

LONG-RUN EFFECTS OF PUBLIC SECTOR SPONSORED TRAINING IN WEST GERMANY

Michael Lechner, Ruth Miquel and Conny Wunsch*

SIAW

Swiss Institute for International Economics
and Applied Economic Research



University of St.Gallen

This version: February, 2006

Date this version has been printed: 06 March 2006

Abstract

We estimate the short-, medium- and long-run effects of different types of government-sponsored training in West Germany using particularly rich data which allows to control for selectivity by matching methods and to measure interesting outcome variable over eight years after programme start. We use distance weighted radius matching with a weighted-regression based bias removal procedure to increase precision and robustness of standard matching estimators. We find negative employment effects in the short-run for all programme types whose magnitude and length of occurrence is directly related to programme duration. In the longer run most types of training seem to increase employment rates by about 10%-points, and for both the shortest and the longest, most intensive programme type these positive effects seem to be sustainable over the eight-year observation period. For the latter programme, however, the positive effects are not (yet?) sufficient to compensate for the large initial lock-in effect.

Keywords: Active labour market policy, matching estimation, programme evaluation, panel data.

JEL classification: J68

Address for correspondence: Michael Lechner, Professor of Econometrics, Swiss Institute for International Economics and Applied Economic Research (SIAW), University of St. Gallen, Bodanstrasse 8, CH-9000 St. Gallen, Switzerland, Michael.Lechner@unisg.ch, www.siaw.unisg.ch/lechner.

* The first author is also affiliated with CEPR, London, ZEW, Mannheim, IZA, Bonn, and PSI, London. Financial support by the Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nuremberg, Germany (project 6-531A), is gratefully acknowledged. Conny Wunsch is also grateful for financial support from the IAB (thesis grant). The data originated from a joint effort with Stefan Bender, Annette Bergemann, Bernd Fitzenberger and Stefan Speckesser to make various German administrative data bases usable for econometric research. The paper has been presented at the SOLE/EALE world congress in San Francisco, at the Econometric Society world congress in London, at the IAB in Nuremberg, the ZEW in Mannheim and at the PPI in Melbourne, as well as in seminars at the Universities of Frankfurt, Queens (Kingston), Sevilla, St. Gallen and Zürich. We thank Stefan Bender, Bernd Fitzenberger, Joseph Hotz and Barbara Sianesi for helpful comments on a previous draft of the paper as well as congress and seminar participants for fruitful discussions. The usual disclaimer applies.

1 Introduction

In the 1990s many continental European countries used active labour market policies (ALMP) as important tools to reduce Europe's notoriously high levels of unemployment, without having to go through the painful side-effects of substantial reforms of the labour markets. Since in many countries a substantial part of the unemployed has below-average or inadequate education, training was considered one of the most important and promising components of this policy (ILO, 1998). This also raised a debate whether substantial human capital investments should be preferred to less intense adjustments of skills in reducing unemployment. From the returns-to-education literature (see e.g. the survey by Card, 1999) it might be conjectured that this, in fact, is a trade-off between short- and long-term gains. However, due to lack of appropriate long-term data an empirical assessment of this question is so far missing in the literature. Most existing evaluation studies, surveyed for example by Fay (1996), Heckman, LaLonde, and Smith (1999), Martin and Grubb (2001), and Kluve and Schmidt (2002), have focused on the short- to medium-term effects of training programmes and are thus not able to fully address this question. The few studies that do estimate long-run effects (Couch, 1992; Hotz, Imbens and Klerman, 2000; Winter-Ebmer, 2001; Jacobson, LaLonde and Sullivan, 2004; Jespersen, Munch and Skipper, 2004) suggest that there are long-term gains from substantial training but in all of these studies a direct comparison with other (less intense) forms of training is missing.

We contribute to filling this gap in the literature by estimating short-, medium- and long-run effects of different types of government-sponsored training for the unemployed (GST) on employment, unemployment and earnings of their participants in West Germany. On the one hand, Germany is one of the few countries that offers a range of training programmes for the unemployed which differ considerably in both the form and the intensity of the human capital investment as well as in their respective duration (ranging from few weeks to more than two years). On the other hand, our administrative data both contain sufficiently detailed

information and large enough sample sizes to capture important aspects of this heterogeneity, and we are able to measure interesting outcome variables on a monthly basis over a period of eight years after a person has entered a programme.

Containing almost 20 years of individual pre-programme employment and earnings histories as well as a rich set of personal, regional and employer information the data allow correcting for selectivity by microeconomic matching methods. We use a matching estimator that accounts for multiple treatments as proposed by Imbens (2000) and Lechner (2001, 2002a, b). Two of the advantages of this estimator are that it is essentially nonparametric and that it allows for unrestricted effect heterogeneity. Accounting for the recent result of Abadie and Imbens (2006) that the usual one-to-one matching estimators may exhibit an asymptotic bias, we implement a weighted-regression based bias removal procedure on top of the matching. Furthermore, we improve efficiency compared to one-to-one matching by predicting the counterfactual outcome by a weighted mean of similar observations in the comparison pool.

For a sample of West German unemployed individuals who enter training in 1993 and 1994 we find that all types of training we evaluate exhibit negative short-run effects on employment whose magnitude and length of occurrence are directly related to programme duration (so-called lock-in effects in the terminology of Van Ours, 2004). However, in the longer run most programme types seem to increase employment rates by about 10%-points. The positive effects for shorter training appear earliest whereas for the longest and most intense programme type it takes almost three years after programme start for the positive effects to appear. But for both programme types these positive effects are sustained in the long run. However, for the most intense programme the cost for obtaining this large positive effect in terms of initial lock-in effects seems to be large. Cumulating the effects over the eight-year period the longest programme type does not succeed in generating a significantly

positive gain in months of employment. The shortest programmes, on the contrary, show a significant gain in employment of nine months over the eight-year window.

Besides adding to the more general discussion of shorter and less intense training versus substantial human capital investments for reducing unemployment, our results also help to resolve a puzzle in the training literature, namely that the estimates for the effects of training programmes for unemployed diverge drastically. This has led to a consensus in the training literature that there is no robust evidence that training programmes for the unemployed have positive effects (as is most vividly illustrated by the quote from Jim Heckman in an article in the *Economist* (April 6, 1996, p.23): "Zero is not a bad number"). In this paper we argue that these diverging findings are most likely strongly influenced by data availability: the different studies are based on very different time horizons over which they are able to measure their respective outcome variables. Given that the effects of training programmes are a mixture of lock-in effects that disappear over time and the long-run effects, it is not at all surprising that the estimated effects are very sensitive to the time window for which the outcome variables are available.

On top of these more general considerations we provide new insights on the effectiveness of GST in West Germany. Perhaps surprisingly given that on average about € 4 billion per year have been spent on GST in West Germany since 1991, so far only little is known about the effectiveness of these programmes, basically because of a lack of appropriate data. Most of the previous studies (see the surveys by Hujer and Caliendo, 2001; Wunsch, 2005) use survey data from the German Socioeconomic Panel (GSOEP).¹ Although with this data it is possible to distinguish GST from other forms of further vocational training, there are not enough observations to appropriately account for effect heterogeneity with respect to participants and

¹ The GSOEP data is a yearly questionnaire-based survey that started 1984. It provides individual data on personal and socioeconomic characteristics as well as retrospective information on the employment history and participation in training programmes. It is the most widely used data source for empirical analyses of the German labour market.

different types of GST.² On the other hand, even the few studies that use richer data do not exploit this information to analyse different forms of GST although there is substantial heterogeneity among them.

Based on a less informative previous version of our data base, Klose and Bender (2000) analyse the effects of GST for a cohort of participants ending programmes in 1986. They construct a control group based on eligibility and a hierarchical matching approach and then estimate hazard rate models. They find ambiguous results concerning the employment effects of GST. Speckesser (2004) and Fitzenberger and Speckesser (2005) draw their sample from the same administrative records as we do. Yet, they restrict their analysis to one special type of GST and observe outcomes for three years after the beginning of the programme only. Based on propensity score matching, they find negative lock-in effects in the first months after programme start. Speckesser (2004) who focuses on training conducted in the period 1993-1994 finds no significant effects thereafter, mainly because of rapidly decreasing sample size towards the end of his observation period. Fitzenberger and Speckesser (2005), who cover a longer period of programme participation (1990-1997), obtain positive employment effects of about 10-20 percentage points from about 10 months after programme start.

None of the studies conducted for West Germany so far has analysed long-term effects of GST beyond three years after the programme. However, since the German programmes are intensive and long by international standards, data that cover considerably more than two or three years after the programme are crucial for understanding their differential impacts.

The remainder of the paper proceeds as follows. The next section provides information on the use of ALMP and in particular training in West Germany and describes the different types of training we evaluate. Section 3 outlines the data and the definition of our evaluation sample and provides descriptive statistics of the data. In Section 4 we discuss our identification and

² Small sample sizes may be one reason why some authors (Hujer, Maurer and Wellner, 1999a, b) using the GSOEP do not

estimation strategy. Our estimation results as well as an extensive assessment of their sensitivity to issues of implementation of the estimator and the definition of nonparticipants are detailed in Section 5. The last section concludes. The appendix contains additional information on the data and on the estimation procedure. Further details are available in an internet appendix to this paper.³

2 Labour market policies in West Germany

2.1 The use of ALMP in West Germany

Besides counselling and job placement services, the most important instruments of ALMP in West Germany are training programmes, rehabilitation measures and subsidised employment. Table 1 displays the expenditure for different measures of passive and active labour market policies in West Germany for the years 1991-2003.

Table 1: Expenditure on active and passive labour market policies 1991-2003

	1991	1993	1995	1997	1999	2001	2003
Total expenditure in million EUR	25,008	35,425	39,275	42,781	41,858	40,745	47,613
	Shares of total expenditure for active and passive labour market policy in %						
Training	13	10	10	8	10	11	7
Temporary wage subsidy	1	<1	<1	<1	2	2	1
Short-time work	1	5	1	1	1	1	1
Job creation schemes	6	4	4	3	3	2	1
Early retirement	1	<1	<1	<1	<1	1	2
Rehabilitation programmes	7	6	4	4	4	5	5
Unemployment benefits	33	43	46	47	39	40	44
Unemployment assistance	14	15	19	23	24	18	19
Other expenditure	25	18	15	13	17	19	19
Unemployment rate in %	6.2	8.0	9.1	10.8	9.6	8.0	9.3

Sources: BA (1992-2004).

Notes: Training: further training, retraining, short training according to §41a EPA (abolished at the end of 1992). Temporary wage subsidies are subsidies during the phase of initial skill adaptation in a new job (*Einarbeitungszuschüsse*). Short-time work: *Kurzarbeit*. Job creation schemes (JCS): *Arbeitsbeschaffungsmassnahmen*. Early retirement: *Vorruhestand/Altersteilzeit/Altersübergangsgeld*. 'Other expenditure' mainly includes counselling and job placement services as well as administrative costs of the public employment service (PES).

In line with the development of the unemployment rate, an increasing and now substantial fraction of expenditure is devoted towards payment of income support during unemployment

distinguish between GST and other forms of further vocational training.

