Evaluation of Training for the Unemployed - New Evidence on Effect Heterogeneity, Dropouts, and Program Duration

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften der Universität Mannheim

Marie Elina Waller

vorgelegt im Juli 2009

Referent:	Prof. Bernd Fitzenberger, Ph.D.
Korreferent:	PD Dr. Friedhelm Pfeiffer
Abteilungssprecher:	Prof. Tom Krebs, Ph.D.
Tag der mündlichen Prüfung:	14. Dezember 2009

Acknowledgements

First and foremost, I am deeply grateful to my supervisor Bernd Fitzenberger. Initially as a graduate student in his classes at CDSE Mannheim and later on working in his team at Goethe-University Frankfurt, at ZEW and finally at Albert-Ludwigs-University Freiburg, I benefited enormously from his guidance, his support, his suggestions and his advice on countless occasions. Our numerous discussions formed my understanding of empirical work and labor economics.

I would like to thank Friedhelm Pfeiffer for helpful comments and for being the second supervisor of my dissertation. Also, I would like to thank my co-authors for our productive cooperation. To the Center for Doctoral Studies in Economics (CDSE) at the University of Mannheim I am thankful for the opportunity to being part of their graduate program I strongly benefited from. The empirical work in this dissertation is based on data from the Institute of Employment Research (IAB). I am grateful to the IAB, especially to Stefan Bender, for making the data available. Financial support by the IAB and the Deutsche Forschungsgemeinschaft is gratefully acknowledged.

My time at CDSE Mannheim was an outstanding experience to me due to the exceptional friendship and mutual support I experienced from the co-students of my year. I would like to thank in particular Stefanie Brilon, Hans-Martin von Gaudecker, Alia Gizatulina, Florian Müller, and Melanie Schienle.

I am grateful to the people I met at ZEW for their suggestions and to my colleagues at Frankfurt, namely Martin Biewen, Karsten Kohn, Aderonke Osikominu, and Robert Völter, from whom I learned a lot, in particular on empirical research. Also at Freiburg, I benefited from working in a friendly and lively environment.

To my parents, I want to express my strong gratefulness. Among countless other things, I am grateful to them for having prepared me so well for university education. Dennis lived with me the whole experience of this dissertation - my deep gratitude goes out to him.

Contents

General Introduction

The effectiveness of training programs for the unemployed, as well as other parts of active labor market policies (ALMP), has been an important topic in the international literature in labor economics over the last decades. There are still many open issues, but fast progress has been made in recent years. This was enabled by advancements in microeconometric methods as well as by a rising interest of policymakers in econometric evaluations which enhanced the composition of data sets based on process generated data made available to researchers (see for example Card, Kluve, and Weber (2009)).

This statement applies not only internationally, but also to the particular situation in Germany. The Federal Employment Office of Germany is offering a wide range of ALMP and in particular different types of training, ranging from short programs which essentially aim at activating the unemployed, to further training programs intending to considerably increase the human capital of participants, to very long retraining programs which lead to a degree in a new profession. Each year there are more than one million entries into public sponsored training programs. Thus, it is of strong interest whether these programs are effective, for whom and under which circumstances. Furthermore, because of the wide range of large-scale programs offered, the German ALMP constitute a fruitful field of study to labor economists interested in analyzing different aspects of the effectiveness of ALMP.

When I started working on the dissertation project at the beginning of 2005, comparatively little was known about the effectiveness of these programs. Most of the existing studies were based on small data sets and relatively restrictive econometric methods. In most cases it was not possible to distinguish between different groups of participants or different types of training programs. These studies mostly found no or only very small positive effects of training programs.[1] While almost all of the early studies had to rely on data suffering from major constraints (i.e. small sample size, poor definition of program participation, no possibility to distinguish between different programs), in 2004 first results had become known from extensive projects in Germany which aimed at producing and utilizing large and rich research-data from different administrative data sources of the Federal Employment Agency.[2] One of

[1]For a literature survey on the evaluation of German training programs until the beginning of 2005, see Schneider et al. (2006). Heckman et al. (1999) survey the early international literature.

[2]First studies using administrative data from the Federal Employment Agency are Fitzenberger,

these projects produced the first versions of the so-called Integrated Employment Biographies Sample (IEBS). The IEBS is a large and rich data set combining data from four administrative data sources. These data allow to study various questions related to unemployment and ALMP. Its availability enhanced research on these topics and also made the empirical work of this dissertation possible.[3]

Recent methodological progress strongly enhanced the quality of microeconometric evaluation studies as well. Let me name three exemplary topics which have been very influential for applied work, including this dissertation. First, the work of Imbens (2000) and Lechner (2001) on pairwise comparison of multiple treatments provides the framework to compare different programs, i.e. to answer the question what would have happened to a participant if he or she had been assigned to a program that differs from the one he or she is actually assigned to. Second, much progress has been made with regard to program evaluation in dynamic settings. In a dynamic setting, the so-called timing-of-events becomes important as discussed by Fredriksson and Johansson (2003) and Sianesi (2004). Static treatment evaluations implicitly condition on future outcomes leading to possibly biased treatment effects. The nontreated individuals in the data might be observed as nontreated because their treatment starts after the end of the observation period or because they exit unemployment before treatment starts (Fredriksson and Johansson (2003)). Third, Abbring and van den Berg (2003) propose an estimation strategy which explicitly uses the timing-of-events in a dynamic setting to identify the treatment effect applying a continuous duration model. Apart from these innovations which relate directly to program evaluation, this dissertation makes use of relatively recent progress in Markov Chain Monte Carlo (MCMC) methods, a technique which has been advanced by Bayesian statisticians in particular in the 1990s (see Chib (2001) for an overview).

The five chapters of this dissertation represent five stand-alone research papers. They cover various aspects of the evaluation of training programs like effect heterogeneity, comparison of different program types, data quality, dynamic selection into programs and out of programs, occurrence and employment perspectives of dropouts,

---

Osikominu, and Völter (2008), Fitzenberger and Speckesser (2007), Hujer, Thomsen, and Zeiss (2006), Lechner, Miquel, and Wunsch (2007, 2009). Klose and Bender (2000) use part of these data even at an earlier time.

[3]Meanwhile there are a couple of other studies using the IEBS and focussing on training: Kluve, Schneider, Uhlendorff, and Zhao (2007), Lechner and Wunsch (2006), Osikominu (2008), Rinne, Schneider, and Uhlendorff (2007), Rinne, Uhlendorff, and Zhao (2008), Schneider and Uhlendorff (2006), and Wunsch and Lechner (2008).

and the effect of different program lengths. Different state-of-the-art econometric methods are applied and some methodological extensions with regard to applied work are made.

Chapter 1 provides an extensive microeconometric evaluation of training for the unemployed in Germany during the period February 2000 to January 2002. It is joint work with Martin Biewen, Bernd Fitzenberger, and Aderonke Osikominu.[4] Building on the work of Sianesi (2004) on dynamic treatments and on the work of Imbens (2000) and Lechner (2001) on pairwise comparison of multiple treatments, we employ a stratified kernel matching approach taking into account the dynamic sorting processes. We compare the effectiveness of different types of programs and consider effect heterogeneity with respect to population subgroups. In addition, we propose a strategy to detect effect heterogeneity within subgroups. From a policy point of view, we address two main questions: the question whether relatively short training programs can compete with more involved medium- to long-term measures, and the question whether practically oriented training programs have advantages over theoretically oriented class-room training. Our methodology allows us to directly compare training programs, i.e. to find out what would have happened if participants in one program type had participated in a different program type. The results suggest that in West Germany both short-term and medium-term programs may have considerable employment effects for certain population subgroups, and that short-term programs are surprisingly effective when compared to the traditional and more expensive longer-term programs. There is evidence that the effects decline for older workers and for low-skilled workers. With a few exceptions, we find little evidence for significant treatment effects in East Germany.

