of training for each starting month in the period January 1986 to July 1995. We find that after 6 months, programs increase the unemployment probability by about 25%-points for participants, and, correspondingly, reduce the employment probability by about 15% points. In the long run, employment is increased by about 10% points, but any effect on unemployment is hard to spot (if there is any, then unem-

⁶ Here 'registered unemployment' is defined as receipt of UB or UA or participation in training.

ployment is increased). Thus, the program effect operates by increasing employment at the expense of the share of unemployed leaving the labor force.⁷ Considering the effects on earnings (nonemployment is counted as zero), we find similar effects with an average long-run monthly earnings gain of about 100 EUR. Although all effects show considerable variation over time, it is hard to spot any relation with the unemployment rate, which is shown in Figure 5.1 as well (for a better exposition, it is presented net of its mean over the 115 months presented in the table).

Figure 5.2: Mean employment rates of participation and nonparticipation for participants



Note: The outcome variables are named as follows: E: employment, 6 or 96: month after program start, 0: nonparticipation, 1: participation, M: mean level. The unemployment rate is presented net of its mean 1986-1995. All effects are smoothed using three-month moving averages.

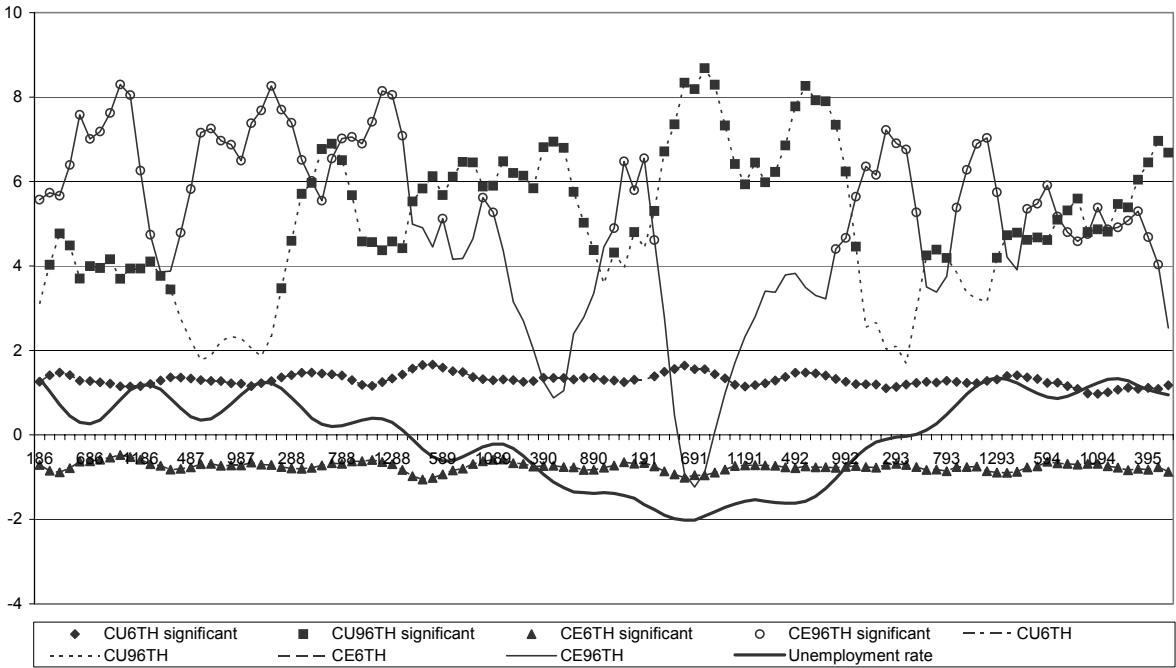
Figure 5.2 shows the estimates for the mean of the potential outcome variable employment that underlies the corresponding effect estimates in Figure 5.1. The short-run outcomes show

⁷ These findings are largely consistent with the studies analysing the effects of post 1992 training programs with these data (i.e. Fitzenberger and Speckesser, 2005; Fitzenberger, Osikominu and Völter, 2006, and Lechner, Miquel, Wunsch, 2005; but note the different definitions of participation and nonparticipation in these studies).

a clear seasonal effect (at least for the first 8 years), whereas, not surprisingly, such relation does not appear for the long-run effects.

Finally, Figure 5.3 shows the cumulated effects in months of (un)employment over time. They imply that the total negative effect in the first 6 months after program start corresponds to a reduction of about 1.5 months of employment as well as an additional month of unemployment. In the long run, there appears to be a gain of about 4-6 months of employment and an additional 4-6 months of unemployment (!), which again suggests that the programs reduce the share of people leaving the labor force drastically. Comparing the cumulated effects with a particular point-in-time estimate after treatment, we find very similar shapes of the effects over time, although obviously the magnitudes and sampling uncertainty differs.

Figure 5.3: Cumulated effects of training on the employment and unemployment probabilities of participants (in months)



Note: The outcome variables are named as follows: CU: cumulated unemployment, CE: cumulated employment, 6 or 96: month after program start, TH: theta (average treatment effect on the treated). For each outcome variable, dots appear if the effect is significant at the 5% level in a particular month. The unemployment rate is presented net of its mean 1986-1995. All effects are smoothed using three-month moving averages.

The next step is to condense the dynamic information about the effects and check their correlation with indicators for the economic development more thoroughly. In Table 5.1, we show the correlation of the effects presented, including earnings, as well as the effects measured after 3 and respectively 6 years after program start, with the quarterly GDP growth rate, the monthly unemployment rate, and the monthly number of participants in training programs. The internet appendix presents correlations with the estimated means of the potential outcomes as well.⁸

The results suggest that the programs are more effective when unemployment is higher at the time when the program starts. This positive dependence of the program effect on the unemployment rate is somewhat larger for the long-run effects than for the lock-in effects. If these correlations have a causal interpretation, their magnitudes imply for example that on average the employment effect of the programs increases by about 0.7-1.8% points when the national unemployment rate is increased by 1%-point (depending on the point in time after program start when the outcome is measured). When we change the perspective and correlate the effects with the unemployment rate in the month when the outcome is actually measured, then not surprisingly the correlations change because unemployment is measured at a later point in the economic cycle. Therefore, the magnitude of the change depends on the distance. Typically, the correlations for the lock-in effects get somewhat smaller, whereas those of the long-term effects change sign. However, the correlation with the unemployment rate at the time of outcome measurement has only limited appeal in a policy sense, because that information is unknown at the time of the participation decision and therefore hard to use to improve the

⁸ The significance levels of the correlations are obtained from a bivariate regression of the effects on a constant and the respective macroeconomic indicator using the Newey-West procedure to correct for the correlation of the program effects over time. The significance level presented corresponds to the two-sided t-test that the coefficient on the unemployment rate is zero.

training policy.⁹ Finally, the quarterly GDP figures appear to be too rough to detect any correlation. Similarly, no systematic correlation can be detected with indicators of program size, like the number of participants.

Table 5.1: Correlation of the program effects with indicators for the macroeconomic situation in %

Outcome	Unemployment rate at program start	Unemployment rate at outcome measurement	Quarterly GDP growth rate	# of participants in training programs	
Unemployment	6 months after prog. start	-43**	-33*	3	19
	3 years after prog. start	-36*	21	8	10
	6 years after prog. start	-27*	24*	15	21
	8 years after prog. start	-1	26	17	17
Employment	6 months after prog. start	25*	5	8	-1
	3 years after prog. start	45**	-45**	2	-3
	6 years after prog. start	43**	-33**	-3	-33**
	8 years after prog. start	31**	-47**	-12	-50**
Monthly earnings	6 months after prog. st.	20	1	7	7
	3 years after prog. start	48**	-58**	5	-2
	6 years after prog. start	53**	-43**	7	-29*
	8 years after prog. start	47**	-50**	1	-40**
Cumulated unemployment	6 months after	-43**	-43**	8	15
	3 years after prog. start	-65**	-27**	20*	24
	6 years after prog. start	-57**	19	16	20
	8 years after prog. start	-50**	27	17	22
Cumulated employment	6 months after	20	20	6	9
	3 years after prog. start	47**	-14	-6	2
	6 years after prog. start	50**	-43**	-2	-10
	8 years after prog. start	52**	-37**	-4	-22
Cumulated earnings	6 months after p.s.	13	13	8	17*
	3 years after prog. start	46**	-18	-2	5
	6 years after prog. start	51**	-51**	4	-5
	8 years after prog. start	56**	-48**	4	-15

Note: The unemployment rate at outcome measurement is the rate measured in the respective month after program start. For the cumulated outcomes, the unemployment rate at outcome measurement is the average unemployment rate over the respective period.