(unemployment benefits, unemployment assistance). The structure of expenditures for ALMP is relatively stable. Training, which has the objective of updating and increasing the human capital of those workers who drop out of the production process and become unemployed, is by far the most utilised instrument followed by rehabilitation programmes, which aim at improving the labour market integration of the disabled and persons with severe health problems, and subsidised employment (temporary wage subsidies and job creation schemes in Table 1).

2.2 *Training as a part of German ALMP*

In Germany, labour market training consists of very heterogeneous instruments that differ largely in the form and intensity of the human capital investment as well as in their respective duration. 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. Usually, participants in such training programmes receive a transfer payment that is of the same amount as unemployment benefits.⁴ Moreover, the public employment service (PES) bears the direct cost of the programme, and it may cover parts of additional expenses for childcare, transportation and accommodation.

For our analysis, we aggregate the different programmes into groups according to their homogeneity with respect to selection of participants, educational contents and organisation, as well as sample size and information available to distinguish different types of programmes. Table 2 shows the resulting groups of training programmes plus a residual category. Ignoring the programme types for which the number of observations is too small, we restrict our analysis to so-called practice firms, general further vocational training and retraining programmes.

³ www.siaaw.unisg.ch/lechner/lmw_fuu.

⁴ So-called maintenance allowance (MA). Before 1994, the replacement rate was somewhat higher than for unemployment benefits.

Table 2: Definition of programmes

Programme	Description
Short training	Further training in the profession held with planned duration ≤ 6 months.
Long training	Further training in the profession held with planned duration > 6 months.
Retraining	Training to obtain a new professional degree in a field other than the profession currently held.
Practice firm	Further training that simulates a job in a specific field of profession
Other programmes	Career improvement: Further training to obtain a higher professional degree, e.g. master craftsman, technician, or a (below university) degree in business administration. Residual category: Various very small and heterogeneous programmes.

Note: Due to insufficient sample size the category *other programmes* is not evaluated in this study.

So-called practice firms simulate - though under very realistic conditions - working in a specific field of profession. There are two forms of practice firms which either simulate the commercial part of a company (administration, accounting, customer relations, etc.) or the manufacturing part.⁵ The mean duration of training in practice firms was six months in our sample (cf. Figure 1).

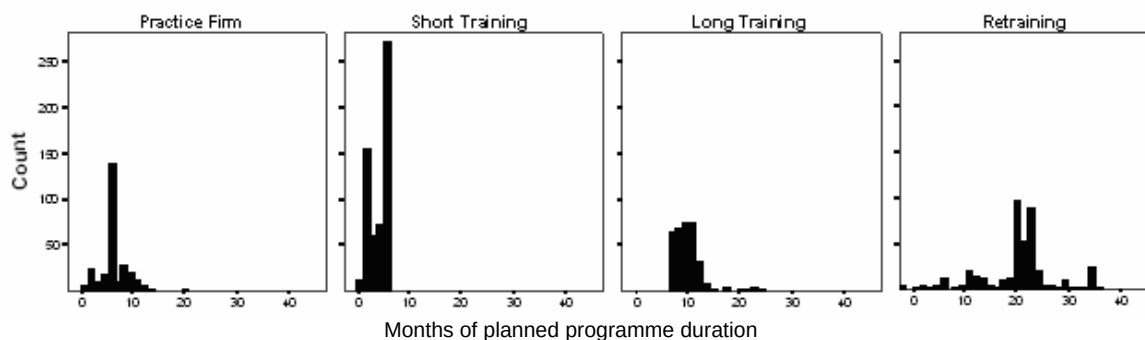
Further training comprises courses that provide a general adjustment of working skills or an additional qualification in the profession currently held, as well as courses that award a first professional degree. Planned durations range from one month to about two years (cf. Figure 1). Since further training is a fairly large and heterogeneous group, we split it into two subgroups based on the planned duration of an individual course. As a characterisation of the programme and not of its participants, *planned* instead of *actual* duration has the advantage that it is independent of the behaviour of the individual during participation (e.g. a short actual duration could be associated with a short course or a long course that a participant left early).

Retraining enables working in a different profession than the one currently held by awarding a new professional degree. Planned durations are long (up to three years, 20 months on average). The acquired skills are equivalent to an apprenticeship in the German apprenticeship

⁵ For the commercial part, there exists a Germany- and Europe-wide network of practice firms that trade 'virtual' goods and services with each other to provide realistic conditions for participants who are the practice firm's employees. The skills acquired correspond to what is required for the specific job held within the practice firm, e.g. that of an accountant. Courses in practice firms representing the manufacturing part, on the other hand, are very heterogeneous ranging from specialist training in technical professions over obtaining a driver's licence for special vehicles to just practising bricklaying.

system. Thus, the human capital investment is quite substantial. Almost two thirds of the participants in our sample were trained towards a profession in the service sector and most of the rest has target professions in the manufacturing sector.

Figure 1: Distribution of the planned programme duration



Note: This is the planned duration of a programme determined before the programme starts.

Note that generally, the German programmes are long on average compared to other OECD countries (e.g. OECD, 1996a, b). Moreover, most training courses are full-time courses, and in addition to classroom training, a course may include on-the-job training which is frequently the case in courses that award a professional degree.

3 Data and definition of the evaluation sample

3.1 The new database

We use administrative data from three different sources which have been made available to the scientific community only relatively recently: the IAB Employment Subsample (ES), the benefit payment register (BPR), and the training participant data (TPD).⁶ Table 3 provides a description of the main features of these data sets. The three data sets were merged to obtain an integrated data base that covers not only participant information but also the full history of insured employment and benefit receipt for both participants and nonparticipants in

⁶ The common German abbreviations for these data sources are IABS, LED and FuU. 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).

government-sponsored training. The merged data base contains information for 208,928 individuals (54,756 of whom registered as training participants in the TPD) from 1975 to 1997. Here, we also use supplementary data from the employment and benefit records up to the year 2002 for the individuals in our sample.⁷

This new database is the most comprehensive one in Germany with respect to training conducted prior to 1998. With this database, we are able to reconstruct more than 25 years of individual employment histories. It contains detailed personal, regional, employer and earnings information (see Appendix A for a complete list of variables). Thus, it allows controlling for many, if not most, of the factors that determine selection into programmes (see the detailed discussion in Section 4.2) as well as a precise measurement of interesting outcome variables (e.g. employment status, earnings). Moreover, we are able to distinguish different programme types and we have a sufficient number of observations for the major programme groups to account for programme heterogeneity.

Of course, there are several drawbacks as well, three of those could be important: First, the data do not cover nonworking recipients of social assistance because they do not receive benefits from the PES but from the local authorities. Second, employment that is not subject to social security contributions is unobserved. On the one hand, this includes self-employment and working as a civil servant ("Beamter"). On the other hand, this regards minor employment below the relevant earnings threshold. Third, it is impossible to distinguish between subsidised employment (like in job creation schemes) and regular employment in the first labour market. This, however, is not a severe problem for West Germany because the number of subsidised employees has always been rather small.

⁷ Following the abolishment of the EPA and introduction of Social Code III on January 1st, 1998, data collection and processing has been changed as well. The new data are similar to the data formerly included in the IABS and in the LED. See Internet Appendix IA.2 for a comparison of the different definitions of the outcome variables before and after this break in data collection.

Table 3: Combined data sources used

	ES	BPR	TPD
Source	Employer supplied mandatory social insurance entries.	Benefit payment register of the PES 1975-1997.	Questionnaires filled in by the labour officer for statistical purposes (ST35).
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. Data until 2002.	Recipients of benefit payments from the PES 1975-2002.	Participants in further training, retraining, short training, German language courses and temporary wage subsidies 1975-1997.
Available information	Personal characteristics and history of employment.	Information about the receipt of benefit payments from the public employment service.	Personal characteristics of participants and information about training programmes.
Important variables	Gender, age, nationality, education, profession, occupational status, industry, firm size, earnings, regional information.	Type and amount of benefits received.	Type, duration and result of the programme, type of income support paid during participation.
Structure	Spells based on daily information.	Spells based on daily information.	Spells based on monthly information.

Note: The merged data is based on monthly information. For detailed information on the merging and recoding procedures see Bender et al. (2005). The creation of this data base 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).

Despite these drawbacks, compared to what has been used in evaluation studies for Germany so far, this database is a substantial improvement in several dimensions, like sample size, selection and outcome information, as well as observable programme heterogeneity. Moreover, relative to what is available for other countries, the database is unique with respect to the length of observable pre- and post-treatment employment histories (up to 19 years before and 8 years after entry in a programme).

3.2 Definition of programme participation, the population and our sample

In this section, we define 'participation in a programme' and our population of interest. We consider programme participation between 1993 and 1994. By choosing this period, we are able to focus on the most recent programmes that still allow for a long enough observation period for detecting long-run effects.⁸ Second, a person is included in our population of interest if he starts an unemployment spell between 1993 and 1994 (so-called 'defining'

⁸ Furthermore, since we observe only training spells after the participant left training, and some courses have a duration of more than two years, and there is no training information after 1997, concentrating on the years 1993 and 1994 does not lead to a selective underrepresentation of long training spells.

unemployment spell). The group of participants in training consists of all persons entering a programme between the beginning of the 'defining' unemployment spell and the end of 1994. If there are multiple participations in this period, then only the first one is included in the analysis. Nonparticipants are all persons who did not enter a programme in this period.

When choosing the appropriate subpopulation from our inflow sample into unemployment, we aim at having a homogenous group of people that covers the prime age part of the West German population who is eligible for participation in training. Therefore, we require that all individuals were employed⁹ at least once before programme participation and that they received unemployment benefits (UB) or assistance (UA) in the month before programme start (as well as in the month of programme start for nonparticipants).¹⁰ This, however, requires the use of variables measured relatively to the start of the programme.¹¹ In this paper, we follow one of the approaches suggested by Lechner (1999). We simulate start dates for nonparticipants by drawing start dates from the empirical distribution for participants. We exclude nonparticipants for whom this date lies before the beginning of the 'defining' unemployment spell, or after 1994, or after the person's last spell that is observed in the data if it ended before 1995.

Given the potential problems with this procedure that have been indicated recently by Fredriksson and Johansson (2003, 2004), we checked the sensitivity of our results with respect to the choice of the treatment window and the procedure for generating programme start dates for nonparticipants fairly extensively. The details of this sensitivity analysis are presented in section 5.4. In summary, our results appear to be robust.

⁹ 'Employed' means that we observe the person at least once in insured employment in the IABS (for the majority of our individuals, the first observation in the data is an employment spell).

¹⁰ In fact, receipt of UB or UA directly before entering a programme is not entirely sufficient to ensure eligibility. Individuals must also have a formal professional degree plus three years of work experience (since 1994, zero years), or alternatively at least six years (since 1994, three years) of work experience. Thus by also requiring individuals to be employed at least once before the programme, the remaining group of participants and nonparticipants is most likely to be eligible.

To avoid most influences coming from retirement, early retirement and primary education, we also impose an age restriction (20-55 years) in the year of the (hypothetical) programme start. Concentrating on the main body of the active labour force, we exclude unemployed who were trainees, home workers, apprentices or without previous employment or whose last employment before the 'defining' UE spell was less intensive than half of the usual full-time work hours. This procedure leaves us with a sample of about 9,000 nonparticipants and about 270 to 570 participants in the four programme groups we consider in the econometric analysis (see Table A.1 in Appendix A).