The second chapter is on data quality.[5] With administrative data becoming increasingly important for empirical research, the quality of crucial variables of process generated data is of growing interest. The IEBS has become the most important data set for microeconometric labor market policy evaluation in Germany, in particular it is the basis for government conducted evaluation of labor market reforms. Being among the first to use the IEBS, our team was involved in comprehensive data

[4] Biewen, M., B. Fitzenberger, A. Osikominu and M. Waller (2007), Which Program for Whom? Evidence on the Comparative Effectiveness of Public Sponsored Training Programs in Germany, IZA Discussion Paper No. 2885. This paper was completed in June 2007 and reflects the state of the literature at that time.

[5] Waller, M. (2008), On the Importance of Correcting Reported End Dates of Labor Market Programs, Schmollers Jahrbuch 128, 213-236. This paper was completed in August 2007 and reflects the state of the literature at that time.

checks. I found that reported end dates of further training programs are a sensitive part of the IEBS. Therefore, I investigated this aspect in detail. The aim is to get insights on how to handle this problem in future studies and - more generally - on how measurement error in program end dates affects evaluation results. A considerable part of end dates of further training programs are later than the actual end of participation. Since measurement error in end dates may influence evaluation results through several channels, it is difficult to predict ex ante how results will be affected. But the IEBS has the advantage that due to its special feature of including data from different administrative processes, it is possible to correct almost all relevant end dates of further training programs. I introduce three approaches to deal with error-prone end dates. Their impact on evaluation results is studied for different estimation frameworks using sensitivity analysis. I come to the conclusion that measurement error in program end dates has negligible to modest effects on estimation results depending on the estimation framework.

Chapter 3 is joint work with Bernd Fitzenberger, Olga Orlyanskaya, and Aderonke Osikominu and focusses on short-term training.[6] In recent years ALMP have placed a greater emphasis on activating the unemployed in the short run (see for example OECD (2007)). In Germany, this is reflected in the introduction of short-term training programs at the end of the 1990s and in particular in a strong increase of entries into these programs in the early 2000s. But programs of this type are not a novelty: similar programs have already been part of German ALMP in the 1980s and in the beginning of the 1990s. Our intention is to find out if short-term training causes lasting positive effects on employment outcomes, if participation in these programs leads to higher participation in longer training programs afterwards, and how the results compare in between both time periods. Our study is the first to estimate the employment effects of the short-term training programs of the 1980s and 1990s using state-of-the-art methods. Furthermore, our work for the first time uses administrative data covering such a long time period, namely 18 years in the 1980s and 1990s and four years in the early 2000s to study the medium- and long-term employment effects of short-term training.[7] Whereas one important goal of modern short-term training is to check the willingness to work of the participants, the older programs focus exclusively on job search assistance, limited training, and guidance

[6]Fitzenberger, B., O. Orlyanskaya, A. Osikominu and M. Waller (2008), Déjà vu? Short-Term Training in Germany 1980-1992 and 2000-2003, IZA Discussion Paper No. 3540. This paper was completed in May 2008 and reflects the state of the literature at that time.

[7]This chapter uses not only the IEBS but also additional data to cover the earlier period of interest.

towards future participation in longer training programs. Thus, to compare the programs of both periods, we distinguish two versions of short-term training in the 2000s: a training variant which focuses on skill provision and the checking variant which focuses on testing the willingness to work. We compare the effects of these two versions of modern short-term training and relate the training variant to the programs of the earlier period. Our main findings are that in most cases short-term training shows persistently positive and often significant employment effects. The checking variant leads to slightly smaller effects compared to the pure training variant. The lock-in periods lasted longer in the 1980s and 1990s compared to the early 2000s. Short-term training leads to an increased future participation in more involved training programs, in particular in the earlier period.

When working on the first chapters of this dissertation, I got the impression from the data that a considerable part of participants drop out of the program instead of completing it. But to my surprise I found that almost nothing is known about the number, the characteristics, and the labor market prospects of dropouts in the literature. Thus, I decided to contribute to filling this gap. Chapter 4 is the first study that sheds light on dropouts from training programs in western countries in a non-experimental setting.[8] Dropouts will have a head start on the labor market, because they may already be employed while the other participants are still attending the program. But how about the medium-term and long-term effects of dropout: does it harm to drop out in the long run? From a policy perspective knowledge about the occurrence and the labor market prospects of dropouts may be important, as institutional settings like benefits during program participation and sanctions may influence the number of those who drop out. Furthermore, studying the labor market prospects of dropouts may provide further insights in understanding the composition of average treatment effects of training programs. To estimate the effect of dropout requires to overcome two main obstacles. First, data allowing to identify which participants drop out of the program are needed. I propose a strategy to identify dropouts of German training programs using the IEBS. It turns out that one out of five participants of further training programs drops out of the program. Second, to estimate the effect of dropping out versus completing the program it seems necessary to take into account observable and unobservable

[8]Waller, M. (2009), Many Dropouts? Never Mind! - Employment Prospects of Dropouts from Training Programs, not published. (An earlier and descriptive version of this study has appeared as: Further Training for the Unemployed - What Can We Learn about Dropouts from Administrative Data?, FDZ Methodenreport 04/08, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg, 2008.)

differences between dropouts and non-dropouts as well as state dependence and duration dependence. I estimate the medium-term to long-term effect of dropout using a bivariate dynamic random effects probit model. The model consists of a dropout equation and an employment equation. Both equations include an unobserved individual effect and these two random effects are allowed to be correlated. I estimate the two equations simultaneously using Bayesian MCMC methods. To implement this, I program a Gibbs sampling algorithm. The MCMC simulation provides Bayesian estimates of the posterior distribution of all parameters of the model including the random effects. In order to estimate the size of the dropout effect, I propose a strategy to calculate average partial effects on the treated which account for the selection based on unobservables. This is possible due to the availability of the predictions of the individual random effects. Results suggest only small effects, thus I conclude that on average the decision to drop out neither harms nor enhances the future employment prospects of participants.

Chapter 5 is joint work with Bernd Fitzenberger and Aderonke Osikominu.[9] We propose to evaluate long-term training programs by using a very flexible discrete-time transition model which we estimate by MCMC methods and which is identified by the timing-of-events in an analogous way as the model proposed by Abbring and van den Berg (2003). Using this approach we are able to estimate employment effects both of the incidence and the duration of training, we account for unobserved heterogeneity, and we are able to estimate various treatment effects of interest. More specifically, we estimate a two-equation dynamic random effects probit model for discrete transitions between employment and non-employment as well as for entry into and exit from training. Selection on unobservables is accounted for by allowing the random effects of both equations to be correlated. We account for observable characteristics, state dependence, duration dependence, and interaction effects. The impact of training is modeled in a very flexible way in order to avoid strict functional form restrictions and to allow for effect heterogeneity of different forms. We estimate the employment equation and the training equation simultaneously using MCMC methods. Next, we suggest a simulation approach which allows to estimate the posterior distribution of various treatment effects of interest: like the average treatment effect on the treated (ATT) of participating instead of not-participating, the ATT of participating versus waiting, and the effect of different planned program lengths. The simulation approach makes use of the predicted random effects provided by

[9] Fitzenberger, B., A. Osikominu and M. Waller (2009), The Heterogeneous Effects of Training Incidence and Duration on Labor Market Transitions, not published.

the MCMC estimation and thus accounts for the selection based on unobservables. Furthermore, with the predicted random effects at hand, we may assess explicitly the selectivity of the treated and the nontreated individuals. Compared to previous studies on long-term training, our results suggest strong positive treatment effects for the treated on unconditional employment rates for men and women living in West and East Germany, respectively. This finding is consistent with an estimated negative selection of training participants. The effect of treatment versus waiting is positive in the medium and long run but it is smaller than the treatment effect of participating instead of not-participating. Finally, increasing the planned duration of training turns out to have a positive effect on the treatment effect.