Newey-West autocorrelation-robust t-values: ** significant at the 1% level, * significant at the 5% level.

In the remainder of the paper, we will try to gain more insights on why there is such a positive correlation between the effectiveness of the programs and labor market conditions as characterized by the monthly unemployment rate.

⁹ When both unemployment rates are simultaneously included in the regression, for the long-run effects both coefficients are typically significant. The coefficients have about the same sign and magnitude as in the

6 The changing composition of program participants and programs

6.1 Participants

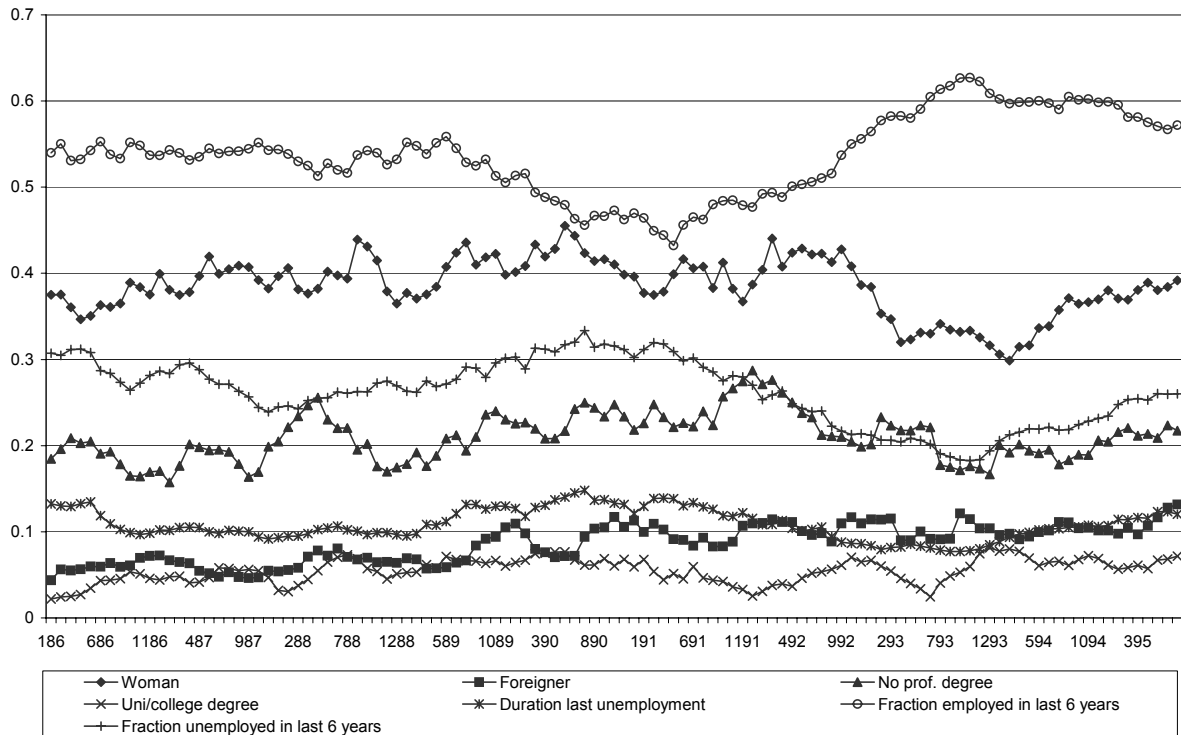
The key question raised by the previously found relation between the effects and the state of the labor market is whether these correlations reflect the fact that the same programs have different effects (different production functions) depending on the state of the economy or whether the correlations are spurious. A spurious correlation could be induced by some other background factor moving the effects in a similar direction as the unemployment rate. Therefore, it is important to 'eliminate' other potentially important factors that change over time, and affect program effectiveness.

The first such potential factor relates to the dependence of the pool of potential participants from which the actual participants are selected on the state of the economy. In a recession, there might be excess supply of unemployed who would benefit from the programs. When the economy recovers fewer of them would be available, but program places still have to be filled (for example because there is a rigidity in the adjustment of the supply of courses due to long-run contracts between the federal employment agency and suppliers). Figures 6.1 and 6.2 show the changes of the composition of participants and nonparticipants over time for some selected characteristics. We see that both groups change, and that they change in a similar fashion. In more detail, the share of women, the employment histories and the education levels fluctuate, whereas the share of foreigners increases more or less continuously.

The impression that the change in the characteristics of participants over time merely reflects changes in the supply of unemployed is confirmed as well by looking at the monthly probit models for program participation that do not show any large difference in the conditional selection model over time (for detailed probit estimates see the internet appendix).

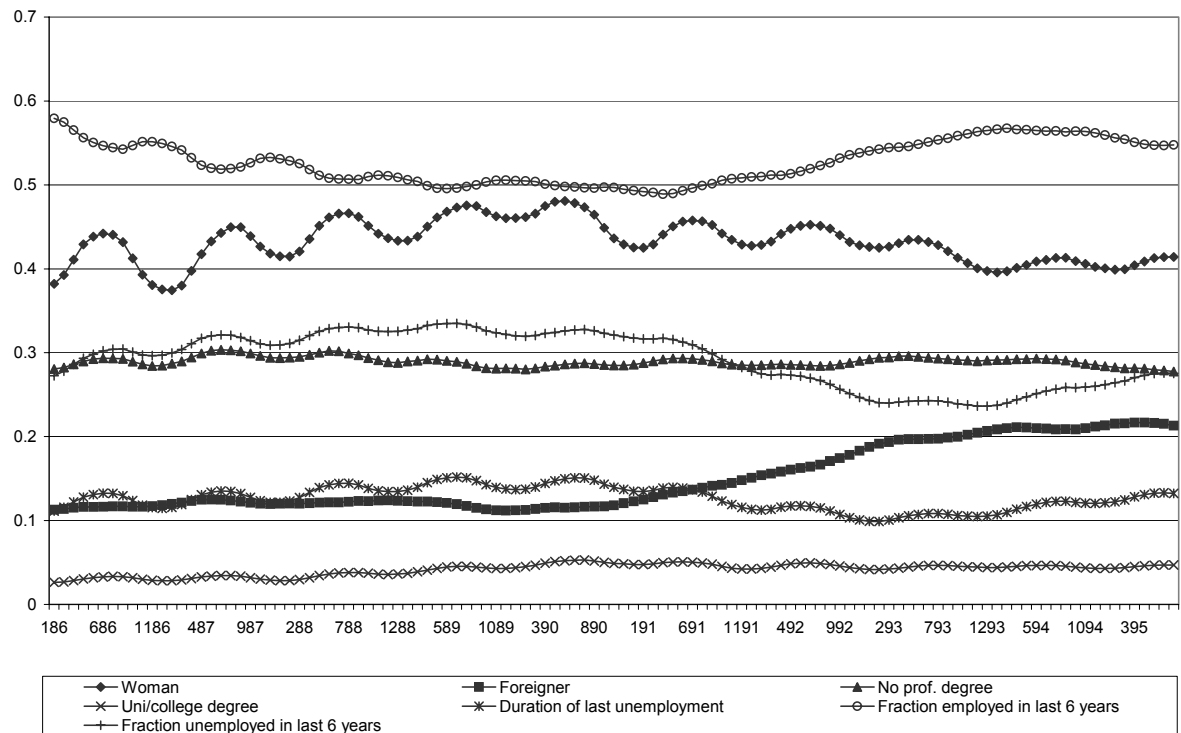
bivariate regressions. This feature remains for the specifications to be discussed in the next sections. For the lock-in effects, the two rates are almost collinear.

Figure 6.1: Composition of participants over time – means of selected variables



Note: Mean of the respective variable in the population of participants. Six month moving averages (to align figures with the pooling of participants in the estimation).