3.3 Descriptive statistics

Table 4 shows descriptive statistics for selected socio-economic variables for the different subsamples defined by treatment status. Concentrating on the first five groups included in the econometric analysis, the results in this table can be summarised as follows: Participants in *retraining* are on average 31 years old, and thus much younger (about five years) than other unemployed which is in line with the idea that substantive human capital investments are most beneficial if the productive period of the new human capital is fairly long. Another interesting feature is that the share of foreigners in the programme is only about half the share of foreigners in the group of nonparticipants. Participants in *practice firms* and *retraining* are less educated and skilled than the rest. The mirror image of this observation is that participants in *short* and *long training* appear to have the best a priori chances on the labour market, although the education level of those in *short training* is somewhat lower than for those in *long training*. Correspondingly, earnings are somewhat higher for participants in *short* and more strongly in *long training* than in *practice firms* and *retraining*. Earnings of the latter two groups are almost the same as for *nonparticipants*. For the two variables indicating remaining UB claims, Table 4 does not show much variation, though the average remaining

¹¹ Moreover, all variables potentially influencing both selection into programmes and outcomes are measured relatively to the start of the programme.

claim for participants in *retraining* is about two months shorter than for the other unemployed. Finally, note that regional and gender differences are fairly small.

Table 4: Descriptive statistics of selected variables by treatment status (shares in %)

	Nonparticipation	Practice firm	Short training	Long training	Retraining	Career improvement	Other
Number of observations	9197	273	572	329	413	110	74
Personal characteristics							
Women	41	33	37	39	38	24	27
Age ⁺⁺ *	37	36	35	35	31	32	32
Nationality: German	81	87	91	92	89	92	92
Education: no university entrance degree, no professional degree	25	18	15	9	24	5	5
Polytechnical or university degree	6	0	6	14	3	10	2
Position in last job							
Salaried employee	28	34	38	57	23	41	39
Unskilled worker	40	36	31	19	53	9	2
Last monthly earnings							
Salary in EUR*	1,680	1,640	1,773	1,889	1,640	2,072	1,781
Remaining UB claim (before entry in the programme)							
Remaining UB claim (months)*	8	7	8	7	5	7	7
Legal UB claim at the beginning of the last unemployment spell before the programme							
Legal claim* (months)	13	11	12	11	9	10	10
Programme information							
Planned programme duration* (months)		6	4	10	20	10	5
Regional information							
North-Rhine-Westphalia	31	21	28	36	35	23	28
Rhineland-Palatinate, Hesse, Saarland	17	23	19	18	17	17	19
Baden-Württemberg, Bavaria	30	26	34	21	22	42	38
UE rate ⁺⁺ ≤ 5%	1	2	2	0	1	2	4
5% < UE rate ≤ 10%	64	67	70	62	61	72	76
UE rate > 10%	35	31	28	38	38	26	20

Note: *Numbers marked by an asterisk are means (rather than proportions). **Measured in the year of the beginning of the programme. The sample used for the table is the one after all selection steps described in Section 3, but before imposing the common support requirement. For a detailed list of variables see the appendix. 'N =' means the number of individuals for whom this information is available at that point in time.

Figure B.1 in Appendix B provides a look at the monthly employment rate centred at the actual or simulated beginning of the programmes as well as the corresponding numbers for the other outcome variable of interest, namely registered unemployment.¹² Prior to training, all curves are on a fairly similar level, although the employment level of the future *retrainees* is the lowest one for most of the time (and among the highest after the programme). Since being unemployed is a precondition for participation, the shapes for all subgroups exhibit the usual Ashenfelter's dip (Ashenfelter, 1978) starting about two years prior to the actual or simulated

¹² Note that we define registered unemployment as receipt of UB or UA in the respective month.

programme starts. It must be pointed out that this dip is a purely statistical phenomenon coming from the way we selected the sample by enforcing the eligibility criterion for participants and nonparticipants.

After training, the different rates recover quickly, the speed mainly depending on the average programme duration. The surprising finding is that for *nonparticipants* the recovery of the employment rate suddenly slows down about one year after the simulated programme start at a level of about 35 to 40%. This timing coincides with the end of the benefit period for this group, thus suggesting that a large group of *nonparticipants* leaves the labour force after benefit exhaustion. This is reinforced by considering unemployment for which no large differences are observed after the initial programme period is over (note that programme participation increases the benefit period), therefore more of the nonworking people remain registered as unemployed thus probably making up the difference in employment rates. Clearly, combining the fact from Table 4 and Figure B.1 that retrainees are no good a priori risks on the labour market together with the observation that their long-run employment rates are among the highest, already suggests the likelihood of a positive effect for this programme.

The left panel of Figure B.2 in Appendix B shows the distribution of starting months in the two-year window we consider. Partly due to the construction of our sample, the treatment probability increases over time. The right panel of Figure B.2 shows the number of months between the beginning of the 'defining' unemployment spell and participation. With the exception of *career improvement*, which is not considered in the evaluation, the start date distribution is rather homogenous across treatments although *retraining* appears to be used very early in the spell (about 45% of retrainees start within in the first three months). Note however, that due to the combination of the 'defining' unemployment spell beginning 1993 or 1994 and training be observed not later than 12/1994, clustering in the first two months is rather natural.

Given our definition of a small treatment window (although often much smaller windows are used in the literature, e.g. Dorsett, 2005), it is particularly important for the interpretation of our results which share of the control groups receives treatment as well (similar to the problem of substitution bias in an experiment). Furthermore, there is the issue of subsequent programme participation over time. The conceptual problem with analysing the effect of e.g. the second participation is that it might be subject to sample selection influenced by the effect of the first programme. Thus, an analysis of the effects of sequences of programmes requires a dynamic evaluation approach as suggested by Miquel and Lechner (2001), or Lechner (2004), which is not feasible with our data without further aggregation of programme types, which is undesirable for obvious reasons.

Table 5 shows the share of observations defined by the treatment status who participate in additional programmes. Only about 11% of nonparticipants receive some sort of training until the end of 1997. For the longest programme *retraining*, subsequent participations occur only in rare cases (4%). For the other programmes between 10% and 19% of trainees participate again with a large fraction participating in the same programme type. To conclude, Table 5 provides convincing evidence that the effects we will estimate are very close to the 'pure' effect of the programme used to define the treatment status.

Table 5: Participation in different programme types in % of participants in subsamples until 1997

Programme participation between 1993 and 1997	Treatment status in study (first treatment)					
	Non- participation	Practice firm	Short training	Long training	Retraining	Other
Practice firm	1.6	3.7	0.9	1.2	0.5	2.3
Short training	4.8	4.0	7.3	2.7	1.2	5.0
Long training	3.2	4.4	4.7	4.9	1.5	5.4
Retraining	1.2	1.5	4.4	1.5	1.0	1.4
Other	0.7	1.1	1.9	0.0	0.2	2.7
Total other treatments than first treatment	11.4	14.7	19.2	10.4	4.4	17.7

Note: Entries show the fraction (%) of members of the subsamples stated in the columns who participated at least once in the treatments stated in rows after their first treatment (programme participants) or after 1994 (nonparticipants). Due to data limitations only training spells completed by the end of 1997 are observable.

4 Identification and estimation

4.1 Conditional independence and matching

We base our analysis on the 'multiple treatments' model of the microeconomic evaluation literature: An individual chooses between several states, like participation in a specific training programme or nonparticipation in such a programme. The different *states* are called *treatments* in the econometric evaluation literature. The potential participant in a programme is assigned a hypothetical outcome $\{Y^0, \dots, Y^M\}$, e.g. employment, in each state $S \in \{0, \dots, M\}$. This notation allows us, under the usual assumptions (see Rubin, 1974), to define our parameters of interest - the average treatment effects on the treated (ATETs) for pair-wise comparisons of the effects of different states (Lechner, 2001, 2002a, b):

$$\theta_0^{m,l} := E(Y^m - Y^l | S = m) = E(Y^m | S = m) - E(Y^l | S = m); \quad m \neq l; m, l \in \{0, \dots, M\}$$

ATET ($\theta_0^{m,l}$) is the expected effect for an individual that is drawn randomly from the population of participants in treatment m only. Without further assumptions, ATETs are not identified from the data. But if we can observe all factors that jointly influence outcomes and the participation decision, then - conditional on these factors (call them X) - the participation decision and the outcomes are independent. This property is exploited by the conditional independence assumption (CIA).

4.2 Is the conditional independence assumption plausible with our data?

Plausibility of the CIA requires that all major factors that jointly determine participation and outcomes are observed in the data. Thus, we should identify these variables and show that they are available in our data. This requires an understanding of the underlying participation decisions. Selection into programmes is determined by three main factors that are discussed in turn: eligibility, selection by caseworkers and self-selection by potential participants.

In Germany, eligibility requires that the potential participant qualifies for or receives unemployment insurance payments. Moreover, he must have a professional degree from the German apprenticeship or higher education system, or a minimum amount of work experience in one field of profession. We control for these factors by use of the education variables, the individual employment records and the entries from the benefit payment register.

Usually, it is the caseworker but it may also be the unemployed himself who proposes participation in training to improve employment prospects. In any case, the unemployed must apply before the beginning of the programme. The caseworker decides whether he will be admitted or not. There is no legal entitlement to participation, and caseworkers have a considerable amount of discretion. Usually, the caseworker decides in consultation with the potential participant whether or not and if so, what kind of training programme would be appropriate based on an assessment of the employment prospects and the specific qualification needs of the unemployed. According to German legislation, caseworkers have to take into account the chances of the unemployed for completing a specific programme successfully, and the situation in the local labour market. Variables capturing information about employment prospects and chances for successful completion of a programme include age, educational attainment, family status and past employment histories including information about past employers, earnings, position in job, specific occupation and industry. Moreover, our data contain detailed regional information that allow us to control for local labour market conditions by use of regional dummies and local unemployment rates. In addition, since unemployment is rising during 1993 and 1994 and participation shows a varying pattern, we condition on the month of the start of the programmes thus netting out seasonal and timing effects.

From the point of view of the unemployed, his decision whether or not to participate in a programme is guided by considerations very similar to those of the caseworker. There are, however, additional reasons for joining or not joining a programme. If, for example, the unem-

ployed sees no chance to find a job anyway, with or without a programme, he may prefer not to join a programme that reduces his leisure time. This again requires controlling for all factors that determine individual employment prospects and labour market conditions. Moreover, legislation provides rather strong incentives to participate in government-supported training. On the one hand, unemployed who refuse to participate in a training course, risk suspension of their benefits. On the other hand, periods during which unemployed receive transfer payments while participating in a training programme count towards acquisition of unemployment benefit claims. Therefore, we constructed variables from the (un)employment histories that indicate the UB claim at the beginning and at the end of a spell.

Although this is much more information than usually available in studies that rely on the CIA (e.g. Heckman and Smith, 1999; Brodaty, Crépon and Fougère, 2001; Larsson, 2003; Dorsett, 2005), there are some potentially important factors missing. In contrast to Gerfin and Lechner (2002) or Sianesi (2004), there is no information about the caseworker's direct assessment of the characteristics and prospects of the unemployed, for example with respect to motivation and ability. Moreover, we do not observe things like jail and health histories. For these variables, we have to rely on their indirect effects, i.e. on their effects on the employment and earnings history that materialised in the past. Since with our data we are able to observe up to almost 20 years of pre-treatment employment and earnings histories of the persons in our evaluation sample, we are confident that we capture these indirect effects sufficiently well.