References

Abbring, J. and G. J. van den Berg (2003). "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica* 71, 1491-1517.

Card, D., J. Kluve, and A. Weber (2009). "Active Labor Market Policy Evaluations: A Meta-Analysis." IZA Discussion Paper No. 4002.

Chib, S. (2001). "Markov Chain Monte Carlo Methods: Computation and Inference." In J. Heckman and E. Leamer (eds.): *Handbook of Econometrics*, Volume 5, Amsterdam: Elsevier Science, 3569-3649.

Fitzenberger, B., A. Osikominu and R. Völter (2008). "Get Training or Wait? Long-run Employment of Training Programs for the Unemployed in West Germany." *Annales d'Economie et de Statistique*, forthcoming.

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

Fredriksson, P. and P. Johansson (2003). "Program Evaluation and Random Program Starts." Institute for Labour Market Policy Evaluation (IFAU), Uppsala, Working Paper, 2003:1.

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

Hujer, R., S. Thomsen and C. Zeiss (2006). "The Effects of Vocational Training Programmes on the Duration of Unemployment in Eastern Germany." *Allgemeines Statistisches Archiv* 90, 299-322.

Imbens, G. (2000). "The Role of the Propensity Score in Estimating Dose-Response Functions." *Biometrika* 87, 706-710.

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*, 421-444.

Kluve, J., H. Schneider, A. Uhlendorff and Z. Zhao (2007). "Evaluating Continuous Training Programs Using the Generalized Propensity Score." IZA Discussion Paper No. 3255.

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 Labour Market Policies in Europe*, Heidelberg: Physica-Verlag.

Lechner, M., R. Miquel and C. Wunsch (2007). "The Curse and Blessing of Training the Unemployed in a Changing Economy: The Case of East Germany after Unification." *German Economic Review* 8, 468-509.

Lechner, M., R. Miquel and C. Wunsch (2009). "Long-Run Effects of Public Sector Sponsored Training in West Germany." *Journal of the European Economic Association*, forthcoming.

Lechner, M. and C. Wunsch (2006). "Active Labour Market Policy in East Germany: Waiting for the Economy to Take Off." IZA Discussion Paper No. 2363.

OECD (2007). "Activating the Unemployed: What Countries do?" Employment Outlook, Chapter 5, Paris.

Osikominu, A. (2008). "Is Short Training Short-Lived and Long Training Long-Lasting? A Multi-State Duration Analysis of the Dynamic Effects of Training Schemes for the Unemployed." Discussion Paper, Albert-Ludwigs-University Freiburg.

Rinne, U., M. Schneider and A. Uhlendorff (2007). "Too Bad to Benefit? Effect Heterogeneity of Public Training Programs." IZA Discussion Paper No. 3240.

Rinne, U., A. Uhlendorff and Z. Zhao (2008). "Vouchers and Caseworkers in Public Training Programs: Evidence from the Hartz Reform in Germany." IZA Discussion Paper No. 3910.

Schneider, H. and A. Uhlendorff (2006). "Die Wirkung der Hartz-Reform im Bereich der beruflichen Weiterbildung." *Journal for Labor Market Research* 39, 477-490.

Schneider, H. , K. Brenke, B. Jesske, L. Kaiser, U. Rinne, M. Schneider, J. Steinwede and A. Uhlendorff (2006). "Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission - Modul 1b: Förderung beruflicher Weiterbildung und Transferleistungen." IZA Research Report No. 7, Bonn.

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

Wunsch, C. and M. Lechner (2008). "What Did All the Money Do? On the General Ineffectiveness of Recent West German Labour Market Programmes." *Kyklos*, 61, 134-174.

Chapter 1

Which Program for Whom? Evidence on the Comparative Effectiveness of Public Sponsored Training Programs in Germany

1.1 Introduction

Recent years have witnessed an enormously increased interest in the evaluation of active labor market policies, both in the US and Europe (for comprehensive overviews see Heckman, LaLonde and Smith (1999), Martin (2000), Martin and Grubb (2001), Kluve and Schmidt (2002) and Kluve (2006)). While, due to methodological and data limitations, earlier studies typically focussed on the evaluation of a single program, recent developments in evaluation methodology and data access have made it possible to gain deeper insights into the possibly very heterogenous effects of different types of programs and their comparative effectiveness. Prominent examples of recent evaluations involving multiple comparisons of different programs are Lechner (2002), Gerfin and Lechner (2002), Sianesi (2003), Hardoy (2005) and Dyke et al. (2006). This progress has been made possible by both methodological developments, in particular the extension of propensity score matching methods to the case of multiple treatments (Imbens (2000), Lechner (2001)), and the increasing availability of large, administrative data sets that provide the necessary sample sizes and program information to carry out in-depth evaluations of narrowly defined sub-programs. Given these new data sources, it is possible not only to evaluate the differential effects of the classical instruments of active labor market policy such as public employment services, job creation in the public sector, or public training programs, but also to evaluate different sub-programs within these categories, for example to study the comparative effectiveness of different forms of employment subsidies or different forms of public training programs.

This paper contributes to the growing evidence on the comparative effects of public sponsored training programs. We focus on the differential effects of public training programs in Germany. The case of Germany provides ideal conditions to study differential effects of public sponsored training for several reasons. First, the country has a long tradition of extensive active labor market programs covering all kinds of approaches.[1] As to public training programs, the Federal Employment Office of Germany has been offered a wide range of different programs ranging from very short measures aimed at minor skill adjustments and job search assistance to medium- and long-term programs with the explicit goal of increasing the human capital of the

[1]The total expenditure on active labor market policies was over 20 billion in 2004 (see Bundesagentur für Arbeit (2005a)). Programs include, among others, job search assistance, employment subsidies, job creation in the public sector, youth measures, measures to promote self-employment, and public training programs.

participants. In fact, the range of programs offered is much wider than in most other countries and the durations of typical programs vary between one or two weeks to several months or even several years. Another reason for using Germany is that the country has recently developed a growing awareness for the need to evaluate active labor market policies, which helped to open up existing administrative data bases to rigorous scientific research.[2] This has led to large, informative data sets merging different administrative sources. These data sets not only contain precise information on individual employment and transfer receipt histories but also comprehensive and detailed information on participation in all public sponsored measures of active labor market policy.[3] Large sample sizes make it possible to address aspects that have hitherto been difficult or impossible to address such as the heterogeneity of programs, the heterogeneity of effects across different groups of participants and the dynamic selection into different programs.

This paper provides a comprehensive and detailed econometric evaluation of public training programs conducted in Germany during the period February 2000 to January 2002. We distinguish different types of programs and consider effect heterogeneity with respect to population subgroups. Building on the work of Sianesi (2003, 2004) on dynamic treatments and on the work of Lechner (2001) on pairwise comparison of multiple treatments, we employ a stratified matching approach based on the propensity score, the elapsed duration of unemployment, and the calendar time. In order to take account of dynamic sorting processes, we stratify treatment effects by elapsed duration of unemployment. Our results show that average effects for too broad populations may hide statistically and economically significant treatment effects for individual subgroups and therefore help to understand why previous evaluation studies often yielded inconclusive results. While in many cases there is no discernible effect heterogeneity between subgroups, there is some evidence that the effects decline for older workers and for low–skilled workers. In these cases, the differences in treatment effects are very pronounced.

From an economic policy point of view, we address two important questions that have recently attracted considerable attention: whether relatively short training

[2] As a part of major labor market reforms, the so-called *Hartz-Reforms*, the need for rigorous scientific evaluation of program effectiveness was explicitly encoded into the law, see e.g. Jacobi and Kluve (2006).