Figure 6.2: Composition of nonparticipants over time – means of selected variables



Note: Mean of the respective variable in the population of nonparticipants. Six-month moving averages (to align figures with the pooling of nonparticipants in the estimation).

A key question that remains is whether these changes in the composition of program participants are correlated with the situation in the labor market as well. Table 6.1 shows that this is indeed the case.

Table 6.1: Correlation of the characteristics of participants with unemployment rate in %

Characteristics of program participants	Unemployment rate at program start
Woman	-52**
Foreigner	-24*
No professional degree	-67**
University/college degree	7
Duration of last unemployment spell	-51**
Fraction of months employed in the last 6 years	82**
Fraction of months unemployed in the last 6 years	-46**

Note: Correlation of monthly mean of respective variable (six-month moving average) with the unemployment rate.

Newey-West autocorrelation-robust t-values: ** significant at the 1% level, * significant at the 5% level.

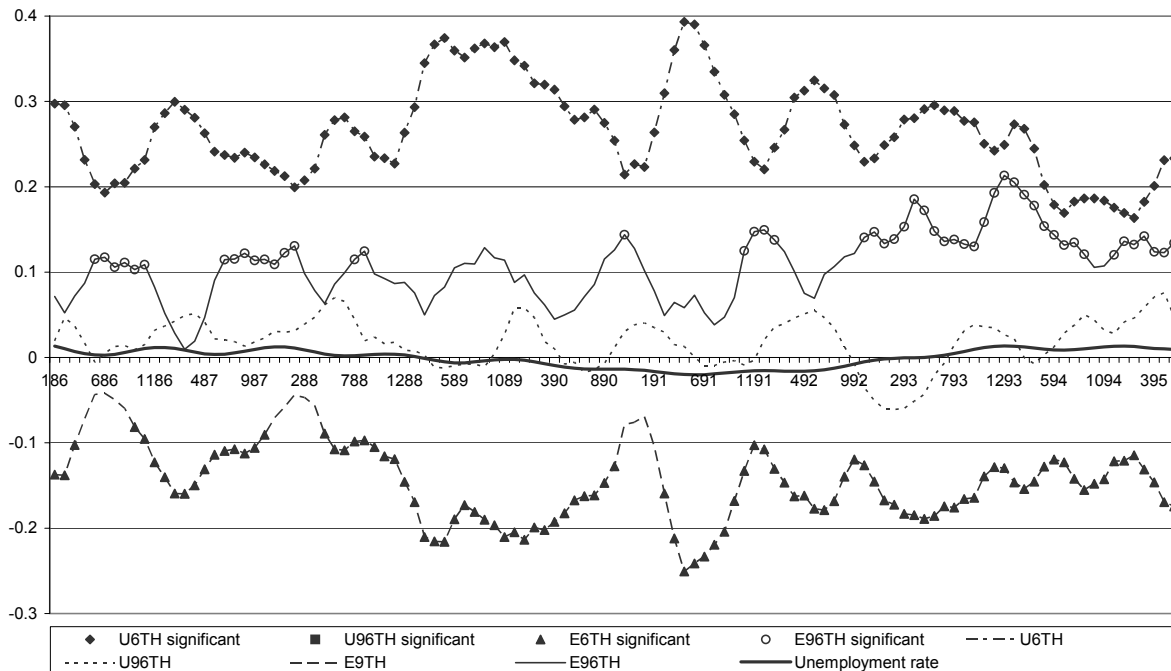
Keeping in mind that current unemployment rates are likely to be negatively correlated with average unemployment rates in the last six years, the negative correlation between unemployment in the past and the positive correlation with past employment is expected. However, participation of women, foreigners, and unemployed with lower education is also lower during times of higher unemployment.

To the extent that there is effect heterogeneity, a fact that is documented in numerous evaluation studies (for West Germany e.g. Lechner, Miquel, and Wunsch, 2005), such systematic relationships between the state of the labor market and the characteristics of participants might influence the correlation with the effects as well. Therefore, Figure 6.3 shows the effects of the training programs for a fixed population of participants. This population is defined as having the average characteristics of the overall population of participants in the period 1986-1995, reduced to the intersection of all common supports over time. That is, more technically speaking, we define a target population of participants with comparable participants and nonparticipants in all months.¹⁰ Month by month, we match participants as well as

¹⁰ Out of 9418 participants in the reference population, only 2101 (22%) fulfil this criterion.

nonparticipants with respect to that target distribution. Since the target distribution is the same for all periods, characteristics of the participants are held constant in the estimation of the effects of training.¹¹

Figure 6.3: *Effect of training on the employment and unemployment probabilities of participants (stable characteristics of participants)*



Note: The outcome variables are named as follows: U: unemployment, E: employment, 6 or 96: month after program start, TH: theta (average treatment effect on the treated). For each outcome variable, dots appear if the effect is significant at the 5% level in a particular month. The unemployment rate is presented net of its mean 1986-1995. All effects are smoothed using three-month moving averages.

Albeit somewhat larger, the results appear to be similar to those for the specification that allows the characteristics of the participants to vary over time. Particularly when we take into account that due to reduced sample size coming from the far more restrictive common support requirement, sampling uncertainty is somewhat larger. Checking the correlation of the effects that result from this specification with different indicators of the macroeconomic situation, it

¹¹ By defining the characteristics used in matching, we carefully avoid that they depend on time or a function of it (e.g. we capture different regional labor market states not by different unemployment rates but by the regional deviation from the national mean at that time).

turns out that, if anything changes, then the correlations increase. Table 6.2 shows the exact values of those correlations for the unemployment rates and selected outcome variables.

Table 6.2: Correlation of the program effects with the unemployment rate in %

(stable characteristics of participants)

Outcome	Unemployment rate at		Previous specification:		
	program start	outcome measurement	program start	outcome measurement	
Unemployment	6 months after prog. start	-49**	-45**	-43**	-33*
	3 years after prog. start	-48**	19	-36*	21
	8 years after prog. start	19	15	-1	26
Employment	6 months after prog. start	36**	24	25*	5
	3 years after prog. start	45**	-56**	45**	-45**
	8 years after prog. start	31*	-30**	31**	-47**
Monthly earnings	6 months after pr. start	40**	26	20	1
	3 years after prog. start	44**	-66**	48**	-58**
	8 years after prog. start	53**	-27*	47**	-50**

Note: The unemployment rate at outcome measurement is the rate measured in the respective month after program start. Newey-West autocorrelation-robust t-values: ** significant at the 1% level, * significant at the 5% level.

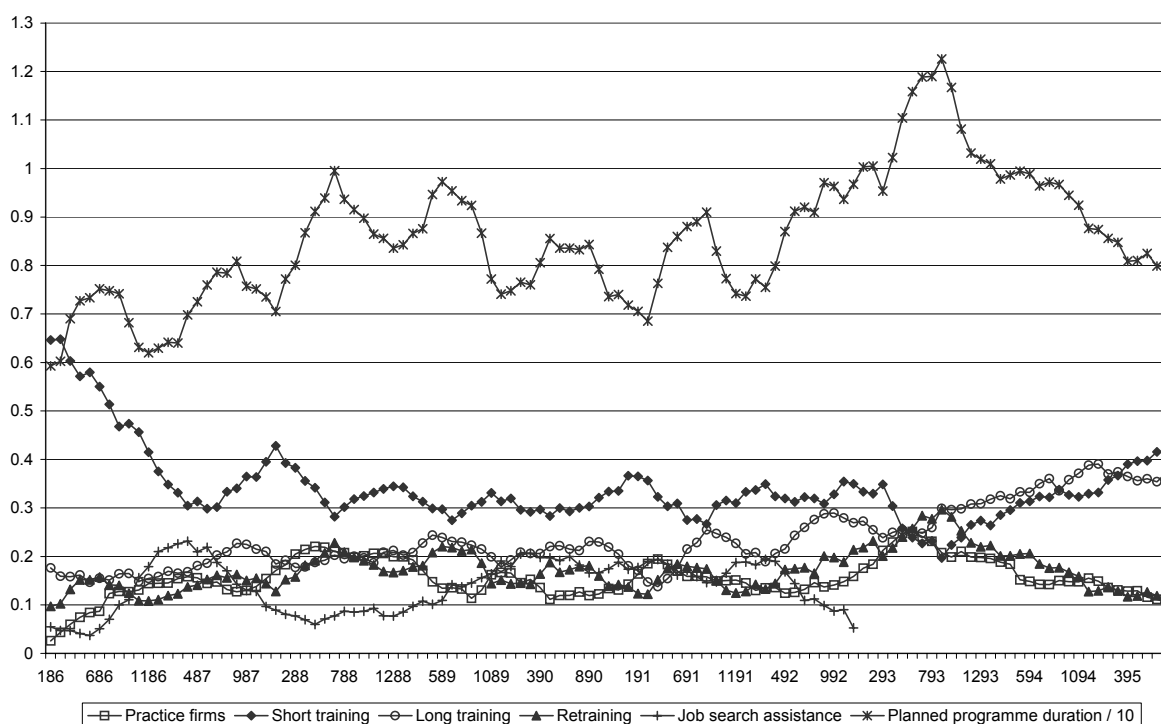
6.2 Programs

Figure 6.4 shows that other import factors that change over time are the composition of the training policy and average planned durations of the training courses.¹²