4.3 *A note on estimation*

Having established identification of the effects, i.e. the general feasibility of consistent estimation of the effects, the question of the appropriate estimator arises. All possible parametric, semi- and nonparametric estimators are implicitly or explicitly build on the principle that for every comparison of two programmes, for every participant of one of those programmes we need a comparison observation from the other programme with the same characteristics re-

garding all factors that jointly influence selection and outcomes (see Section 4.2 for these variables). Here, we use propensity score matching estimators for multiple treatments as our baseline estimator to produce such comparisons. A clear advantage of these estimators is that they are essentially nonparametric and that they allow arbitrary individual effect heterogeneity (see Heckman, LaLonde and Smith, 1999, for matching with a binary treatment, and Imbens, 2000, and Lechner, 2001, for multiple treatments; Imbens, 2004, provides an excellent survey of the recent advances in this field). Gerfin and Lechner (2002) and Larsson (2003) used this estimator, among many others.

Internet Appendix IC contains the detailed results of a multivariate analysis that models selection into the different treatment groups based on a multinomial probit model that is estimated by simulated maximum likelihood. Beyond providing useful descriptive statistics, the output from this selection model plays a key role in our selection correction mechanism since we use the estimated conditional on X choice probabilities (the so-called propensity scores) to form our comparison groups. The analysis revealed that age, nationality, education, and last occupation are important individual characteristics that determine participation. Furthermore observed employment and unemployment histories are significantly correlated with participation choice. Regional information which entered the probit to capture the specifics of supply and demand in the local labour market, also play important roles in the selection process.

In addition to the propensity scores, we include as matching variables gender and the date of the beginning of the programme as well as three dummies which indicate whether the participant is observed employed 12, 24 and 48 months before the programme to make sure that we match correctly on these important variables. However, excluding all additional matching variables except gender, or adding elapsed unemployment duration before the

programme does not change the qualitative results that we obtain for the effects of training (see Section 5.4 and the internet appendix for details).

We improve the standard matching estimator for multiple treatments in two dimensions: To allow for higher precision when many 'good' comparison observations are available we incorporate the idea of calliper or radius matching (e.g. Dehejia and Wahba, 2002) into the standard algorithm. Furthermore, we increase the matching quality 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 as well as asymptotic bias of matching estimators (see Abadie and Imbens, 2006) and thus increase robustness of the estimator in this dimension as well.

The actual matching protocol is presented in detail in Appendix C. All issues concerning the estimation and tests of its operational characteristics are contained in the internet appendix (see in particular Appendix IB for the implications of imposing common support, and match quality). The longer working paper version (Lechner, Miquel and Wunsch, 2004) contains more technical information about the estimator that is omitted here for the sake of brevity. The most important sensitivity checks are discussed in Section 5.4. In summary, the estimator performed reasonably well.

5 The effects of training

5.1 *Measurement of the outcomes in the labour market*

According to German legislation one of the objectives of active labour market policy is to increase reemployment chances and to reduce probabilities of remaining unemployed.

Therefore, important outcome variables are those relating to the employment status, like registered unemployment and different types of employment.¹³ Some types of employment require a certain quality, approximated for example by the job's duration and earnings compared to the previous job. Furthermore, gross earnings are considered as a crude measure for individual productivity.

All effects are measured beginning in the month after the programme started (with simulated start dates for nonparticipants). Focusing on the beginning instead of the end rules out that programmes appear to be successful, just because they keep their participants busy by making them stay in the programme. We consider a programme to be most successful if everybody would leave for 'good' employment immediately after starting participation. Whenever a person participates in any of the programmes he is considered as registered unemployed (and not employed).

5.2 Mean effects of programmes for their participants

Table 6 panel (a) shows the means of the outcomes in the various groups, $E(Y^m | S = m)$, on the main diagonal and the estimated counterfactual expectations, $E(Y^l | S = m)$, off the diagonal. Panel (b) displays all pair-wise comparisons between the programmes and nonparticipation, $E(Y^m - Y^l | S = m)$. Panel (c) states the sample size in each group after imposing common support. Here, we concentrate on the long-term employment effects by looking at the outcome *employment eight years after participation started*.

The columns in panel (a) show the (estimated) outcomes in counterfactual state l for the different subpopulations characterised by m . Thus, they reveal the magnitude of the selection bias corrected for by the estimation procedure. It is up to a magnitude of about 10%-points for some comparisons. From the direction of the selection correction, we can infer whether one

¹³ Here 'registered unemployment' denotes all individuals receiving UB or UA.

group has a priori better or worse chances on the labour market than the other. It turns out that participants in *short training* and to a lesser extent in *long training* have better a priori chances than *nonparticipants* and participants in *practice firms*. The same holds true for participants in *retraining* compared to *nonparticipants*. Note that due to sampling error in the estimates, the results do not allow for a complete ranking of all populations.

Table 6: Estimated employment effects eight years after the beginning of the programme

<i>l</i>	Nonparticipation	Practice firm	Short training	Long training	Retraining
<i>m</i>	(a) Estimate of $E(Y^l S = m)$				
Nonparticipation	38.5	50.2	50.1	46.2	54.6
Practice firm	44.8	45.3	52.8	52.8	60.0
Short training	45.0	49.0	52.5	46.4	59.3
Long training	49.1	52.5	52.4	53.6	56.8
Retraining	50.6	47.3	52.3	50.2	59.1
<i>m</i>	(b) Estimate of $\theta_0^{m,l} = E(Y^m - Y^l S = m)$				
Nonparticipation	0	-11.7	-11.6*	-7.7	-16.1*
Practice firm	0.5	0	-7.5	-7.5	-14.7*
Short training	7.5	3.5	0	6.1	-6.8
Long training	4.5	1.1	1.2	0	-3.2
Retraining	8.5	11.8	6.8	8.9	0
(c) Sample size after imposing common support					
	6964	247	503	267	386

Note: Bold numbers indicate significance at the 5% level, numbers in *italics* relate to the 10% level and * to the 1% level.

The estimates of the mean effects of treatment *m* compared to treatment *l* for the subpopulation observed in state *m* in panel (b) can be computed directly from panel (a) by subtracting the value in a cell from the value in the shaded cell of the same row. They show that in the long run all programmes have positive effects compared to *nonparticipation*, although not always significant. *Retraining* has the largest long-run effect of about 9%-points compared to *nonparticipation* suggesting a substantial and sustainable impact of this expensive programme. In the long run, *retraining* also dominates *practice firms*, but although it seems to have larger effects compared to *short* and *long training*, the latter are not significant. Among the programmes, the effects for *practice firms* appear to be hardest to pin down. Finally, note that the comparison between *short* and *long training* reveals hardly any positive returns of the additional investment in time and money required for *long training*. In

fact, the results are quite positive for a 'short' programme with a maximum duration of six months and a much lower average duration, although due to lacking cost data, other than very rudimentary cost-benefit considerations are impossible.

Table 6 is also informative about the question whether, on average, caseworkers send those types of unemployed into a specific programme that can expect the highest return from it. If this presumption is true, the effect must be larger for the participants in that programme than for any of the comparison groups, i.e. $\theta_0^{m,l} > -\theta_0^{l,m} \forall l \neq m$. Even without exact standard errors for this test, it is obvious that, if there is evidence at all, it suggests that caseworkers do not send those unemployed into the programme for which the highest return is expected.

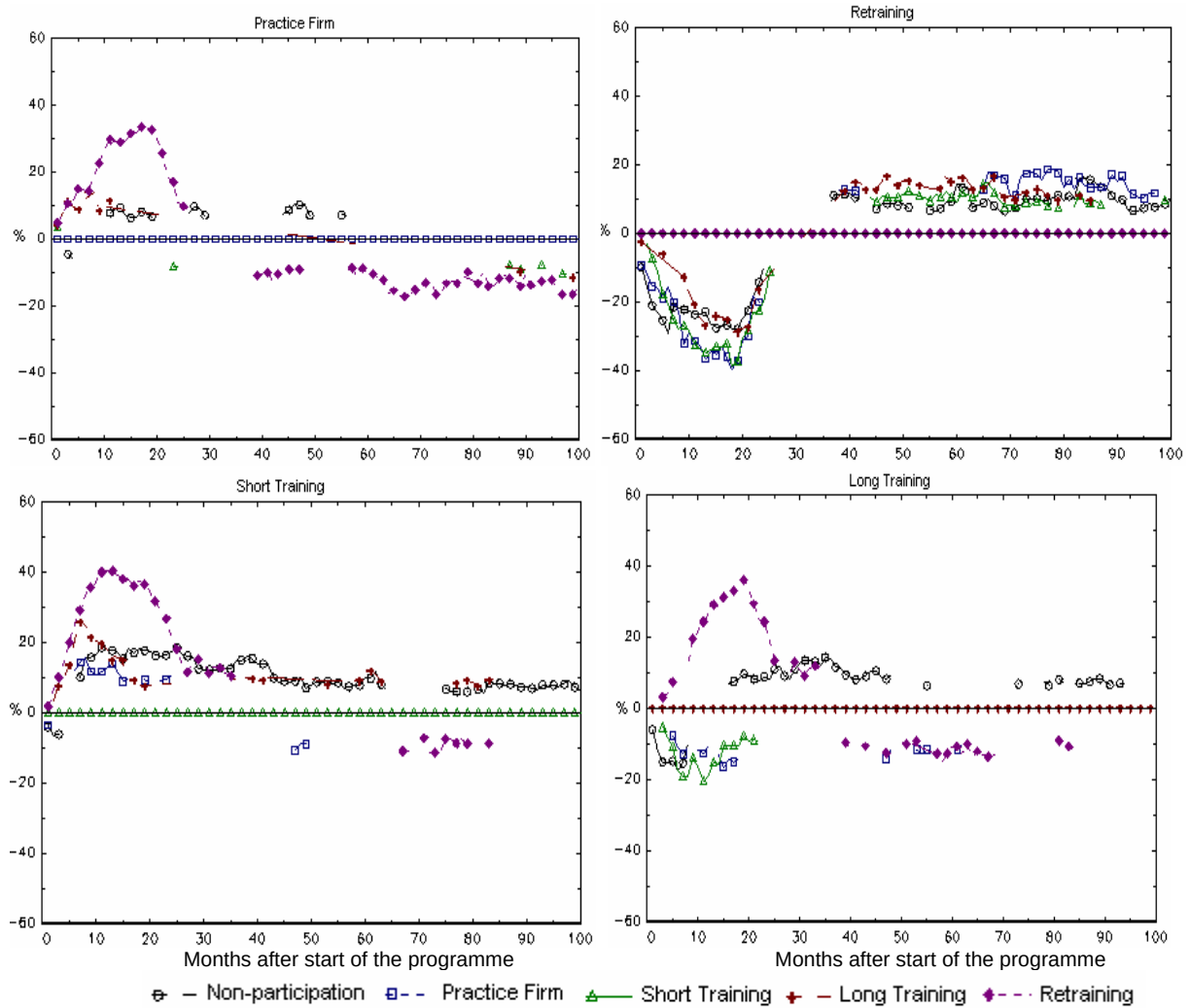
Figure 2 displays monthly estimates of the effects of the different programmes (compared to the other states) for participants in the respective programme (ATET) for *employment*. A line above zero indicates that the programme has a positive effect relative to the state associated with that particular line. In other words, a line above zero is good news for the programme appearing in the header of the respective graph and bad news for the one associated with the particular line. Only effects significant at the 5% level are displayed.