[3] In fact, part of the project leading to this paper was the design and the validation of a merged administrative data base in cooperation with the Institut für Arbeitsmarkt- und Berufsforschung of the Federal Employment Office. This data base has subsequently been used for most of the policy evaluations in the context of the *Hartz-Reforms*.

measures can compete in effectiveness with more involved medium- to long-term measures, and the question whether practically oriented training programs have advantages over theoretically oriented class-room training. Our main motivation for the first question is that, traditionally, the focus of German public training programs was on medium- to long-term measures lasting several months to several years. Following criticism that such programs may not be effective as they 'lock-in' the participants for a long time, there has been a drastic shift towards short-term programs recently.[4] In terms of the number of participants, short-term training measures are by now the largest program of German active labor market policy. One of our aims is to evaluate whether or not this policy change can be justified *ex-post*. The specific form of short-term measures in Germany is also interesting from another point of view, as these measures often comprise elements of job search assistance, profiling or monitoring of the unemployed, apart from the provision of specific skills. By evaluating these kinds of programs we therefore also contribute to the literature that has focused on these specific forms of active labor market policy (see e.g. Martin (2000), Dolton and O'Neill (2002) and OECD (2005)).

The second question we address is also of considerable policy interest. It concerns the contents of training programs and focuses on the aspect of whether practically oriented training measures are better suited to provide unemployed workers with the skills and qualifications needed to improve labor market chances. Our results support hypotheses put forward in the literature (see e.g. Martin and Grubb (2001) and OECD (2005)) that practically oriented training may have advantages over pure classroom training. In this regard, our findings are in contrast to earlier findings for Germany during the 1990's, see Lechner et al. (2005a), Fitzenberger et al. (2006a), and Fitzenberger and Völter (2007).

A key advantage of our methodology is that it allows us to directly *compare* training programs, i.e. to ask the question of what would have happened if participants in short-term programs had participated in longer-term programs, or if participants in classroom training had taken part in more practically oriented training. This leads to more informative results than if one compares the effectiveness of different types of training when compared to not taking part in training at all. These results can directly be used for policy purposes, as they provide information on which programs are most advantageous for whom.

[4]See e.g. Bundesagentur für Arbeit (2005b), and figure 1.1 below.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. In section 3 we describe the main institutional features of the German system of public sponsored training. Section 4 presents details on the data used in this study. In section 5, we describe our econometric evaluation strategy. Section 6 discusses our empirical results, and section 7 concludes.

1.2 Literature Review

Although there exists a vast literature evaluating different aspects of active labor market policies in different countries (see the overview studies by Heckman, LaLonde and Smith (1999), Martin (2000), Martin and Grubb (2001), Kluve and Schmidt (2002) and Kluve (2006)) there are relatively few studies that focus on the comparative effects of different forms of training programs.

One of the first studies to consider differences in the outcomes of training programs was Gerfin and Lechner (2002). Using data for Switzerland, Gerfin and Lechner distinguished between five forms of public sponsored training programs with durations ranging between 5 and 13 weeks. Their results were negative in the sense that, one year after program start, the employment rate of participants was lower than that of comparable non-participants. However, longer, more involved training courses seemed to produce less negative results than shorter ones.

Most recent studies that focus on the differential effects of training programs use data for Germany. For example, Lechner et al. (2005a,b) evaluate the effects of a variety of training programs employed in East and West Germany in the 1990s. They distinguish between medium-term programs (mean duration 4 months), longer programs (mean duration 9 to 12 months) and long programs with specific contents such as retraining or training in a practice firm. Lechner et al. conclude that most of the programs had positive effects in the long run, even in East Germany. An important finding is that medium-term programs seem to outperform longer programs as they exhibit a much shorter lock-in period with otherwise similar employment effects after the end of the program. These findings are shared by Fitzenberger and Speckesser (2007), Fitzenberger et al. (2006a), and Fitzenberger and Völter (2007) who use the same data source but different econometric methods. Contrary to common hypotheses about the effectiveness of more practically oriented training programs (see e.g. Martin and Grubb (2001) or OECD (2005)), Lechner et al.

(2005a), Fitzenberger et al. (2006a), and Fitzenberger and Völter (2007) do not find that practical training as implemented in the 1990s dominates other kinds of training.

Using more recent and more informative data, Hujer et al. (2004) study the effectiveness of training programs in the early 2000s depending upon the duration of the programs. The study distinguishes programs of short (1-3 months), medium (6-12 months), and long (over 12 months) duration and estimates a multivariate mixed proportional hazard model. The results imply strong lock-in effects for the time the programs are attended but no significant effects on the exit rate from unemployment after completion of the program. Schneider et al. (2006) present policy evaluation results commissioned by the federal government in the context of the *Hartz-Reforms*. Although their focus is on the changes caused by these reforms, they also provide some results on the comparative effectiveness of a number of medium-term and long-term training programs. Their results also confirm the finding that shorter programs may be more effective than longer ones.

A drawback of all of these studies is that they omit the by now most important type of public sponsored training in Germany, so-called short-term training ('Trainingsmaßnahmen') - this program is not to be confused with short further training programs as analyzed by Hujer et al. (2004) or Lechner et al. (2005a,b). Short-term training courses typically last only 2 to 12 weeks and often combine elements of job search assistance with the provision of specific skills (see more detailed description below). In light of the policy debate (Martin and Grubb (2001) or OECD (2005)), short-term training seems attractive since it may serve the purpose of activating the unemployed without locking them in lengthy training programs. Furthermore, a number of recent contributions from the evaluation literature suggest that increased job search assistance may be an inexpensive way to help unemployed individuals back into employment (see e.g. Blundell et al. (2004), Weber and Hofer (2004), Fougère et al. (2005), Hujer et al. (2005), Crépon et al. (2005), and Van den Berg and Van der Klaauw (2006)).

The only other two studies we are aware of that consider short-term training in Germany are Hujer et al. (2006), and Lechner and Wunsch (2006). Hujer et al. examine whether participation in short-term training measures reduces the unemployment duration of West German job-seekers. They do not compare short-term training to other measures of active labor market policy. Lechner and Wunsch (2006) evaluate a large number of different training and non-training measures in East Germany, among them short-term training. Their results suggest no or even negative effects for all programs considered. Lechner and Wunsch explain their finding by the difficult situation in the East German labor market.

1.3 Training as Part of Active Labor Market Policy

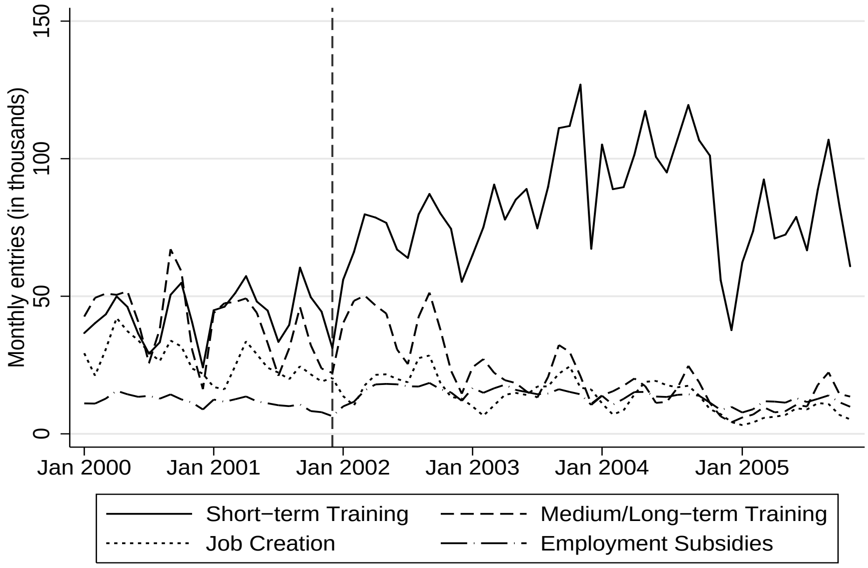
The main goal of German active labor market policy is to permanently reintegrate unemployed individuals (and individuals who are at risk of becoming unemployed) back into employment. The policy instruments cover a wide range of different measures such as employment subsidies, job creation in the public sector, measures directed at youth unemployment, measures to promote self-employment, and public training programs. For an overview over the different kinds of policies and their quantitative importance, see figure 1.1 and table 1.2 in the appendix.[5]