Long training increases over time, whereas the job search assistance programs were terminated after 1992. The shares of the other program groups fluctuate in an unsystematic matter. Similarly, the planned program duration of all participants fluctuates considerably. It reaches its peak of more than 12 months for programs beginning in the second part of 1993, where the rather long retraining courses have been used quite extensively. The lowest level of about 6 months appears in 1986, where short training was most important.

¹² This figure is based on the participants used in Section 5. The plot for the target population defined in Section 6.1 is very similar and therefore relegated to the internet appendix. That appendix shows also a plot of the program type specific planned durations.

Figure 6.4: Shares of different programs and average planned duration



Note: Units of measurement: Shares for the different programs; months /10 for program duration.

Table 6.3: Correlation of the characteristics of the training policy with the unemployment rate in %

Characteristics of programs	Unemployment rate at program start	
	Participants	Stable participants
Fraction of participants in practice firms	5	-4
Fraction of participants in short training	28**	26**
Fraction of participants in long training	34**	25**
Fraction of participants in retraining	-1	6
Fraction of participants in job search assistance	-42**	-27*
Planned program duration	6	1
Planned duration of practice firms	-34**	20
Planned duration of short training	-20*	27**
Planned duration of long training	-13	6
Planned duration of retraining	-21*	-54**
Planned duration of job search assistance	34**	46**

Note: Correlation of monthly mean of respective variable (six-month moving average) with the unemployment rate.

Participants are those participants used in Section 5, whereas stable participants are those used in Section 6.1.

Newey-West autocorrelation-robust t-values: ** significant at the 1% level, * significant at the 5% level.

As before, the key question is whether these changes are related to labor market conditions as well. Table 6.3 shows the correlation of those variables with the unemployment rate. These results suggest that this correlation exists, at least for short and long training (positive) and job

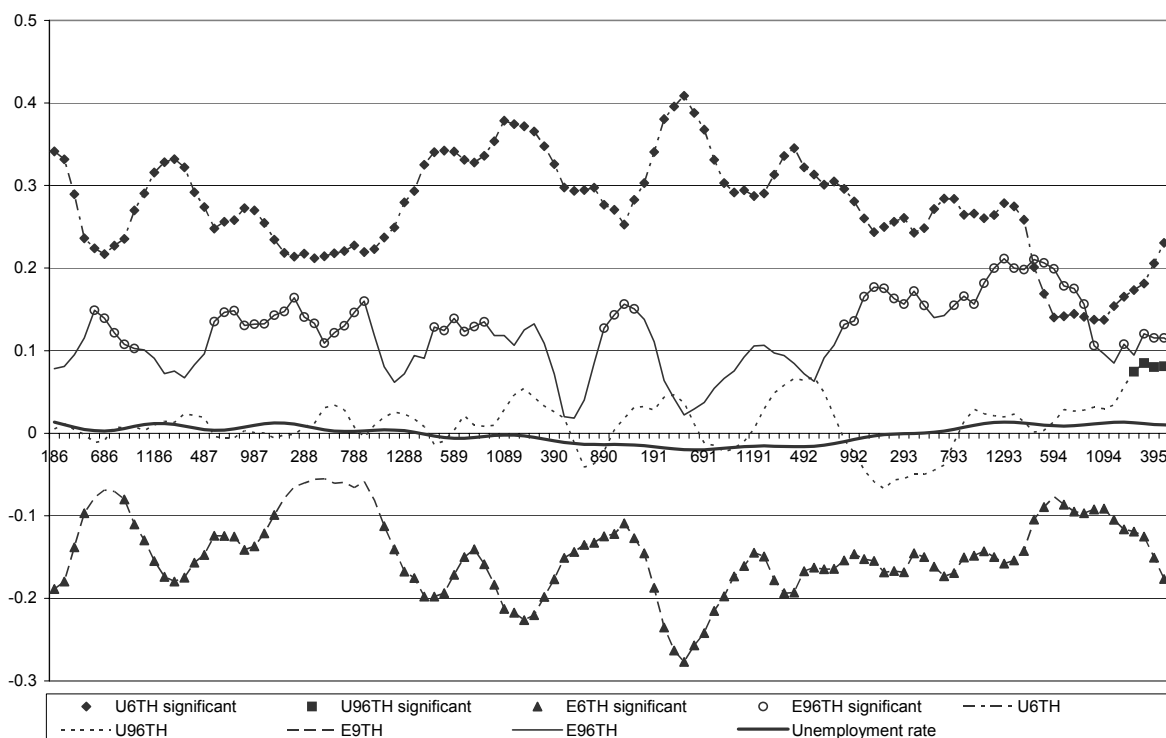
search assistance (negative). This finding holds for all participants as well as those used in the previous section.

To take out the effects of changing program shares and planned durations over time, we keep the characteristics of participants (as in the previous section) as well as the program shares and planned program durations constant over time by following exactly the same approach as described in Section 6.1. We also add the type of program as well as its planned duration as additional matching variables for program participants (obviously, nothing changes for nonparticipants). Figure 6.5 shows the results. They are based on a population of participants with an average duration of programs of 9.4 months (standard deviation is 7.5 months). 46% of those participants take part in short training, 34% in long training, and 20% in retraining. Participants receiving job search assistance are omitted because the program is terminated after 1992. Since participants in job search assistance are no longer part of the reference population to which participants and non-participants are matched, they are removed from the sample and estimates of the effects for this program do not appear in this figure.

Although the effects seem to be somewhat larger than in the previous specifications, these changes are most likely within a range that could be attributed to sampling error. Analyzing the correlations of the effects with the unemployment rate (Table 6.4), we find that, at least for employment, the correlations increase further compared to the previous two specifications.

Overall, we conclude that keeping the characteristics of the participants and the training policy constant, reaffirms the findings of the previous section that programs are more effective when unemployment is high.

Figure 6.5: Effect of training on the employment and unemployment probabilities of participants
(stable characteristics of participants, program types, and programme durations)



Note: The effects are named as follows: U: unemployment, E: employment, 6 or 96: month after program start, TH: theta (average treatment effect on the treated). For each outcome variable, dots appear if the effect is significant at the 5% level in a particular month. The unemployment rate is presented net of its mean 1986-1995. All effects are smoothed using three-month moving averages.

Table 6.4: Correlation of the program effects with the unemployment rate %

(stable characteristics of participants, program types, and programme durations)

Outcome		Unemployment rate at		Previous specification:	
		program start	outcome measurement	Unemployment rates at program start	
				participants constant	participants change
Unemployment	6 months after prog. start	-58**	-58**	-49**	-43**
	3 years after prog. start	-39**	28*	-48**	-36*
	8 years after prog. start	3	2	19	-1
Employment	6 months after prog. start	46**	37*	36**	25*
	3 years after prog. start	31*	-56**	45**	45**
	8 years after prog. start	40**	-22	31*	31**
Monthly earnings	6 months after pr. start	52**	42**	40**	20
	3 years after prog. start	37*	-63**	44**	48**
	8 years after prog. start	55**	-13	53**	47**

Note: The unemployment rate at outcome measurement is the rate measured in the respective month after program start. Newey-West autocorrelation-robust t-values: ** significant at the 1% level, * significant at the 5% level.