Figure 2 suggests that all programmes have some negative lock-in effect (in the terminology of Van Ours, 2004) due to reduced job search or received job offers during participation in the programme. The length of this negative effect is very much tied to programme duration, being about two to three years for *retraining*, six months to one year for *long training*, three to six months for *short training* and *practice firms*.

After the lock-in effect, the reemployment chances for participants in *practice firms* are somewhat better than in case of *nonparticipation*, but the effects are only mildly significant and appear not to be sustainable. Participation in *retraining* would have been superior in terms of employment chances. After about three years there are substantial positive employment effects of *retraining* for its participants compared to *nonparticipation* and the other types of

training. A positive effect of similar size appears for *short training* compared to *nonparticipation* although it starts much earlier (nine months) and seems to decline slowly over time. The positive effect for *long training* compared to *nonparticipation* begins even later (18 months) and is not very incisive after about four years.

Figure 2: Dynamics of the effects ($\hat{\theta}_t^{m,l}$): Employment differences in %-points



Note: Only effects that are significant at the 5% level (point wise) appear in the figures. Sample size declines quickly after 8 years (96 months).

To check whether jobs are (somewhat) stable, we use an outcome variable which requires at least seven months of continuous employment (six months is the usual probation period in Germany, within that period termination of a job is very easy for both sides). We obtained comparable results. In a similar vain we coded somebody as employed only if the person

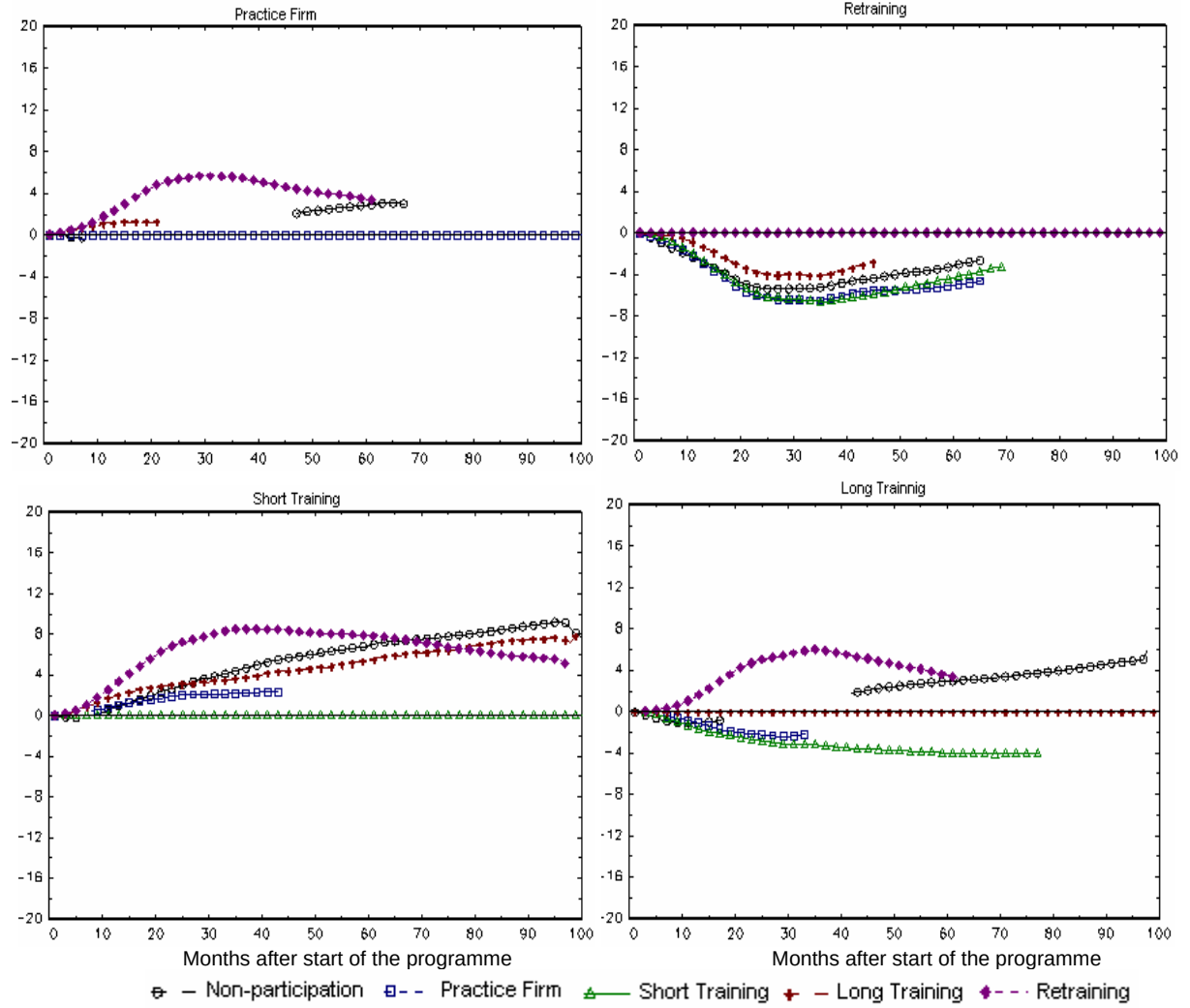
received at least 90% of the earnings of his last job prior to training. It does not change the conclusions (detailed results are available in Internet Appendix ID.1).

It is likely that a substantive programme like *retraining* may not only affect the employment probability, but the productivity of the new job as well. For the latter, earnings are a convenient summary measure. However, the earnings differences (zero earnings if not employed) are very much driven by the employment dynamics and hence, it is not surprising that they confirm the previous results: Eight years after programme start there is a gain in monthly earnings from *retraining* compared to *nonparticipation* of about € 350. The gains from *short* and *long training* compared to *nonparticipation* are positive as well, but about € 150-200 lower than for *retraining* (see Internet Appendix ID.1 for details).

At least from the point of view of the unemployment insurance system, considering the outcome variable registered unemployment is relevant (see Figure ID.4 in the Internet Appendix). We find that none of the programmes dominates nonparticipation systematically, probably because programme participation increases the maximum unemployment benefit entitlement period, so that non-workers have an incentive of remaining registered. Thus, the major effect of the programmes compared to nonparticipation is that they bring those unemployed back to work that would otherwise leave the labour force. For inter-programme comparisons, the previous findings are by and large confirmed.

Figure 2 showed that there are indirect costs of the programmes in terms of the initial negative effects most likely due to lock-in, i.e. a reduced job finding probability during programme participation. A first step of a cost benefit analysis is to compare the initial negative effects to the positive effects that may occur later. To do so, we cumulate the effects over time, starting with the first month of the programme. Figure 3 displays the respective total effects for employment at any point in time during the eight year interval for which we have reliable data. Not surprisingly, the effects appear in a somewhat different light.

Figure 3: Cumulated employment effects ($\hat{\theta}_t^{m,l} = \sum_{\tau=1}^t \hat{\theta}_{\tau}^{m,l}$) in months



Note: See note below Figure 2. Read entry for $\hat{\theta}_t^{m,l}$ as: "On average for participants in m , t months after beginning participation in m , it increased the total time in employment compared to l for by $\hat{\theta}_t^{m,l}$ months."

Figure 3 shows that from this perspective *short training* is the most attractive programme. Short programmes have only a small lock-in effect, and thus their positive effect accumulates much longer, suggesting a gain of about nine months of employment over the eight years following programme start compared to *nonparticipation* and a corresponding gain of almost six months compared to *retraining*. A similar pattern shows up for *long training* compared to nonparticipation, but the level of the effects is somewhat different. There appear to be some positive cumulated effects for *practice firms* compared to *nonparticipation* as well, but they fade out after six years. For *retraining*, eight years are not sufficient to recover fully from the

initial lock-in effect and to create an overall significantly positive effect (compared to all programmes and nonparticipation). Assuming a continuing trend, it seems likely that positive effects appear after about ten years, but of course this projection remains a speculation. Nevertheless, after eight years for participants in retraining it is impossible to conclude which of the available training schemes would be overall most effective for them. For cumulated earnings, which shows a similar pattern as cumulated employment, there is however, a positive and significant gain of about € 10,000 for *retraining* compared to nonparticipation (not discounted). For short and long training the gain is about € 20,000 and € 15,000, respectively. For *practice firms* there is no significant cumulative earnings gain after eight years (see Internet Appendix ID.1 for details).

5.3 Heterogeneity by types of unemployed

We investigate whether groups defined by different exogenous socio-economic characteristics exhibit different effects by stratifying the sample along the dimensions local unemployment rate, town size, gender, unemployment duration, occupational status and education and match within these strata. Internet Appendix ID.1 Table ID.2 displays the results. The number of observations given in the lower part of this table indicates that in many cases the subsample estimates will be too imprecise to uncover significant differences.¹⁴ Despite the problem of precision, some conclusions can be drawn. Compared to *nonparticipation*, all programmes appear more effective for short-term unemployed (less than one year). The most significant differences show up with respect to *practice firms* and gender. They are ineffective for men, but highly effective for women. An explanation could be that the types of *practice firms* men and women mainly attend are different (men: mainly manufacturing; women: mainly commercial), so that we measure different effectiveness for the two types of programmes

¹⁴ We use the MNP estimate from the joint model, but the remaining steps of the estimation are performed in the subsamples. Therefore, the observations do not add up to the number of the observations in the full sample, because the common support criterion must delete more observations if used in subsamples separately (a table with detailed numbers is available in the internet appendix).

instead of a male-female difference. The results for local unemployment rate, town size, occupational status and education suggest that effect heterogeneity does not appear to be present on a massive scale in these dimensions.

5.4 Sensitivity analysis

We performed a profound sensitivity analysis to check the robustness of our results to various implementational issues. Table D.1 in Appendix D summarises the tests we did and their results. The Internet Appendix includes additional details. Here, we discuss the main points.

First, taking up the arguments in Fredriksson and Johansson (2003, 2004) the definition of the nonparticipation status and of the treatment window is important for the interpretation and the validity of our results. The fact that we condition nonparticipation on ending unemployment without entering a programme (or not ending unemployment at all) before 1995 might lead to some bias in our results in favour of the so-defined nonparticipation status. This bias should be severe if nearly every unemployed has to participate. As stated before this is, however, by far not the case in our sample (see Table 5).

To test the sensitivity of our result with respect to the choice of the treatment window we first reduce it by excluding the first six months to the period 07/1993-12/1994 which has mainly the effect of reducing sample size. Second and more importantly with respect to the Fredriksson and Johansson (2003, 2004) arguments, we only considered the period 01/1993-10/1994 which corresponds to reducing the period in which we require nonparticipants not to enter a programme by two months. In both cases we obtain the same qualitative estimation results but loose some precision due to reduced sample sizes (up to 40% in the second case).