As shown in figure 1.1, public training programs have traditionally been the most important part of German active labor market policy. There are three main categories of training programs: *short-term training* ('Trainingsmaßnahmen'), *further training* ('Berufliche Weiterbildung'), and *retraining* ('Umschulung').[6] Apart from the fact that all three types of training require full-time participation, they differ considerably in length and contents. Recently, *short-term training* has become the largest training program regarding the number of participants - for the following, see Kurtz (2003). Short-term training measures last only two to twelve weeks (the mean duration is slightly over four weeks, see table 1.1) and typically pursue one or several of the following three aims. A first potential aim is aptitude and qualification testing, i.e. the program is used to assess job seekers' labor market opportunities and their suitability for different jobs. This may also entail profiling activities on the side of the Federal Employment Office and preparation of more detailed work plans

[5]This paper focuses on public training programs attended in the period 2000 to 2002. The following paragraphs describe the relevant institutional settings up to the end of 2002, before the *Hartz-Reforms* were enacted. The reforms also changed some of the rules on public training programs. These changes are not relevant to our study but they will be important for future evaluations (see e.g. Biewen and Fitzenberger (2004) or Schneider et al. (2006)).

[6]In addition, there are specific training schemes for youth unemployed and disabled persons, as well as German language courses for asylum seekers and ethnic Germans returning from former German settlements in Eastern Europe. These training measures are not considered here.

Figure 1.1: Active Labor Market Policies in Germany



Source: Statistics of the Federal Employment Office of Germany.

to reintegrate the job seeker into the labor market. A second aim is to test the job seeker's willingness to work and to improve job search skills. This may be achieved through activities such as job-application training, simulation of job interviews or general counseling on job search methods. The third and final aim of short-term training measures is the provision of specific skills that are necessary to improve the job seeker's labor market prospects. Typical examples for this type of measures are computer courses or courses providing commercial training. In 2001, 22 percent of short-term training measures belonged to the first type, 19 percent to the second type, and some 28 percent to the third type. About 31 percent were combinations of the different types. In most cases, these were combinations of job search assistance and the provision of specific skills, or aptitude testing and the provision of specific skills (Kurtz, 2003, tables A3 and A6).

In comparison to short-term training, the more substantial *further training* programs typically take much longer and are more involved. With durations ranging between several months and one year, further training measures can be classified as medium-term programs. Their aim is to maintain, update, adjust, and extend professional skills and qualifications. Further training programs cover a wide range of courses in a variety of fields and may also comprise practical elements such as on-the-job training, internships or working in practice firms. In our evaluation we will distinguish between practically-oriented further training programs (which are typically of shorter duration) and pure class-room training. Apart from short-term and further training, employment offices also offer *retraining*. Retraining programs last two to three years and typically lead to a new vocational education degree within the German apprenticeship system. Retraining may involve vocational training in a profession that was not the original profession of the job seeker. In addition, retraining may be granted to job seekers who face difficult labor market prospects because they lack a vocational degree in the first place. In general, retraining programs are similar to regular apprenticeships and typically combine class-room training with on-the-job training.

To become eligible for participation in one of the training programs, job seekers have to register personally at the local labor office. This involves a counseling interview with the caseworker. Besides being registered as unemployed or as a job seeker at risk of becoming unemployed, candidates for short-term training do not have to fulfil any additional eligibility criteria. In the case of medium- and long-term training, individuals are typically eligible only if they also fulfil a minimum work requirement of one year and if they are entitled to unemployment compensation. However, there are several exceptions to these requirements. The really binding criterium is that the training scheme has to be considered necessary in order for the job seeker to find a new job. This is, for example, the case if the employment chances in the target occupation of a job seeker are good but require an additional adjustment of skills. Training measures are usually assigned by the caseworker. Depending on regional and local circumstances, caseworkers may exercise a great deal of discretion when allocating the different programs. Suitable programs are chosen from a pool of certified public or private institutions or firms.

If a person is admitted to one of the training measures, the employment office pays all direct training costs. In addition, the participants of short-term training may continue to receive unemployment benefits or means-tested unemployment assistance, if they are eligible for such transfer payments. Participants of short-term training are still registered unemployed during the program. In contrast, participants of further training or retraining do not remain registered unemployed during the program. Participants of further training and retraining usually also receive a subsistence allowance provided they fulfill a minimum work requirement of twelve months within the last three years. This subsistence allowance is usually of the same amount as unemployment benefits or unemployment assistance. Overall, there are no significant financial incentives for unemployed individuals to participate in a training program, in contrast to the situation in Germany before 1998, see Fitzenberger et al. (2006a).

Table 1.1: Average Expenditures per Participant in Short-term, Further and Retraining in Germany from 2000-2003

	2000		2001		2002		2003	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Short-term training	580	1,2	570	1,1	658	0,9	538	1
Further/retraining	1627	8,2	1668	9,3	1686	9,1	1555	10,5
– subsistence allowance	1152		1178		1188		1156	
– training costs	640		664		681		631	

Note: Columns labeled with a (1) contain the average monthly expenditures (in Euro) per participant, columns labeled with a (2) display the average duration of the program in months. Source: Bundesagentur für Arbeit, Daten zu den Eingliederungsbilanzen 2001,2002a-2004a.

Table 1.1 shows that the average monthly training costs per participant are lower for short-term training courses (about 570 Euros in 2001) than for the longer-term measures (664 Euros). Given that the average length of short-term measures is only 1.1 months while that of longer-term measures is some 9.3 months, this results in training costs for short-term measures (627 Euros) that amount to only about one tenth of those for medium- and long-term measures (6175 Euros).[7] Since 2002, in light of huge differences in costs, the Federal Employment Office has been drastically increasing the share of short-term training measures at the expense of longer-term measures (see figure 1.1). Of course, the higher training costs may be justified if the medium- to long-term measures lead to correspondingly higher gains in employment probabilities. This is one of the main questions motivating our evaluation.

[7]In addition to the direct costs, participants in longer-term training schemes usually receive the subsistence allowance. However, the subsistence payments simply replace the ordinary unemployment compensation the participants would have otherwise received.

1.4 Data

1.4.1 Integrated Biographies Sample

Our study uses a new and exceptionally rich administrative data base, the so-called Integrated Biographies Sample (IEBS). This data base has only recently been made available by the Federal Employment Office of Germany.[8] The IEBS is a merged 2.2% random sample of individual data drawn from the universe of data records collected through four different administrative processes. Our version of the IEBS has been supplemented with additional information which is not publicly available (especially information on health). The IEBS contains detailed daily information on employment subject to social security contributions, receipt of transfer payments during unemployment, job search, and participation in different programs of active labor market policy. In addition, the IEBS comprises a large variety of covariates including socio-economic characteristics (information on family, health and educational qualifications), occupational and job characteristics, extensive firm and sectoral information, as well as details on individual job search histories and assessments of case workers.[9] For evaluation purposes, a rich set of covariates is essential as it can be used to reconstruct the circumstances that did or did not lead to the participation in a particular program thus making it possible to control for the selection of individuals into programs.

We give a brief description of the IEBS in order to underscore its value for evaluation purposes. The IEBS is based on four different administrative sources the so-called Employment History ('Beschäftigten-Historik'), the Benefit Recipient History ('Leistungsempfänger-Historik'), the Supply of Applicants ('Bewerberangebot'), and the Data Base of Program Participants ('Massnahme-Teilnehmer-Gesamtdatenbank').