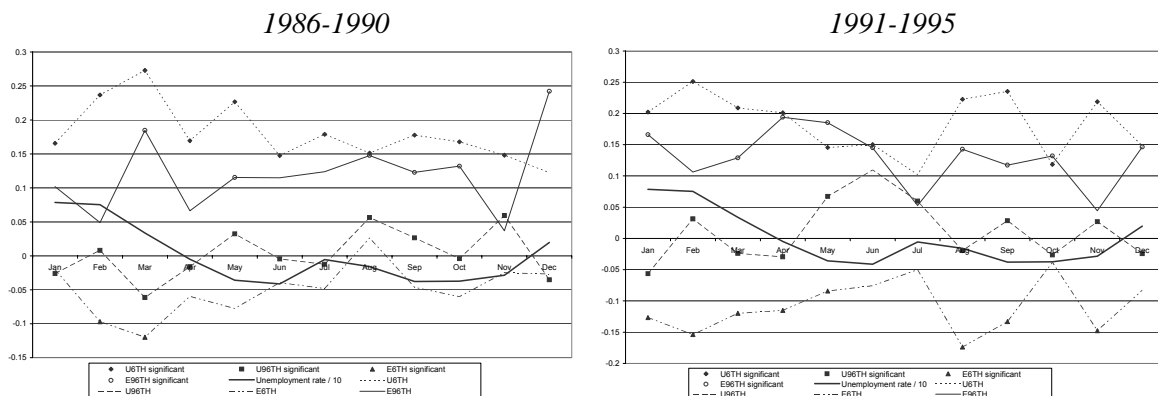
7 Sensitivity analysis

7.1 Seasonal patterns

Visual inspection of the effect estimates in the various specifications presented before may suggest that the correlation with the unemployment rate is merely a reflection of some seasonal variation, instead of a more long-term macroeconomic trend. To understand whether this may be a valid interpretation, we analyze the seasonal pattern of the effects directly.

To do so, we aggregate monthly participation along the starting month of the programs over the years and consider the effects conditional on that particular starting month leading to twelve subsamples for which we estimate the effects (January to December). Since visual inspection suggests a different seasonal pattern for the first and the second half of the sample, those two periods are considered separately. As for the estimation of the development over time, three different specifications are estimated. Figure 7.1 presents the results for the specification with constant population and program shares. Since the results for the other two specifications are similar, their results are relegated to the internet appendix.

Figure 7.1: Seasonal effects of training on the employment and unemployment probabilities of participants (stable characteristics of participants, program types, and durations)



Note: The effects are named as follows: U: unemployment, E: employment, 6 or 96: month after program start, TH: theta (average treatment effect on the treated). For each outcome variable, dots appear if the effect is significant at the 5% level in a particular month. The unemployment rate is presented net of its mean in each time window. Unemployment rate is aggregated for particular month over the respective time window.

It is hard to detect any systematic pattern in Figures 7.1, which is probably related to the fact that the variation in the unemployment rate over the year is rather small. This view is confirmed by considering the correlations between the effects and the unemployment rate given in Table 7.1. In that table, which is based on 12 months and thus only 12 data points, all correlations are insignificant. Furthermore, it is very hard to detect any systematic pattern in these correlations.

*Table 7.1: Correlation of the program effects with the unemployment rate in %
(participants and program compositions do not change over time)*

Outcome	Average monthly unemployment rate program start	
	1986-1990	1991-1995
Unemployment 6 months after prog. start	45	39
3 years after prog. start	11	-11
8 years after prog. start	-46	-29
Employment 6 months after prog. start	-45	-32
3 years after prog. start	-26	-2
8 years after prog. start	6	-5
Monthly earnings 6 months after pr. start	-61	-25
3 years after prog. start	-18	-8
8 years after prog. start	-5	-12

Note: 12 observations for each cell. All correlations are insignificant at the 5% level.

Therefore, we conclude that seasonal correlation cannot be an important part of the explanation for the correlation between the effects and labor market conditions.

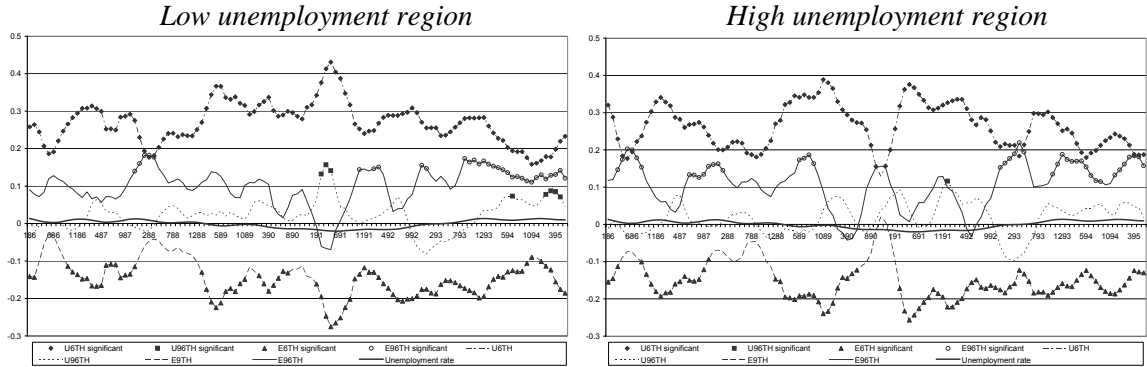
An additional check of our results is to use the six-month rolling window and base the seasonal analysis on the first month of that window, as this corresponds most closely to the procedure used in Section 6. Again, the results do not suggest that the above found correlation of the effects with monthly unemployment rates may be related to seasonal effects (for detailed results see the internet appendix).

7.2 Regional variation

If it is true that the effectiveness of the training programs increases with unemployment, then one should expect that programs are more effective in regions with higher unemployment than

in regions with lower unemployment. Therefore, the analysis is done separately for high and low unemployment regions. Of course, just comparing effects in regions with low and high unemployment is not satisfactory, because many aspects of the labor market that might influence program effects, like the industry structure and the characteristics of the participants differ across local labor market. To minimize these problems, we use a similar matching strategy as before. We specify the same target distribution of characteristics of participants in both subsamples and then perform matching as before. Since splitting the sample increases the noise in our estimates considerably, we choose an overlapping split (60% of all unemployed facing the lowest regional unemployment rates vs. 60% of those unemployed facing the highest ones).¹³ Figure 7.2 shows the corresponding results.

Figure 7.2: Employment and unemployment probabilities of participants in regions with low and high unemployment (stable characteristics of participants, program types, and durations)



Note: The effects are named as follows: U: unemployment, E: employment, 6 or 96: month after program start, TH: theta (average treatment effect on the treated). For each outcome variable, dots appear if the effect is significant at the 5% level in a particular month. The unemployment rate is presented net of its mean 1986-1995. Unemployment rate is national rate normalized as before.

Again, it is difficult to spot the difference between the two plots. In other words, our test has not much power, because splitting the sample led to an increased variability of the estimates. However, when we consider the mean effects over time we find that the long-term effects are indeed lower in regions with lower unemployment, albeit only by about 1%-point for em-

¹³ We used a classification in terms of deviation of the local unemployment rate from its 10-year mean to rule

ployment. Table 7.2 shows that within the two types of regions, the effects are again positively correlated with the development of unemployment over time.

Table 7.2: Correlation of the program effects with the unemployment rate % for regions with low and high unemployment (stable characteristics of participants, program types, and durations)

Outcome	Monthly unemployment rate program start		
	Low UE rate regions	High UE region	
Unemployment	6 months after prog. start	-56**	-37**
	3 years after prog. start	-47**	-37**
	8 years after prog. start	-13	-14
Employment	6 months after prog. start	38**	15
	3 years after prog. start	40*	36*
	8 years after prog. start	49**	46**
Monthly earnings	6 months after pr. start	48**	27
	3 years after prog. start	44**	39**
	8 years after prog. start	59**	60**

Note: Low / high UE rate regions are those where 60% of the unemployed facing the lowest / highest regional unemployment rate live.

Newey-West autocorrelation-robust t-values: ** significant at the 1% level, * significant at the 5% level.