We also test a different procedure for simulating programme start dates for nonparticipants. We randomly draw from the empirical distribution of elapsed unemployment durations before programme start of participants and check whether the respective nonparticipant has at least

the same unemployment duration. If so, we assign as hypothetical starting date the month after the drawn unemployment duration.¹⁵ Our results remain almost unchanged.

Third, since the effects for men and women based on the common estimation of the MNP model showed considerable effect heterogeneity (especially for practice firms), more flexibility may be required when estimating the participation probabilities. Therefore, we estimate separate MNPs for men and women but do not find significant differences in the effects compared to the common MNP model. Moreover, we simulate the effects of some important covariates on the estimated participation probabilities by changing the values of these covariates. The results are consistent with our MNP estimates.

Fourth, the common support criterion is made stricter by defining the upper and lower bounds as 10th largest and smallest observation instead of the minimum or maximum, leading to a better match in the tails of the propensity score distribution. Due to the smaller sample sizes (up to 40% of observations per group are deleted from the original sample) some effects are no longer significant, but the conclusions do not change.

Fifth, the additional matching variables other than gender used to define the distance metric in the matching algorithm are not used. The results are qualitatively the same, but in particular for *retraining*, the effects are somewhat smaller and fewer of them are significant. On the other hand, adding elapsed unemployment duration before the programme to the list of matching variables somewhat increases precision but does not change overall results.

Sixth, for selected outcome variables we redid the analysis with simple one-to-one matching as in Gerfin and Lechner (2002). The main change is that several effects are estimated more noisily but the overall conclusions do not change.

Seventh, in the Internet Appendix we present estimates of the asymptotic bias we adjust for in the weighted matching estimator used here. A large and volatile value of it would clearly raise

¹⁵ We thank Barbara Sianesi for suggesting this.

concerns about the adjustment procedure. This is not the case, however, compared to the magnitude of the effects.

Finally, we used three-month moving averages instead of monthly observations for the respective outcome variables. Not surprisingly, the results are a bit 'smoother', but the efficiency gains appear to be very small.

6 Conclusion

For a sample of West German unemployed who enter government-sponsored training in 1993 and 1994 we estimate the short-, medium- and long-run effects of different types of training on employment, unemployment and earnings of their participants. These programmes differ considerably in both the form and the intensity of the human capital investment as well as in their respective duration. The administrative data we use is particularly rich which allows us to correct for selectivity by microeconomic matching methods which accounts for multiple treatments and to measure interesting outcome variables on a monthly basis for a period of eight years after programme entry. We implement a matching procedure that is more efficient and more robust than one-to-one matching because we use a weighted mean of similar comparison observations and we apply a weighted-regression based bias removal procedure. Extensive sensitivity checks show that our estimation results are not sensitive to issues concerning the implementation of the estimator.

Similar to earlier studies (e.g. Fitzenberger and Speckesser, 2005, for Germany; Lechner and Gerfin, 2002, for Switzerland; Jespersen, Munch and Skipper, 2004, for Denmark), we find negative short-run (lock-in) effects on employment whose magnitude and length of occurrence are directly related to programme duration for all types of training. In the longer run after about three to four years, however, most programmes increase employment rates and earnings. But the results for the various programmes differ quite substantially when considering employment eight years after programme start (the longest horizon which is

observable). Retraining, involving the most substantial investment in human capital dominates all other programmes as well as nonparticipation. The gain in employment probability over nonparticipation is about 10%-points. This does not only hold for participants in retraining, but also for participants in other programmes had they participated in retraining. Short and long training dominate nonparticipation with a somewhat smaller gain of about 5-8%-points after eight years. Positive effects for practice firms, if any, appear to be too small to be detectable with our sample sizes.

Focussing on the overall performance over the eight-year period, i.e. netting out positive and negative effects over time, the findings change somewhat, because the different programmes exhibit very different lock-in effects. In this comparison, shorter programmes (up to six months duration) outperform the rest. Compared to nonparticipation, the gain after eight years would be about nine months of additional employment. It is about six months for longer training courses other than retraining. For retraining, the initial lock-in effects are so large that a period of eight years is too short to allow significant positive effects to be detected, although there is a clear trend towards positive overall effects. Finally, no sustainable positive effects are visible for practice firms.

With respect to registered unemployment, even after eight years all programmes except short training increase the duration of benefit receipt compared to nonparticipation. The increase due to retraining is about nine months, for long training three months (though not significant), and practice firms increase benefit receipt by about four months. These numbers point again to the fact that the positive and sustainable employment (and earnings) effects of retraining come at a considerable cost.

Besides providing new insights on the effectiveness of government-sponsored training in West Germany, our results contribute to the more general discussion of shorter and less intense training versus substantial human capital investments for reducing unemployment. In

contrast to the presumption of a potential trade-off between short- and long-term gains of less versus more intense training, we find that both the least intense and the most intense programme types have sustainable positive long-run effects on employment and earnings. However, there is a trade-off in terms of initial lock-in effects. For shorter training this initial negative effect is small while for the most intensive programme, retraining, it is large. Cumulating the monthly effects over the eight-year observation period, for the latter the positive long-run effects are not sufficient to compensate for the large initial lock-in effect. Short training, however, leads to a substantial and significant gain in employment and earnings after the eight-year period.

These findings also help to understand a puzzle that occurred in the previous literature, namely that for most training programmes significantly positive effects cannot be found. If we believe the kind of consensus of the education literature that returns in earnings of one year of full-time schooling are between five and ten percent (Card, 1999), then it would be surprising if programmes for unemployed with a duration of a few months or much less can have effects large enough to be detectable by noisy data, even if the samples are large. Yet, even the group of shorter German programmes are fairly long by international comparisons, and particularly retraining involves rather significant human capital investments that can be expected to be above the 'noise-threshold' level, and are thus detectable in our study.

References

- Abadie, A. and G. W. Imbens (2006): "Large Sample Properties of Matching Estimators for Average Treatment Effects", *Econometrica*, 74(1), 235-267.
- Ashenfelter, O. (1978): "Estimating the Effect of Training Programs on Earnings", *Review of Economics and Statistics*, 6(1), 47-57.
- 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*, 120/4, 649-662.
- Bender, S., J. Hilzendegen, G. Rohwer, and H. Rudolph (1996): "Die IAB-Beschäftigtenstichprobe 1975-1990", *Beiträge zur Arbeitsmarkt- und Berufsforschung*, 197.

- Brodaty T., B. Crépon and D. Fougère (2001): "Using Kernel Matching Estimators to Evaluate Alternative Youth Employment Programs: Evidence from France, 1986-1988", in M. Lechner and F. Pfeiffer (eds.): *Econometric Evaluations of Labour Market Policies*, Heidelberg: Physica Verlag, 85-124.
- Bundesanstalt für Arbeit (BA, 1992-2004): "Arbeitsstatistik - Jahreszahlen", *Amtliche Nachrichten der Bundesanstalt für Arbeit*, 40-52, Special Issue, Nuremberg.
- Card, D. (1999): "The causal effect of education on earnings", in: O. Ashenfelter and D. Card (eds.), *Handbook of Labour Economics*, Vol. 3, 1801-1863, Amsterdam: North-Holland.
- Couch, K.A. (1992): "New Evidence on the Long-Term Effects of Employment Training Programs", *Journal of Labour Economics*, 10(4), 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.
- Dorsett, R. (2005): "The New Deal for Young People: Effect on the Labour Market Status of Young Men", *Labour Economics*, forthcoming.
- Fay, Robert 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.
- 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.
- 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, Institute for the Study of Labor, Bonn.
- Gerfin, M., and M. Lechner (2002): "Microeconomic Evaluation of the Active Labour Market Policy in Switzerland", *Economic Journal*, 112, 854-893.
- Heckman, 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.
- Heckman, J. and J. Smith (1999): "The Pre-Program Earnings Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies", *Economic Journal*, 109, 313-348.
- 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.
- Hujer, R. and Caliendo, M. (2001): "Evaluation of Active Labour Market Policy - Methodological Concepts and Empirical Estimates", in: I. Becker, N. Ott and G. Rolf (eds.): *Soziale Sicherung in einer dynamischen Gesellschaft*, Campus-Verlag, Frankfurt, 583-617.
- Hujer, R., K.-O. Maurer and M. Wellner (1999a): "Analyzing the Effects of On-the-Job vs. Off-the-Job Training on Unemployment Duration in Westdeutschland", in: L. Bellmann and V. Steiner (eds.): *Panelanalysen zur Lohnstruktur, Qualifikation und Beschäftigung, Beiträge zur Arbeitsmarkt- und Berufsforschung*, 229, 203-237.
- Hujer, R., K.-O. Maurer and M. Wellner (1999b): "Estimating the Effect of Training on Unemployment Duration in West Germany - A Discrete Hazard-Rate Model with Instrumental Variables", *Jahrbücher für Nationalökonomie und Statistik*, 218, 619-646.
- ILO (1998): "World Employment Report 1998-1999 – Employability in the Global Economy: How Training Matters". ILO, Geneva.
- Imbens, G. W. (2000): "The Role of the Propensity Score in Estimating Dose-Response Functions", *Biometrika*, 87, 706-710.
- 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. Ten Have, H. I. Feldman, and St. Kimmel (2004): "Model Selection, Confounder Control, and Marginal Structural Models", *The American Statistician*, November, 58-4, 272-279.
- Klose, C. and S. Bender (2000): "Berufliche Weiterbildung für Arbeitslose - Ein Weg zurück in Beschäftigung? Analyse einer Abgängerkohorte des Jahres 1986 aus Massnahmen zur Fortbildung und Umschulung mit einer ergänzten IAB-Beschäftigtenstichprobe 1975-1990", *Mitteilungen aus der Arbeitsmarkt- und Berufsforschung*, 33/3, 421-444.
- Kluve, J. and C. Schmidt (2002), "Can training and employment subsidies combat European unemployment?", *Economic Policy*, 35, 411-448.
- Larsson, L. (2003): "Evaluation of Swedish Youth Labor Market Programs", *Journal of Human Resources*, 38(4).
- Lechner, M. (1999): "Earnings and Employment Effects of Continuous Off-the-job Training in East Germany after Unification", *Journal of Business Economics and Statistics*, 17, 74-90.
- Lechner, M. (2001): "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption", in: M. Lechner and F. Pfeiffer (eds.), *Econometric Evaluation of Active Labour Market Policies*, 43-58, Heidelberg: Physica.
- Lechner, M. (2002a): "Programme Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labour Market Policies", *Review of Economics and Statistics*, 84, 205-220.
- Lechner, M. (2002b): "Some Practical Issues in the Evaluation of Heterogeneous Labour Market Programmes by Matching Methods", *Journal of the Royal Statistical Society - Series A*, 165, 59-82.
- Lechner, M. (2004): "Sequential Matching Estimation of Dynamic Causal Models", Discussion Paper 2004-06, Department of Economics, University of St. Gallen.
- Lechner, M., R. Miquel and C. Wunsch (2004): "Long-run effects of Public Sector Sponsored Training in West Germany", Discussion Paper 2004-19, Department of Economics, University of St. Gallen.
- 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. and M. Lechner (2001): "Identification of Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions", Discussion Paper, Department of Economics, University of St. Gallen.
- Miquel, R., C. Wunsch and M. Lechner (2002): "Die FuU-Teilnehmer-Datei 1976-1997", Graues Papier des Instituts für Arbeitsmarkt- und Berufsforschung, Nuremberg.
- OECD (1996a): "Lessons from Labour Market Policies in the Transition Countries", OECD, Paris.
- OECD (1996b): "The Public Employment Service: Austria, Germany, Sweden", OECD, Paris.
- Rubin, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66, 688-701.
- 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(1), 133-155.
- Speckesser, S. (2004): "Using Social Insurance Data for the Evaluation of Active Labour Market Policy: Employment Effects of Further Training for the Unemployed in Germany", University of Mannheim, Unpublished Manuscript.
- 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: Sample selection rules

Table A.1 shows the different sample selection steps we apply and how this shrinks our sample accordingly. Simulating programme start dates for nonparticipants and ensuring that this date lies after the beginning of the 'defining' unemployment spell and within the period the person is observed (if this ends before December 1994) only reduces the number of nonparticipants. The strict definition of eligibility affects all groups but nonparticipants to the largest extend. But since this group is large anyway and these steps have the objective of making them comparable to participants, this is no problem.