The *Employment History* involves register data comprising employment information for all employees subject to contributions to the public social security system. It covers the time period 1990 to 2004. The main feature of this data is detailed daily information on the employment status of each recorded individual. We use this information to account for the labor market history of individuals as well as to

[8]For more information on the IEBS, see Osikominu (2005, section 3) and Hummel et al. (2005).

[9]The IEBS lacks direct information on cognitive and non-cognitive skills of the unemployed as well as information on caseworkers and instructors. There is no German data set available which is suitable for the evaluation of training programs and which includes this information.

measure employment outcomes. For each employment spell, in addition to start and end dates, data from the Employment History contains information on personal as well as job and firm characteristics such as wage, industry, or occupation.

The second data source, the *Benefit Recipient History*, includes daily spells of all unemployment benefit, unemployment assistance and subsistence allowance payments between January 1990 and June 2005. It also contains information on personal characteristics. The Benefit Recipient History is important as it provides information on the periods in which individuals were out of employment and therefore not covered by the Employment History. In particular, the Benefit Recipient History includes information about the exact start and end dates of periods of transfer receipt. We expect this information to be very reliable since it is, at the administrative level, directly linked to flows of benefit payments. The Information on benefit payments allow us to construct individual benefit histories dating back several years. Moreover, we use additional information contained in the Benefit Recipients History involving sanctions and periods of disqualification from benefit receipt that may serve as indicators for a lack of motivation.

The third administrative data source of the IEBS is the so-called *Supply of Applicants*, which contains data on individuals searching for jobs. The Supply of Applicants data cover the period January 1997 to June 2005. In our study they are used in two ways. First, they provide additional information about the labor market status of a person, in particular whether the person in question searches for a job but is not (yet) registered as unemployed or whether he or she is sick while registered unemployed. Second, the job search episodes include additional information about personal characteristics, in particular about educational qualifications, nationality, and marital status. They also provide information about whether the applicant wishes to change occupations, about health problems that might influence employment chances, and about the labor market prospects of the applicants as assessed by the case worker. Finally, the data on applicants include regional and local identifiers, which we use to link regional and local information, for example unemployment rates at the district level.

The fourth data source in the IEBS is the *Data Base of Program Participants*, which is particularly important for evaluation purposes. This data base contains detailed information on participation in public sector sponsored labor market programs covering the period January 2000 to July 2005. Similar to the other sources, information comes in the form of spells indicating the start and end dates at the daily level, the type of the program as well as additional information on the program such as the planned end date, whether the participant entered the program with a delay, and whether the program was successfully completed. The Data Base of Program Participants not only contains information on the set of training measures evaluated in this paper, but also on other programs such as employment subsidies. This is important, as it enables us to distinguish between regular and subsidized employment when evaluating employment outcomes.[10]

Being among the first to use the IEBS, we were involved in comprehensive data checks.[11] We ran extensive consistency checks of the records coming from the different sources, making use of additional information on the data generating process provided to us by the Institute for Employment Research.[12] Our conclusion is that on the one hand the employment and benefit data are highly reliable concerning employment status, wage and transfer payments, and the start and end dates of spells. The likely reason for this is that contribution rates and benefit entitlements are directly based on this information. On the other hand, information not needed for these administrative purposes can be less reliable. For example, in the employment data base the educational variable appears to be affected by non-negligible measurement error as it is not directly relevant for social security entitlements (see Fitzenberger et al. (2006b) for imputation methods to correct the education variable). Personal characteristics exhibit a higher degree of reliability in the program participation and job seeker data, because they are relevant for the purpose of assigning job offers or programs to the unemployed. In our evaluation, we exploited the available information as efficiently as possible by choosing the data source that is most reliable for a given purpose.

Although the data in the IEBS generally seem very reliable, there is some need for data corrections. In particular, we corrected in some cases the end dates of program spells if there was evidence that the end dates recorded in the data base of program

[10]A disadvantage of the data covering labor market training in German in the 1990s used in studies such as Fitzenberger et al. (2006a), Fitzenberger and Speckesser (2007), Fitzenberger and Völter (2007), and Lechner et al. (2005a,b) is that it is not possible to distinguish whether participants found employment in the regular labor market or whether they took part in job creation measures. Note that for the time period from the year 2000 onwards, Lechner et al. (2005a,b) use the information based on the IEBS whether an individual is employed in a subsidized job.

[11]Given the non-trivial task of merging four large scale administrative data sources of very different designs such checks were indispensable.

[12]This work is documented in Bender et al. (2004, 2005).

participation was wrong. For details on measurement error in program end dates in the IEBS and correction procedures, see Waller (2007).

1.4.2 Evaluation Sample and Training Programs

We follow an evaluation strategy (see below) that is based on comparisons with (multiple) control groups. A common feature of control group approaches is that they partition the group of potential participants into a group of participants and a group of non-participants. As a consequence, the first question that has to be answered when selecting the evaluation sample is that of who is a potential program participant.

For several reasons, we decide to focus on individuals who become unemployed after having been continuously employed for at least three months, instead of individuals who are *observed* unemployed at a given point of time. This is to avoid the case of individuals registering as unemployed from being out of labor force because they want to participate in a training program. In interviews, case workers told us that especially women returning from maternity leave, divorcees, or university graduates who have difficulty finding a job may contact the local employment office inquiring about the possibility of participating in public training programs. However, these individuals often only register as unemployed if the chances of actual participation are high enough. An evaluation sample based on observed unemployment status (instead of an inflow sample into unemployment) would therefore suffer from the problem of an incompletely observed control group, because it would be difficult to find comparable non-participants for those individuals who endogenously register as unemployed (due to their non-registering as unemployed, non-participating counterparts would not appear in the sample). Analyzing an inflow sample into unemployment, we focus on individuals who have been attached to the labor market, which helps to construct the control group based on the labor market relevant information in the data. Furthermore, the beginning of unemployment defines a natural time scale to align treated and nontreated individuals.

In the following, we focus on an inflow sample into unemployment consisting of individuals who became unemployed between the beginning of February 2000 and the end of January 2002, after having been continuously employed for at least three months. Entering unemployment is defined as quitting regular (not marginal), non-subsidized employment and subsequently being in contact with the employment office (not necessarily immediately), either through benefit receipt, program participation or a job search spell.[13] In order to exclude individuals eligible for specific labor market programs for the youth and individuals eligible for early retirement schemes, we only consider persons aged between 25 and 53 years at the beginning of their unemployment spell. Our evaluation focusses on the first training program that is attended in the course of an unemployment spell.

Based on the description of program types in section 1.3, we analyze four different types of training, which closely follows the legal grouping of program types:

- short-term training (STT),
- classroom further training (CFT),
- practical further training (PFT), and
- retraining (RT).

In some cases, we grouped programs whose planned duration and contents did not really fit into the category defined by the law into the category that was most appropriate from an economic point of view. According to the same criteria, we also grouped measures of 'discretionary support' (Freie Förderung) and measures financed through the European Social Fund (Europäischer Sozialfond, ESF) into one of the four program categories. We carry out our evaluations for men and women, for East and West Germany, and (for reasons explained in the next section) for different durations of elapsed unemployment separately. This results in a total number of twelve evaluation samples, the sample sizes of which are shown in table 1.3 in the appendix. Table 1.6 in the appendix and figure 1.16 provide descriptive information on the duration of different program types. STT is the shortest program and RT the longest program. Durations for CFT are fairly uniformly distributed between 1 and 12 months with a strong spike at 12 months. PFT is shorter than CFT and shows a strong spike at 7 months.

[13]Note that this implies that the same individual may appear more than once in our evaluation sample. About ten percent of the individuals in our sample are represented by more than one unemployment spell according to the above definition. We take account of multiple inclusion of the same individual in the sample when calculating standard errors, see section 1.5.