A note of caution is however appropriate, because our sample sizes for participants are clearly too small for a credible regional analysis that should disaggregate the regions much more to arrive at a more powerful result.

7.3 Stability of the correlation between the effects and unemployment over time

There may be the concern that the relation between unemployment and program effects holds only for the early participants and is no longer relevant for recent programs. Therefore, we repeat our correlation analysis for the first and second half of the ten year period to see whether the correlations between the effects of the programs and the labor market conditions remain constant before and after German unification.

The findings shown in Table 7.3 confirm again that short and long-term employment and earnings outcomes are positively related to the unemployment rate. However, one of the measures for medium-run outcomes (3 years after program start) is large and significant in the

out conditioning on the business cycle.

first period, whereas the correlation for the other long-run outcome (8 years after program start) is large and significant in the second period. We conjecture that this appears most likely because of the additional sampling uncertainty coming from reducing the sample by half.

Table 7.3: Correlation of the program effects with indicators for the macroeconomic situation in % for the first and second half of the sample (stable characteristics of participants, program types, and durations)

Outcome	Monthly unemployment rate at				
	program start		outcome measurement		
	1986-1990	1991-1995	1986-1990	1991-1995	
Unemployment	6 months after training	-38*	-75**	-44**	-65**
	3 years after training	-70**	-20	53**	-24
	8 years after training	2	6	20	33
Employment	6 months after training	33*	55**	38*	40*
	3 years after training	62**	4	-48**	5
	8 years after training	23	56**	10	-29
Monthly earnings	6 months after training	41**	62**	43**	48*
	3 years after training	55**	19	-63**	20
	8 years after training	29	68**	8	-38*

Note: The last month in the first period is September 1990. The first month in the second period is October 1990. The unemployment rate at outcome measurement is the rate measured in the respective month after program start. Newey-West autocorrelation-robust t-values: ** significant at the 1% level, * significant at the 5% level.

7.4 Further sensitivity checks

This section summarizes further checks to improve the credibility of our key result that the effects of the training programs are positively correlated with the unemployment rate over time. For the sake of brevity, all the details are relegated to the internet appendix.

Before discussing the different checks, the reader should be aware of the limitations of the data. Given that we are interested in the dynamic evolution of the effects in relation to the starting dates of the program, the sample sizes for participants quickly become too small to powerfully check for individual heterogeneity. For example, it would be interesting to investigate the correlations of the effects of the different types of programs with the labor market conditions. Clearly, there are not enough observations for practice firms, retraining and job search assistance, but even the estimates for the larger groups of short and long

training courses are too noisy to allow any firm conclusions. In a similar vein, it is not possible to investigate the issues of participant subgroup heterogeneity much further.

A crucial issue that comes up in our implementation is how observations are aggregated over time. For each month, the results above are based on the participants and nonparticipants of one and the next five months. For a sensitivity check that window is, first, reduced to four months and, second, increased to nine months. The results are detailed in the internet appendix. Qualitatively, the results do not change, but, again, in the first case sample size becomes an issue. Thus, in the first case the precision of the estimated coefficients is reduced whereas in the second case precision increases. With respect to the correlation of the effects with the unemployment rate over time, we find somewhat smaller (larger) correlations when the pooling window is reduced (increased) but the overall conclusions do not change.

Furthermore, there might be an issue on how a nonparticipant is defined (see the papers by Frederickson and Johansson, 2003, 2004, and Sianesi, 2004). In all results presented above, nonparticipants are required not to participate for the 12 months following and including their potential program start. We checked the sensitivity of our results by requiring 6 (24) months instead which reduces (increases) potential selection bias but makes nonparticipants in a particular month more (less) similar to participants. In both cases, we find very similar results to the ones presented above.

Another issue is that future participation rates of nonparticipants might be related to the business cycle. We find that future participation rates for both participants and nonparticipants are decreasing over the ten-year period we consider and that they are uncorrelated with the unemployment rate (see the internet appendix for all details). Thus, the correlation of the program effects with the unemployment rate we find is not due to differential future program participation of nonparticipants over the business cycle.

Finally, the fact that we require all persons not to have participated in a program in the 48 months before (potential) program start might affect our results. For the most important specification with stable population and program characteristics, this choice does not matter at all since the common support of the reference population we choose only includes persons who have never participated in a program before.

The estimation and inference of the correlations may be questioned. Using regression based inference based on the Newey-West t-values as implemented in EViews should take care of any autocorrelation and heteroscedasticity that is inherent in the effects (by construction of the moving 6-months defining participation). In addition, the dependent variable in that regression is estimated and thus mismeasured. Since it is consistently estimated, this type of measurement error in the dependent variable should not matter. Nevertheless, as a sensitivity check we estimated a weighted regression in which the weights are proportional to the precision of the effects. Again, the results confirmed our findings.

It remains to check the sensitivity of the results with respect to some operational characteristics of the chosen matching estimator, like the bias correction procedure or the choice of the caliper width. Such checks have been extensively performed and documented by Lechner, Miquel, and Wunsch (2005) who use an identical estimator, but apply it only at one point in time. The reader is referred to their results which indicate a low sensitivity of the estimator with respect to not too large changes of these parameters.

8 Conclusions

We analyze the effects of training programs for the unemployed over a ten-year period based on newly available very informative German administrative data. We generally find negative lock-in effects as well as positive medium to long-run employment and earnings effects of the training programs. We also detect considerable variation of the effects over time. This variation remains even when we artificially (econometrically) keep the characteristics of partici-

pants and the composition of the programs (which show considerable variation) constant over time. We find that this variation is related to unemployment at the start of the program, in the sense that the negative lock-in effects are larger in times of low unemployment and the positive long-run effects are larger in times of high unemployment.

At least for the first part of this finding the explanation appears to be obvious. The negative lock-in effects occur because, while in the program, the unemployed show reduced job search effort and receive less job offers from the caseworker. Therefore, unemployed not 'locked-in' a program find jobs faster. However, if unemployment is high, it takes longer to find a job. Hence, the cost of reduced job search because of attending a program is lower. Since this affects the current participants in the program, we expect the lock-in effect to become worse when the labor market situation improves. For the long-run effects, it is not so obvious why this correlation between effects and labor market conditions exists. One immediate explanation is, however, that the negative lock-in effects continue to influence labor market outcomes, although they are dominated by the positive effects of the additional human capital received in the programs. Thus, even if the human capital effect is more or less unrelated to labor market conditions, the respective correlation of the lock-in effect is sufficient to induce the same correlation in the medium and long-run effects as found for the lock-in effects.

In conclusion, our results suggest that when the economy picks up and unemployment falls, one may want to reduce the volume of training programs by more than the proportional reduction in unemployment would suggest. On the other hand, when labor market conditions worsen, then the share of unemployed in the programs might be increased.

Literature

Abadie, A., and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects", *Econometrica*, 74, 235-267.

- Bender, S., A. Bergemann, B. Fitzenberger, M. Lechner, R. Miquel, S. Speckesser, and C. Wunsch (2005): "Die Wirksamkeit von FuU-Maßnahmen: Ein Evaluationsversuch mit prozessproduzierten Daten aus dem IAB", *Beiträge zur Arbeitsmarkt- und Berufsforschung*, 289.
- Bender, S., Haas, A., and C. Klose (2000): "The IAB employment subsample 1975-1995", *Schmollers Jahrbuch, Zeitschrift für Wirtschafts- und Sozialwissenschaften*, 120, 649-662.
- Bender, S., J. Hilzendegen, G. Rohwer, and H. Rudolph (1996): "Die IAB-Beschäftigtenstichprobe 1975-1990", *Beiträge zur Arbeitsmarkt- und Berufsforschung*, 197.
- Boone, J., and J. van Ours (2004): "Effective Active Labour Market Policies", IZA Discussion paper 1335.
- Couch, K.A. (1992): "New Evidence on the Long-Term Effects of Employment Training Programs", *Journal of Labour Economics*, 10, 380-388.
- Dehejia, R. H., and S. Wahba (2002): "Propensity-Score-Matching Methods for Nonexperimental Causal Studies", *Review of Economics and Statistics*, 84, 151-161.
- Fitzenberger, B., A. Osikominu and R. Völter (2006): "Get Training or Wait? Long-Run Employment Effects of Training Programs for the Unemployed in West Germany", Department of Economics, Goethe-University, Frankfurt am Main.
- Fitzenberger, B., and S. Speckesser (2005): "Employment Effects of the Provision of Specific Professional Skills and Techniques in Germany", Department of Economics, Goethe-University, Frankfurt am Main.
- Fay, R. G. (1996): "Enhancing the Effectiveness of Active Labour Market Policies: Evidence from Programme Evaluations in OECD countries", *Labour Market and Social Policy Occasional Papers*, 18, OECD.
- Fredriksson, P., and P. Johansson (2003): "Program Evaluation and Random Program Starts", IFAU Discussion Paper, 2003(1), Uppsala.
- Fredriksson, P., and P. Johansson (2004): "Dynamic Treatment Assignment - The Consequences for Evaluations Using Observational Data," IZA Discussion Paper 1062, Bonn.
- Frölich, M., and M. Lechner (2006): "Exploiting Regional Treatment Intensity for the Evaluation of Labour Market Policies", CEPR Discussion Paper 5728.
- Gerfin, M., and M. Lechner (2002): "Microeconometric Evaluation of the Active Labour Market Policy in Switzerland", *Economic Journal*, 112, 854-893.