Table A.1: Sample selection rules and remaining numbers of observations

	Nonparticipation	Practice firm	Short training	Long training	Retraining	Career improv.	Other
Persons entering unemployment between Jan. '93 and Dec. '94	36965	324	644	380	497	130	103
Simulated programme start after the entry in unemployment (UE) and before the end of the observation period							
Remaining observations	26022	324	644	380	497	130	103
Eligibility: Only individuals receiving UB or UA in the month of and before the programme start							
Remaining observations	13091	309	618	350	450	118	92
Personal characteristics : a) $20 \leq \text{age} \leq 55$; b) no trainees or apprentices; c) at least one observation of employment; d) no home workers; e) no part-time worker less than half of a full-time work							
Final sample	9197	273	572	329	413	110	74

Note: All variables are measured before or in the same year as the start of the programme.

Appendix B: Data

Table B.1: Descriptive statistics

	Nonparticipation	Practice firm	Short training	Long training	Retraining	Career improve.	Other
Number of observations	9197	273	572	329	413	110	74
Proportions in %							
Personal characteristics							
Women	41	33	37	39	38	24	27
Older than 50 years	15	7	6	5	1	3	1
Younger than 26 years	14	19	15	15	21	21	22
Age*	37	36	35	35	31	32	32
Nationality: German	81	87	91	92	89	92	92
Western European	12	7	6	5	7	3	5
Eastern European	4	3	2	1	2	3	1
Other	3	3	1	3	2	3	1

Table B.1 to be continued

Table B.1: Descriptive statistics (continued-1)

	Nonparticipation	Practice firm	Short training	Long training	Retraining	Career improve.	Other
Marital status: Single	48	60	58	59	62	64	53
Married	52	40	42	41	38	36	47
Children: No child	62	73	64	65	61	72	58
At least one child	38	27	36	35	39	28	42
Education							
No university entrance degree, no professional degree (PD)	25	18	15	9	24	5	5
No university entrance degree, with PD	65	78	75	71	68	80	81
University entrance degree, no PD	1	1	1	2	1	1	1
University entrance degree and PD	3	2	3	5	3	5	8
Polytechnical degree	2	0	3	5	1	5	1
University degree	4	0	3	9	2	5	1
Position in last job							
Salaried employee	28	34	38	57	23	41	39
Part-time worker	10	5	8	7	8	4	4
Master craftsman	1	1	1	0	0	1	1
Unskilled worker	40	36	31	19	53	9	2
Skilled worker	21	23	23	16	16	45	35
Industrial sector							
Construction	8	5	6	5	5	14	9
Commerce	16	16	20	16	12	12	15
Banking, insurance	1	3	1	3	2	1	3
Local, regional authorities, social insurance	4	5	5	7	6	4	1
Non-profit organisations, private households	3	3	2	3	4	3	0
Agriculture, forestry, fishing	1	3	1	1	1	0	0
Energy and supply industry, mining	0	1	0	1	0	1	0
Manufacturing (without construction)	38	40	38	38	41	48	49
Transportation, telecommunications	5	4	5	4	7	3	7
Other services	23	20	21	22	22	15	16
Last monthly earnings							
Salary in Euros*	1680	1640	1773	1889	1640	2072	1781
No information++	11	3	4	4	3	5	3
Regional information							
Big city (at least 300,000 inhabitants)	26	18	24	29	20	26	7
North (Hamburg, Bremen, Schlesw.-H.)	22	30	19	25	26	18	15
North-Rhine-Westphalia	31	21	28	36	35	23	28
Rhineland-Palatinate, Hesse, Saarl.	17	23	19	18	17	17	19
Baden-Wuerttemberg, Bavaria	30	26	34	21	22	42	38
Local UE rate ≤ 5% **	1	2	2	0	1	2	4
Local 5% < UE rate ≤ 10% **	64	67	70	62	61	72	76
Local UE rate > 10% **	35	31	28	38	38	26	20
Last occupation in/as							
Agriculture, forestry, fishing	2	2	2	3	3	0	0
Plumbing, metal construction technology	9	15	11	7	9	19	7
Food and nutrition	4	2	4	1	4	2	4
Construction, woodworking	10	8	8	5	9	15	15
Merchant (goods and services)	9	12	12	9	6	5	7
Transportation, storage	13	12	14	9	15	3	15
Administration, office work, business, social scienc.	14	22	22	31	13	17	16
Health services	2	1	1	1	3	2	3

Table B.1 to be continued

Table B.1: Descriptive statistics (continued-2)

	Nonparticipation	Practice firm	Short training	Long training	Retraining	Career improve.	Other
Last occupation in/as							
Hairdressing, guest assist., housekeeping, cleaning	8	6	3	2	7	3	1
Chemical worker, polymer processing	3	1	1	2	6	1	0
Unskilled worker	2	4	2	1	2	0	1
Metal production and processing	4	5	4	2	6	2	8
Textile, leather, clothing	3	1	1	2	2	0	1
Security services	1	1	1	1	1	4	0
Paper manufacture and processing, printing	2	2	2	2	2	4	1
Social services, education, counselling	3	0	0	2	4	2	1
Media, humanities, arts	1	0	1	1	0	0	1
Mining	0	0	0	1	0	0	0
Technology, natural sciences	4	1	6	16	2	15	14
Machinist	0	0	0	0	0	1	0
Electronics	4	2	3	4	4	7	4
Stone, ceramics, glass making and/or processing	1	1	1	0	1	0	0
Remaining unemployment (UE) benefits claim at the end of the last unemployment spell before entry in the programme							
Remaining UE benefits claim (in months)*	8.1	7.4	7.8	7.2	5.8	6.8	6.5
No information or no claim	44	48	44	42	38	27	31
Legal UE benefits claim at the beginning of the last unemployment spell before the programme							
Legal claim* (months)	12.6	11.0	11.5	10.6	9.3	10.1	9.6
No information	26	29	28	24	25	17	20
No claim	4	5	3	4	3	3	3
Unemployment benefits or assistance in the month before the beginning of the programme							
UE benefits	73	71	72	76	75	83	80
UE assistance	27	29	28	24	25	17	20
Various historical un/ out-of/employment information before the "first unemployment period"							
Months of last employment spell*	50	44	51	49	40	48	45
Proportion of employment months (in %)*	72	70	72	72	70	80	75
Proportion of out-of-labour months (in %)*	13	12	13	13	16	11	10
Proportion of UE months (in %)*	11	12	10	9	9.03	43	8
# of programs up to 2 years before the UE period (UEP)*	0.09	0.11	0.09	0.12	0.12	0.15	0.08
# of programs up to 5 years before the UEP*	0.18	0.25	0.23	0.22	0.26	0.19	0.22
# of programs from entry in the data up to UEP*	0.28	0.45	0.39	0.40	0.34	0.25	0.41
Mean duration of UE spells up to 2 years before UEP*	1.48	1.74	1.30	1.19	1.23	0.69	0.87
up to 5 years before UEP*	2.77	3.05	2.31	2.03	2.30	1.16	1.77
from entry in data to UEP*	3.95	4.09	3.5	3.3	2.95	2.13	3.04
Mean duration of employment spells up to 2 years before UEP*	3.7	3.5	3.3	4.6	4.8	4.8	3.8
up to 5 years before UEP*	12.1	12	12.4	14.8	13.2	13	16.5
from entry in data to UEP*	48.6	42.0	48.6	47.5	35.7	47.3	41.5
Mean duration of out-of-labour spells up to 2 years before UEP*	2.20	1.91	1.70	1.70	1.39	1.11	1.62
up to 5 years before UEP*	3.61	3.22	2.95	2.94	3.60	3.19	3.86
from entry in data to UEP*	7.58	5.64	7.66	9.25	8.82	6.72	6.58
Total months in all programmes up to 2 years before the UEP*	0.73	0.52	0.47	0.76	0.70	0.63	0.46
5 years before the UEP*	1.32	1.54	1.55	1.53	1.48	0.95	1.54
before entry in the sample*	2.07	2.89	2.85	2.73	2.04	1.65	2.58

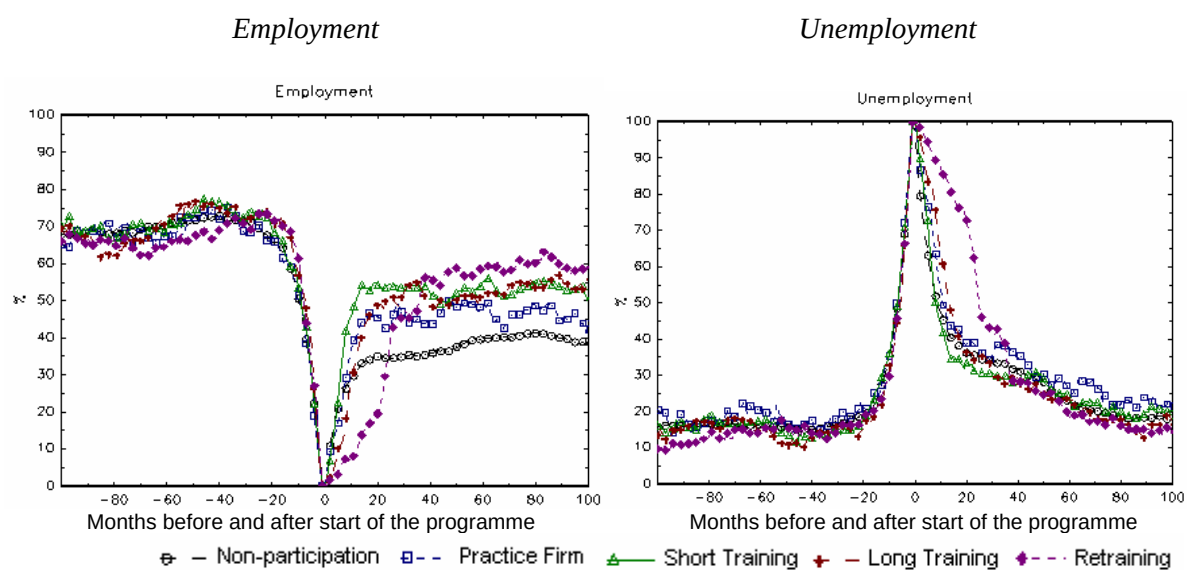
Table B.1 to be continued

Table B.1: Descriptive statistics (continued-3)