1.5 Econometric Implementation

Our goal is to analyze the effect of the $K = 4$ different training programs (STT, CFT, PFT, RT) on monthly employment at the individual level. In a situation where individuals have multiple treatment options, we estimate the average treatment effect on the treated (ATT) of one training program against nonparticipation in any of the three programs and of pairwise comparisons of two programs. Extending the static multiple treatment approach to a dynamic setting, we follow Sianesi (2003, 2004) and apply the standard static treatment approach recursively depending on the elapsed unemployment duration. The implementation builds upon the approach for binary treatment in Fitzenberger and Speckesser (2007) and for multiple treatments in Fitzenberger et al. (2006a). In contrast to these earlier papers, we also analyze the heterogeneity of the estimated ATT by various socio-economic characteristics of the treated individuals.

1.5.1 Multiple Treatments in a Dynamic Context

Our empirical analysis is based upon the potential-outcome-approach to causality, see Roy (1951), Rubin (1974), and the survey of Heckman et al. (1999). Lechner (2001) and Imbens (2000) extend this framework to allow for multiple, exclusive treatments. Let the potential outcome $Y^k, k = 1, ..., 4$, represent the outcome associated with training program k and Y^0 is the outcome when participating in none of the 4 training programs. For each individual, only one of the $K+1$ potential outcomes is observed and the remaining K outcomes are counterfactual. We estimate the average treatment effect on the treated (ATT) of participating in treatment $k = 1, 2, 3, 4$ against nonparticipation $k = 0$ (treatment versus waiting) and the differential effects of the programs (program k versus program l where $k, l \neq 0$), see Lechner (2001).

Fredriksson and Johansson (2003, 2004) argue that a static evaluation analysis, which assigns unemployed individuals to a treatment group and a nontreatment group based on the treatment information observed in the data, yields biased treatment effects. This is because the definition of the control group conditions on future outcomes or future treatment. For Sweden, Sianesi (2004) argues that all unemployed individuals are potential future participants in active labor market programs, a view which is particularly plausible for countries with comprehensive systems of active labor market policies (like Germany). This discussion implies that a purely static evaluation of the different training programs is not warranted. Following Sianesi (2003, 2004), we analyze the effects of the first participation in a training program during the unemployment spell considered *conditional on the starting date of the treatment*. We distinguish between treatment starting during months 0 to 3 of the unemployment spell (stratum 1), treatment starting during months 4 to 6 (stratum 2), and treatment starting during months 7 to 12 (stratum 3).

We analyze treatment conditional upon the unemployment spell lasting at least until the start of the treatment k and this being the first treatment during the unemployment spell considered. Therefore, the ATT parameter (comparing treatments k and l) of interest is

$$\theta(k,l;u,\tau) = E(Y^k(u,\tau)|T_u = k, U \geq u-1, T_1 = ... = T_{u-1} = 0) \tag{1.1}$$

$$-E(Y^l(\tilde{u},\tau-(\tilde{u}-u))|T_u = k, u \leq \tilde{u} \leq \bar{u}, U \geq u-1, T_1 = ... = T_{u-1} = 0) \ ,$$

where T_u is the treatment variable for treatment starting in month u of unemployment. $Y^k(u,\tau)$, $Y^l(u,\tau)$ are the potential treatment outcomes for treatments k and l, respectively, in periods $u+\tau$, where treatment starts in period u and $\tau = 0, 1, 2, ...$, counts the months since the beginning of treatment. When $l = 0$, we compare treatment k versus waiting (nonparticipation in the stratum) and when $l \geq 1$, we do a pairwise comparison between treatment k and l. U is the duration of unemployment, $\tilde{u}$ is the random month when alternative treatment l starts, and $\bar{u} = 2, 4, 8$ is the last month in the stratum of elapsed unemployment considered. Then, $\tau - (\tilde{u} - u)$ counts the months since start of treatment l yielding alignment of unemployment experience, because $u + \tau = \tilde{u} + (\tau - (\tilde{u} - u))$, and $Y^l(\tilde{u}, \tau - (\tilde{u} - u))$ is the outcome of individuals who receive treatment l between period u and $\bar{u}$. For starts of l later than u, we have $\tilde{u} - u > 0$ and therefore, before l starts, $\tau - (\tilde{u} - u) < 0$. Then, these individuals are still unemployed, i.e. $Y^l(\tilde{u}, \tau - (\tilde{u} - u)) = 0$ when the second argument of $Y^l(.,.)$ is negative. This way, we account for the fact that alternative treatments, for which the individual receiving treatment k in period u is eligible, might not start in the same month u. The treatment parameter we actually estimate is the average within a stratum

$$\theta(k,l;\tau) = \sum_u g_u \theta(k,l;u,\tau) \ ,$$

with respect to the distribution g_u of starting dates u within the stratum.

Our estimated treatment parameter (1.1) mirrors the decision problem of the case worker and the unemployed who recurrently during the unemployment spell decide whether to start any of the programs now or to postpone participation to the future.

We evaluate the differential effects of multiple treatments assuming the following dynamic version of the conditional mean independence assumption (DCIA)

$$(1.2) \quad E(Y^l(\tilde{u}, \tau - (\tilde{u} - u))|T_u = k, u \leq \tilde{u} \leq \bar{u}, U \geq u-1, T_1 = ... = T_{u-1} = 0, X)$$

$$= E(Y^l(\tilde{u}, \tau - (\tilde{u} - u))|T_{\tilde{u}} = l, u \leq \tilde{u} \leq \bar{u}, U \geq u-1, T_1 = ... = T_{u-1} = 0, X) \ ,$$

where X are time-varying as well as time-invariant (during the unemployment spell) characteristics, $T_{\tilde{u}} = l$ indicates treatment l between u and $\bar{u}$ ($\bar{u}$ is the end of the stratum of elapsed unemployment considered), and $\tau \geq 0$, see equation (1.1) above and the analogous discussion in Sianesi (2004, p. 137). We effectively assume that conditional on X, conditional on being unemployed at least until period $u-1$, and conditional on not receiving any treatment before u (both referring to treatment in period u) individuals are comparable in their outcome for treatment l occurring between u and $\bar{u}$.

Building on Rosenbaum and Rubin's (1983) result on the balancing property of the propensity score in the case of a binary treatment, Lechner (2001) shows that the conditional probability of treatment k, given that the individual receives treatment k or treatment l, $P^{k|kl}(X)$, exhibits an analogous balancing property for the pairwise estimation of the ATT's of program k versus l. This allows to apply standard binary propensity score matching based on the sample of individuals participating in either program k or in program l. For this subsample, we simply estimate the probability of treatment k and then apply a bivariate extension of standard propensity matching techniques. Implicitly, we assume that the actual beginning of treatment within a stratum is random conditional on X.

To account for the dynamic treatment assignment, we estimate the probability of treatment k given that unemployment lasts long enough to make an individual 'eligible'. For treatment during months 0 to 3, we take the total sample of unemployed, who participate in k or l during months 0 to 3 (stratum 1), and estimate a Probit model for participation in k. For $l = 0$, the group of nonparticipants in k includes those unemployed who either never participate in any program or who start some treatment after month 4. For treatment during strata 2 and 3, the basic sample consists of those unemployed who are still unemployed in the first month of the stratum.

We implement a stratified local linear matching approach by imposing that the matching partners for an individual receiving treatment k are still unemployed in the month before treatment k starts, i.e. we exactly align treated and nontreated individuals by elapsed unemployment duration in months. For the comparison of training against waiting, we align treated and controls in addition by the elapsed duration of unemployment benefit receipt in months. The expected counterfactual employment outcome for nonparticipation is obtained by means of a bivariate local linear regression on the propensity score and the starting month of the unemployment spell. We use a product kernel in the estimated propensity score and the calendar month of entry into unemployment

$$KK(p,c) = K\left(\frac{p-p_j}{h_p}\right) \cdot h_c^{|c-c_j|}, \tag{1.3}$$

where $K(z)$ is the Gaussian kernel function, p and c are the propensity score and the calendar month of entry into unemployment of a particular treated individual, p_j and c_j are the estimated propensity score and the calendar month of entry into unemployment of an individual j belonging to the comparison group of individuals treated with l. h_p and h_c are the bandwidths. Taken together, we impose three matching requirements: i) similarity of the pairwise propensity score, ii) *exact* match of the elapsed unemployment (and benefit receipt) duration, and iii) similarity of beginning of unemployment.