- Heckman, J. J., R. LaLonde, and J. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs", in: O. Ashenfelter and D. Card (eds.), *Handbook of Labour Economics*, Vol. 3, 1865-2097, Amsterdam: North-Holland.
- Hotz, J., G. Imbens, and J. Klerman (2000): "The Long-Term Gains from GAIN: A Re-Analysis of the Impacts of the California GAIN Program", NBER Working Paper, 8007.
- Imbens, G. W. (2004): "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review", *Review of Economics and Statistics*, 86(1), 4-29.
- Jacobson, L., R. LaLonde, and D.G. Sullivan (2004): "Estimating the Returns to Community College Schooling for Displaced Workers", IZA Discussion Paper, 1017.
- Jespersen, S., J. R. Munch, and L. Skipper (2004): "Costs and Benefits of Danish Active Labour Market Programmes", *Danish Economic Council*, Working Paper 2004:1.
- Joffe, M. M., T. R. Have, H. I. Feldman, and S. Kimmel (2004): "Model Selection, Confounder Control, and Marginal Structural Models", *The American Statistician*, 58-4, 272-279.
- Johansson, K. (2001): "Do labour market programs affect labor force participation", *Swedish Economic Policy Review*, 8(2), 215-234.
- Kluve, J. (2006): "The Effectiveness of European Active Labour Market Policy", IZA Discussion Paper 2018.
- Kluve, J., and C. Schmidt (2002), "Can training and employment subsidies combat European unemployment?", *Economic Policy*, 35, 411-448.
- Lechner, M., R. Miquel, and C. Wunsch (2005): "Long-Run Effects of Public Sector Sponsored Training in West Germany", CEPR Discussion Paper 4851.
- Martin, J. P., and D. Grubb (2001): "What Works and for Whom: A Review of OECD Countries' experiences with active labour market policies", *Swedish Economic Policy Review*, 8(2), 9-56.
- Miquel, R., C. Wunsch, and M. Lechner (2002): "Die FuU-Teilnehmer-Datei 1976-1997", *Graues Papier des Instituts für Arbeitsmarkt- und Berufsforschung*, Nuremberg.
- OECD (2004): "Employment Outlook", Paris.
- Rubin, D. B. (1979): "Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies", *Journal of the American Statistical Association*, 74, 318-328.
- Sianesi, B. (2004): "An evaluation of the Swedish system of active labour market programmes in the 1990s", *Review of Economics and Statistics*, 86, 133-155.

van Ours, J. (2004): "The Locking-in Effect of Subsidized Jobs", *Journal of Comparative Economics*, 32, 37-52.

Winter-Ebmer, R. (2001): "Evaluating an Innovative Redundancy-Retraining Project: The Austrian Steel Foundation", IZA Discussion Paper 277.

Wunsch, C. (2005): "Labour Market Policy in Germany: Institutions, Instruments and Reforms since Unification", Discussion Paper 2005-06, Department of Economics, University of St.Gallen.

Appendix A: Data

A.1 Further details on the data

Table A.1 briefly describes the three different administrative data sources used to compile the data that are the basis of this study.

Table A.1: Combined administrative data sources used

	IAB Employment subsample (ES)	Benefit payment register (BPR)	Training participant data (TPD)
Source	Employer supplied mandatory social insurance entries 1980-2003.	Benefit payment register of the PES 1980-2003.	Questionnaires filled in by PES staff for statistical purposes.
Population	1% random sample of persons covered by social insurance for at least one day 1975-1997. Self-employed, civil servants, university students are not included.	Recipients of benefit payments from the PES.	Participants in further training, retraining, short training, German language courses and temporary wage subsidies 1980-1997.
Available information	Personal characteristics and history of employment.	Personal characteristics and information on the receipt of benefit payments from the PES.	Personal characteristics of participants and information about training programs.
Important variables	Gender, age, nationality, education, profession, occupational status, industry, firm size, earnings, regional information.	Family status, number of children, type, and amount of benefits received.	Type, duration, and result of the programme, type of income support paid during participation.

Note: The merged data is based on monthly information. For detailed information on the merging and recoding procedures, see Bender et al. (2005). The construction of this database is a result of a three-year joint project of research groups at the Universities of Mannheim (Bergemann, Fitzenberger, Speckesser) and St. Gallen (Lechner, Miquel, Wunsch) as well as the Institute for Employment Research of the FEA (Bender). A detailed description of the ES is provided by Bender et al. (1996) and Bender, Haas and Klose (2000). For the TPD see Miquel, Wunsch, and Lechner (2002).

The administrative data we use are the most comprehensive database with respect to training in Germany prior to 1998. Unfortunately, collection of training participation data is

discontinued in 1998. It covers an exceptionally long time horizon for observing not only individual employment histories before and after program participation but also for participation itself. Furthermore, the data are very rich in covariates that can be used to control for selectivity.

A.2 Evaluation sample and definition of participation status

Our evaluation sample consists of the prime-age part (age 20-55) of the West German population observed at least once in employment subject to social insurance before a (potential) program start. We exclude persons who were last employed as home workers, apprentices, trainees, or part-time workers below half of the full-time equivalent, because we want to focus on the most common forms of regular employment.

Participants are all unemployed who start a program in a particular month between 1986 and 1995 (120 months). Nonparticipants are all unemployed who do not start a program but receive UB/UA in that month. As long as they fulfil all sample selection criteria individuals can be participants in one and nonparticipants in another month. To ensure that all persons we consider are eligible for participation, we require that they received UB/UA in the month before (potential) program start.

To ensure that we do not use unemployed who completed a program shortly before (potential) program start (are still in an earlier unemployment-participation-unemployment spell), we require that nobody participated in a program in the four years before the (potential) program start we consider. For nonparticipants, we require in addition that they did not participate in the 11 months following potential program start to ensure that they are not too similar to participants with respect to program participation while keeping potential selection bias small.

To obtain a sufficient number of participants we pool participants and nonparticipants over a six-month window in the estimation. Thus, we estimate effects for 115 program starts in the

period 1986-1995. These conditions are all subjected to a sensitivity analysis (see Section 7 for details).

A.3 Measurement of the outcomes

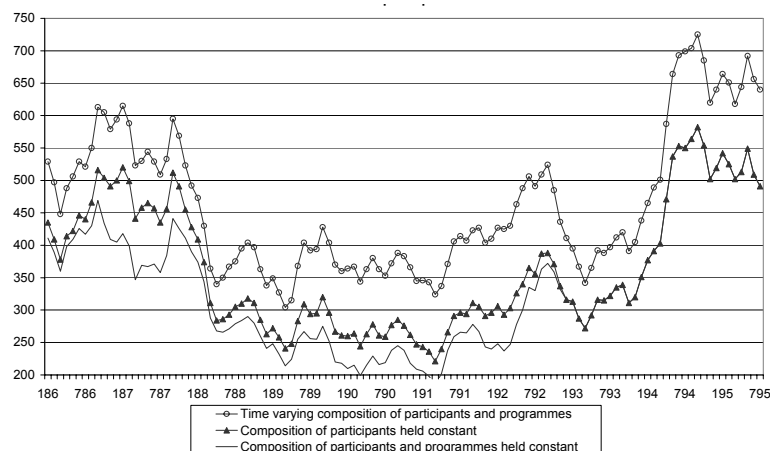
We measure all outcomes relative to the (potential) program start. Whenever a person participates in a program, he is considered as registered unemployed (and not employed).