	Nonpar- ticipation	Practice firm	Short training	Long training	Retraining	Career improve.	Other	
Programme information								
Planned programme duration* (months)		5.95	3.82	9.74	20.43	9.91	5.09	
----- Various un/employment information from the "first unemployment period"								
Duration of the "first UE spell"*	7.60	6.55	6.91	6.59	5.56	5.26	4.42	
Duration of last UE spell before prog.*	6.75	5.49	5.57	5.30	4.49	4.25	3.95	
Time since beginning of last UE spell (before prog.) even if other state be- tween UE and prog.*	6.77	6.60	7.08	6.60	5.61	5.60	4.85	
Time between the prog. and last job* time between programme - last job ≤ 3 months	15.0	13.4	13.8	12.3	11.0	9.7	10.8	
	16	11	16	17	22	26	20	
≤ 6 months	35	37	39	43	43	46	59	
≤ 12 months	62	67	67	67	74	75	78	
≤ 24 months	84	88	91	90	92	95	92	
Transition in 6 months before prog.								
UE. → UE	58	56	61	56	54	44	42	
empl. → UE	26	27	27	33	32	39	45	
out → UE	14	16	10	10	10	8	11	
prog. → UE	2	1	1	1	4	9	3	
# of prog. in year before actual progr.*	0.04	0.03	0.02	0.03	0.07	0.14	0.05	
# of prog. in 6 months before act. pr.*	0.02	0.01	0.01	0.02	0.04	0.10	0.03	
----- Firms size of the last employer								
No information	4	4	4	3	3	4	5	
1 to 9 employees	21	21	21	19	18	27	27	
10 to 99 employees	36	36	38	36	35	31	30	
100 to 499 employees	21	25	22	23	23	24	26	
500 employees or more	19	14	15	18	21	15	12	
----- Timing of programme and appearance in data on average (average month)								
Date of entry in the data*	Oct. 83	Dec. 83	Jun. 83	Aug. 83	May 85	Jul. 84	Apr. 84	
Date of UE spell defining treat. status*	Oct. 93	Aug. 93	Sep. 93	Sep. 93	Sep. 93	Jul. 93	Sep. 93	
Date of begin of prog.*	Jul. 94	Mar. 94	May 94	May 94	Mar. 94	Jan. 94	Mar. 94	
----- Outcome+								
Unemployment in Jan. 1990	9559	17.7	21.5	20.4	16.6	15.3	5.9	18.0
1992	10609	19.1	21.3	17.9	15.5	15.9	9.3	17.4
1993	10870	24.7	30.8	27.9	21.9	31.5	30.0	27.4
1994	10962	66.2	77.7	72.9	70.5	74.8	78.2	52.7
1995	10968	64.9	60.4	64.2	73.6	86.4	60.9	16.2
1997	10968	36.8	39.9	32.0	30.4	40.7	18.1	14.9
1999	10968	28.7	30.8	28.7	23.7	24.9	15.5	20.3
2001	10968	19.4	24.5	21.3	15.5	17.0	10.0	12.2
Employment in Jan. 1990	9559	69.5	68.0	67.0	71.3	66.9	81.4	70.5
1992	10609	69.1	67.5	70.0	73.3	68.5	80.6	68.1
1993	10870	60.9	55.7	60.5	64.7	57.8	62.7	63.0
1994	10962	24.1	17.2	22.7	22.5	20.3	14.6	41.9
1995	10968	17.9	28.6	29.9	19.8	9.4	32.7	73.0
1997	10968	30.1	44.0	49.3	52.6	46.3	60.0	70.3
1999	10968	33.5	46.5	49.5	51.4	56.7	56.4	59.5
2001	10968	36.2	46.2	53.2	55.0	60.0	60.9	59.5

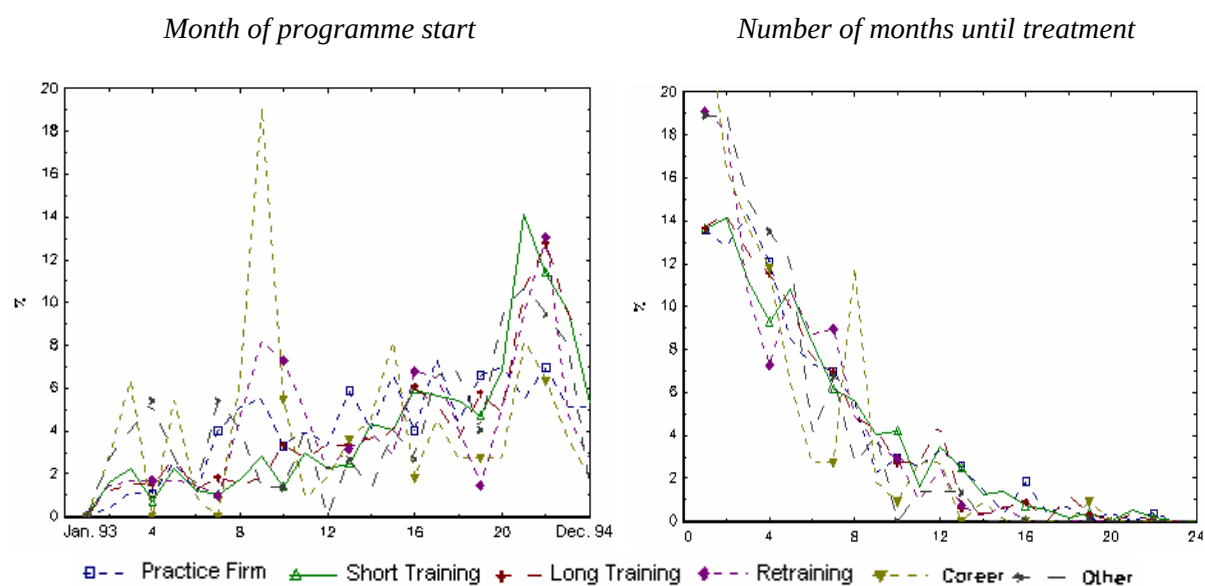
Note: Sample before imposing the common support requirement. *Means rather than proportions. **Local unemployment rates for the 141 local labour office districts. +The different outcomes do not add up to 100% because of some missing values. ++ The category 'No information' includes cases with missing earnings information and with the entry '0'. Zero entries are made for so-called inactive employment which includes women on maternity leave, men in the military or civil service, as well as employees having been ill for more than six weeks. The first column: number of observations used to compute the proportions. The sample size decreases due to different entry dates into the sample (first UE spell in 93/94) and exit dates from the sample.

Figure B.1: Employment and unemployment rates by participation status



Note: Timing relative to observed or simulated starting dates of programmes. Note that after 80 months the sample size starts to decline rapidly.

Figure B.2: Monthly distribution of start dates (in %)



Note: The treatments *other* and *career* are not considered in the evaluation below. The right panel shows the number of months until participation after the beginning of the 'defining' unemployment spell.

Appendix C: Technical details of the matching estimator used

Table C.1: A matching protocol for the estimation of $\theta_0^{m,l}$

Step 1	Specify and estimate a multinomial probit model to obtain the marginal choice probabilities (see Internet-Appendix): $[\hat{P}_N^0(x), \hat{P}_N^1(x), \hat{P}_N^2(x), \hat{P}_N^3(x), \hat{P}_N^4(x)]$; $\hat{P}_N^l(x) := \hat{P}_N(S = l X = x)$
Step 2	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 S.
Step 3	<i>Estimate the respective (counterfactual) expectations of the outcome variables.</i> For a given value of m and l the following steps are performed: Standard propensity score matching step (multiple treatments) a-1) Choose one observation in the subsample defined by participation in m and delete it from that pool. b-1) Find an observation in the subsample of participants in l that is as close as possible to the one chosen in step a-1) in terms of $[\hat{P}_N^m(x), \hat{P}_N^l(x), \tilde{x}]$. '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 participant in m is left. Exploit thick support of X to increase efficiency (radius matching step) d-1) Compute the maximum distance (d) obtained for any comparison between treated and matched comparison observations. a-2) Repeat a-1). b-2) Repeat b-1). If possible, find other observations in the subsample of participants in l 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 m is left. d-2) For any potential comparison observation, add the weights obtained in a-2) and b-2). 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 $y^l(x_i)$ of every observation in l and m using the coefficients of this regression: $\hat{y}^l(x_i)$. f-2) Estimate the bias of the matching estimator for $E(Y^l S = m)$ as: $\sum_{i=1}^N \frac{1(S = m) \hat{y}^l(x_i)}{N^m} - \frac{1(S = l) w_i \hat{y}^l(x_i)}{N^m}$. g) Using the weights obtained by weighted matching in d-2), compute a weighted mean of the outcome variables in l. Subtract the bias from this estimate. Final estimate h) Compute the treatment effect by subtracting the weighted mean of the outcomes in the comparison group (l) from the weighted mean in the treatment group (m).
Step 4	Repeat Step 3 for all combinations of m and l.

Note: Lechner (2001) suggests an estimator of the asymptotic standard errors for $\hat{\theta}_N^{m,l}$ conditional on the weights that we use here. $\tilde{x}$ includes the date of the beginning of the programme, sex, three dummies indicating if the individual is employed (and observed) 12, 24 and 48 months before the programme. $\tilde{x}$ 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). Note that once we estimate all $E(Y^l | S = m)$ for all m, they can be directly used to obtain $E(Y^l)$.

Appendix D: Sensitivity Analysis

Table D.1: Sensitivity tests

Issue	In study	Alternative tested / sensitivity check	Result
Treatment window	Jan. 1993 - Dec. 1994	(a) Jul. 1993 - Dec. 1994	Same qualitative results but loss of some precision (up to 8% less observations per group)
		(b) Jan. 1993 - Oct. 1994	Same qualitative results but loss of precision (up to 25% less observations per group)
Programme start dates for non-participants	Random draw from empirical distribution of start dates of participants	Random draw from the empirical distribution of elapsed unemployment durations before the programme of participants, assign month after this duration	Same qualitative results
Estimation of the propensity scores	Multinomial probit (MNP)	(a) Separate MNP estimation for men and women	No significant differences compared to the common model
		(b) Simulation for different values of the covariates: (bi) all men vs. all women; (bii) all German vs. all foreigner; (biii) all low-skilled vs. all high-skilled; (biv) all unemployed vs. all employed 5 years before 1993	Consistent with the results from the original MNP estimation
Common support for the propensity scores	Maximum of the minima as lower bound, minimum of the maxima as upper bound	10th smallest observation as lower bound, 10th largest observation as upper bound	Same qualitative results but loss of precision (up to 40% less observations per group)
Matching variables	Propensity scores, gender, month of programme start, employed in month 12/24/48 before the programme	(a) Same as before plus elapsed unemployment duration before the programme	Same qualitative results, gain of some precision
		(b) Propensity scores, gender	Same qualitative results but loss of some precision (especially for retraining)
Matching procedure	Radius matching with bias adjustment	(a) One-to-one matching without bias adjustment	Same qualitative results but loss of some precision (especially for practice firms)
		(b) Analyse estimated bias from adjustment procedure	Estimated bias small and stable over time
Outcome variables	Monthly observations	Three-month moving averages	Same qualitative results, gain of some precision