To capture the short-, medium- and long-run effects of training, we measure all outcome variables at four points in (process) time: 6, 36, 72, and 96 months after program start. However, as choosing one particular month may be a noisy measurement of these effects we calculate the short-run outcome as the mean of months 5-7, the medium-run outcome as the mean of months 33-39, the long-run outcome after six years as the mean of months 61-72 and the long-run outcome after eight years as the mean of months 85-96.

In this study, we focus on the outcome variables registered unemployment (defined as receipt of UB or UA or participation in training) and employment subject to social insurance. We also consider gross earnings as a crude measure for individual productivity.

A.4 Sample sizes of participants

Figure A.1: Number of participants over time



Note: From January 1993 on the number of participants in the specification where the composition of both participants and programmes is held constant is the same as in the specification where only the composition of participants is held constant (because the latter deselected participants in job search assistance that ceases to exist after 1992).

Sample sizes decrease when keeping first the characteristics of participants and then the composition of programs constant over time due to enforcing common support with our reference population (see A.5 for details). They never fall below 200 participants pooled over 6 months. The average is about 400 participants. While these numbers are sufficient for many interesting analyses, it is clear that insights from further disaggregation of the data are clearly limited.

A.5 Characteristics of the reference population

The following table provides descriptive statistics for selected characteristics of the reference population we use to keep the composition of participants and programs constant over time. As reference population, we chose the pool of all training participants in our sample in the period 1986-1995, reduced to the intersection of the common support in all starting months.

Table A.2 Descriptive statistics for the reference population

	All participants		Common support	
	Mean	Std. deviation	Mean	Std. deviation
Age in years	34	9.1	33	8.6
Woman	39	0.49	39	0.49
Married	39	0.49	34	0.48
At least one child	33	0.47	26	0.44
Foreigner	9	0.28	3	0.16
No professional degree	21	0.40	18	0.38
Completed apprenticeship training	74	0.44	77	0.42
University/college degree	6	0.23	6	0.23
Blue-collar worker	39	0.49	38	0.48
High-skilled	20	0.40	15	0.36
Duration of last unemployment spell	11	0.14	7	0.06
Duration of last employment	36	0.41	33	0.30
Fraction employed in last 6 years	54	0.29	60	0.28
Fraction unemployed in last 6 years	26	0.24	19	0.19
Gross earnings of last employment ≤ 1000 €	57	0.50	60	0.49
1000-1500 €	29	0.45	32	0.46
1500-2000 €	6	0.23	6	0.23
> 2000 €	2	0.15	3	0.17
Practice firms	14	0.34	13	0.33
Short training	36	0.48	36	0.48
Long training	25	0.43	25	0.43
Retraining	17	0.37	18	0.39
Job search assistance	9	0.28	7	0.26
Planned program duration in months	8.5	7.3	8.9	7.5
Number of observations	9418		2101	

Note: All variables are measured at or relative to program start. If not stated otherwise the means are percentages. Only the common support is used in the estimation.

Appendix B: Technical details of the matching estimator used

We use a matching procedure that incorporates the improvements suggested by Lechner, Miquel, and Wunsch (2005). These improvements aim at two issues: (i) To allow for higher precision when many 'good' comparison observations are available, they incorporate the idea of calliper or radius matching (e.g. Dehejia and Wahba, 2002) into the standard algorithm. (ii) Furthermore, matching quality is increased by exploiting the fact that appropriate weighted regressions that use the sampling weights from matching have the so-called double robustness property. This property implies that the estimator remains consistent if either the matching step is based on a correctly specified selection model, or the regression model is correctly specified (e.g. Rubin, 1979; Joffe, Ten, Have, Feldman, and Kimmel, 2004). Moreover, this procedure should reduce small sample bias as well as asymptotic bias of matching estimators (see Abadie and Imbens, 2006) and thus increase robustness of the estimator.

The actual matching protocol is shown in Table B.1. Lechner, Miquel, and Wunsch (2005) contains more technical information about the prototypical estimator.

To keep the characteristics of participants and nonparticipants constant over time, we separately match both groups to a reference population to obtain estimates of the potential outcomes of both participation and nonparticipation for a population that resembles the chosen reference population. The average treatment effect on the treated is then obtained by subtracting these estimated potential outcomes.

Table B.1: A matching protocol for the estimation of a counterfactual outcome and the effects

Step 1	Specify a reference distribution defined by X .
Step 2	Pool the observations forming the reference distribution and the participants in the respective period. Code an indicator variable W , which is 1 if the observation belongs to the reference distribution. All indices, 0 or 1, used below relate to the actual or potential values of W .
Step 3	Specify and estimate a binary probit for $p(x) := P(W = 1 X = x)$
Step 4	Restrict sample to common support: Delete all observations with probabilities larger than the smallest maximum and smaller than the largest minimum of all subsamples defined by W .
Step 4	<p><i>Estimate the respective (counterfactual) expectations of the outcome variables.</i></p> <p>Standard propensity score matching step (multiple treatments) a-1) Choose one observation in the subsample defined by $W=1$ and delete it from that pool. b-1) Find an observation in the subsample defined by $W=0$ that is as close as possible to the one chosen in step a-1) in terms of $p(x), \tilde{x}$. 'Closeness' is based on the Mahalanobis distance. Do not remove that observation, so that it can be used again. c-1) Repeat a-1) and b-1) until no observation with $W=0$ is left.</p> <p>Exploit thick support of X to increase efficiency (radius matching step) d-1) Compute the maximum distance (d) obtained for any comparison between member of reference distribution and matched comparison observations. a-2) Repeat a-1). b-2) Repeat b-1). If possible, find other observations in the subsample of $W=0$ that are at least as close as $R \cdot d$ to the one chosen in step a-2) (to gain efficiency). Do not remove these observations, so that they can be used again. Compute weights for all chosen comparisons observations that are proportional to their distance. Normalise the weights such that they add to one. c-2) Repeat a-2) and b-2) until no participant in $W=1$ is left. d-2) For any potential comparison observation, add the weights obtained in a-2) and b-2).</p> <p>Exploit double robustness properties to adjust small mismatches by regression e) Using the weights $w(x_i)$ obtained in d-2), run a weighted linear regression of the outcome variable on the variables used to define the distance (and an intercept). f-1) Predict the potential outcome $y^0(x_i)$ of every observation using the coefficients of this regression: $\hat{y}^0(x_i)$. f-2) Estimate the bias of the matching estimator for $E(Y^0 W = 1)$ as: $\sum_{i=1}^N \frac{\mathbb{1}(W=1)\hat{y}^0(x_i)}{N^1} - \frac{\mathbb{1}(W=0)w_i\hat{y}^0(x_i)}{N^1}$. g) Using the weights obtained by weighted matching in d-2), compute a weighted mean of the outcome variables in $W=0$. Subtract the bias from this estimate to get $E(Y^0 W = 1)$.</p>
Step 5	Repeat Steps 2 to 4 with the nonparticipants playing the role of participants before. This gives the desired estimate of the counterfactual nonparticipation outcome.
Step 6	The difference of the potential outcomes gives is the desired estimate of the effect with respect to the reference distribution specified in Step 1.

Note: We use the fixed-weight heteroscedasticity robust standard errors suggested by Lechner, Miquel, and Wunsch (2005). Since participants and nonparticipants are independent, variance of the effect is the sum of the variances of the potential outcomes. \tilde{x} includes gender, elapsed unemployment duration until program start, and whether a person is employed in month 12 or month 24 before program start. In some specifications, we also match on education. In the specification where program composition is held constant, we also match on the type of program and planned program duration. \tilde{x} is included to ensure a high match quality with respect to these critical variables. R is fixed to 90% in this application (different values are checked in the sensitivity analysis).