We use a bivariate crossvalidation procedure to obtain the bandwidths h_p and h_c by minimizing the squared prediction error for the average of the l-outcome for the nearest neighbors of the participants in program k.[14] An estimate for the variance of the estimated treatment effects is obtained through bootstrapping based on 250 resamples. This way, we take account of the sampling variability in the estimated propensity score.

As a balancing test, we use the regression test suggested in Smith and Todd (2005) to investigate whether the covariates are balanced sufficiently by matching on the estimated propensity score using a flexible polynomial approximation. Furthermore, we investigate whether treated and matched nontreated individuals differ significantly in

[14] This method is also used in Fitzenberger et al. (2006a) and it is an extension of the crossvalidation procedure suggested in Bergemann et al. (2004).

their outcomes before the beginning of unemployment, in addition to those variables already used as arguments of the propensity score. We estimate these differences in the same way as the treatment effects after the beginning of the program. By construction, treated individuals and their matched counterparts exhibit the same unemployment duration until the beginning of treatment.

1.5.2 Specification of the Propensity Scores

First, we need to discuss the plausibility of the DCIA (1.2) for our application. For propensity score matching to be a valid procedure one needs to control for the variables that jointly influence participation and outcomes such that, when conditioning on these variables, potential outcomes are mean independent of treatment status. It is therefore essential to base the estimation of the propensity scores on all relevant information. Given our data base, we are in the lucky position to construct a large set of time-constant as well as time-varying (within the unemployment spell) variables to model the selection into the different training programs.

As Sianesi (2004), we argue that the participation probability depends upon the variables determining re-employment prospects once unemployment began. Consequently, all individuals are considered who have left employment in the same two years (matching controls for beginning of unemployment) and who have experienced the same unemployment duration before program participation. Furthermore, observable individual characteristics and information from the previous employment and benefit history have been included in the propensity score estimation. E.g., we consider skill information, regional information, occupational status, and industry which should be crucial for re-employment chances. In addition, we use subjective assessments of the unemployed by case workers, which should proxy for further relevant unobserved characteristics. In addition to matching on the begínning of unemployment and the elapsed duration of unemployment, we argue that the variables used in the estimation of the propensity score are rich enough to control for the selection into treatment. This is particularly plausible because participation occurred at a fairly large scale, assignment was not very targeted and driven by the supply of programs, and case workers had little guidance on 'what works for whom'. Supporting our point of view, Schneider et al. (2006) argue that until 2002 assignment to training was strongly driven by the supply of available courses.

Concretely, we use the following variables and their interactions for the specification of the propensity score.[15]

Personal characteristics

As personal characteristics, we consider age, disability status, schooling and professional qualification, family status, whether there are children, whether there are children under 10 years, nationality other than German, and whether the person in question is an ethnic German who has migrated back into Germany (usually from Eastern European countries).

Labor market and benefit histories

We use information on occupation and industry of the last job before unemployment, whether this last job was less than full-time, whether it was a white-collar or blue-collar position, the reason why this last job was ended, the quarter of the beginning of the unemployment period, whether there were any periods of incapacity in the last three years, the total length of employment (all durations are measured in days) during the last three years, the duration of transfer payments during the last three years (i.e. unemployment benefits, unemployment assistance, subsistence allowance), times without any information in the data set, times of contact with the employment office during the last three years before unemployment, whether the person was employed 6, 12, 24 months before the beginning of the unemployment period, log daily wage in the last job before unemployment, an indicator whether this wage was censored, the log average wage in the year before unemployment and censoring dummies related to this variable.

Case worker reported assessments

As to the assessment of the case workers with regards to the motivation, plans and labor market prospects of the unemployed, we consider current health status, past health problems, information on whether a program was canceled within the last three years, penalties and disqualification from benefits within the last three years, participation in a program with a social work component, indication of lack of motivation within the last three years, the number of job proposals made the unemployed received from the employment office, and information on the desired

[15]See the appendix for summary statistics and a more detailed description of the variables used. Time-varying covariates are updated at the beginning of each stratum. For time-varying variables, information from spells starting more than a few days later than the beginning of the respective time window is not used in order to avoid endogeneity problems.

job.

Regional information

We use different unemployment rates in the home district of an individual, the districts which share the labor market classification of the region, the federal state, and all of Germany.

Using these variables as possible regressors, we fit the propensity scores separately for each of the twelve evaluation samples (men/women, East/West, stratum 1/2/3), and each treatment comparison pair. In each case, we run an extensive specification search. The final specification is chosen based on economic considerations, statistical significance of the variables included, and the balancing tests described above.[16] The final specification typically includes 20 to 35 covariates.

1.5.3 Estimating Effect Heterogeneity

The estimation of the ATT provides a semiparametric, aggregate impact measure for a possibly heterogeneous treatment group. However, it is conceivable that a zero average hides positive and negative treatment effects for different subgroups of the treated. Although, one could estimate the ATT for each subgroup of interest[17] using the dynamic matching estimator as described in section 1.5.1, such a strategy is limited by the curse-of-dimensionality. The sample sizes of the subgroups need to be sufficiently large to do so with reasonable precision.

As a simple alternative, we propose to run linear regressions of the estimated individual cumulated treatment effects in the matched samples on covariates which could cause effect heterogeneity. Such regressions after matching are used in the literature to adjust for possible remaining mismatch between treated individuals and matched controls (see e.g. Lechner (1999)). However, we are not aware of any recent study in the evaluation of active labor market policy that uses regressions after matching to study effect heterogeneity. We focus on the cumulated treatment effect after the end of the lock-in period.

To be specific, we run the following regressions in the matched sample for the cu-

[16] Estimation results are available from the authors upon request.

[17] E.g. Lechner et al. (2005a) do so for a small number of subsets of the treatment sample to investigate effect heterogeneity.

mulated treatment effect over the months $\tau = T_1, ..., T_2$

$$\sum_{\tau=T_1}^{T_2} (Y_i^k(\tau) - \hat{Y}_i^l(\tau)) = \alpha + x_i\beta + (x_i - \hat{x}_i)\gamma + u_i\,, \tag{1.4}$$

where individual i with covariates x_i receives treatment k with observed outcome Y_i^k, $\hat{Y}_i^l$ and $\hat{x}_i$ are the predicted, counterfactual l-outcome and the predicted covariates based on the local linear regression in the matching procedure, and $x_i - \hat{x}_i$ represents the mismatch between the average covariates of the matched comparison individuals and the covariates of indvidual i. We estimate the cumulated effect over the time interval $[T_1, T_2]$ after the beginning of treatment. T_1 and T_2 are chosen specifically for the program and the stratum under consideration. T_1 is a proxy for the end of the lock-in period and T_2 is the last post treatment month observed. As a benchmark, if $\beta = \gamma = 0$, then the estimated α corresponds to the estimated ATT for the entire treatment sample and there is no systematic effect heterogeneity by the level of covariates.

We test systematically for significant effect heterogeneity by covariates. The standard errors of the estimated regression coefficients are obtained through the bootstrap procedure for the matching estimator by rerunning the regression (1.4) for all resamples. In some cases, the regression (1.4) in the matched sample suffers from multicollinearity problems due to the mismatch terms being highly correlated with the covariates. In cases of strong multicollinearity, we exclude the mismatch from the regression. As final results we only report those specification of the effect heterogeneity regressions with significant covariates.

1.6 Empirical Results

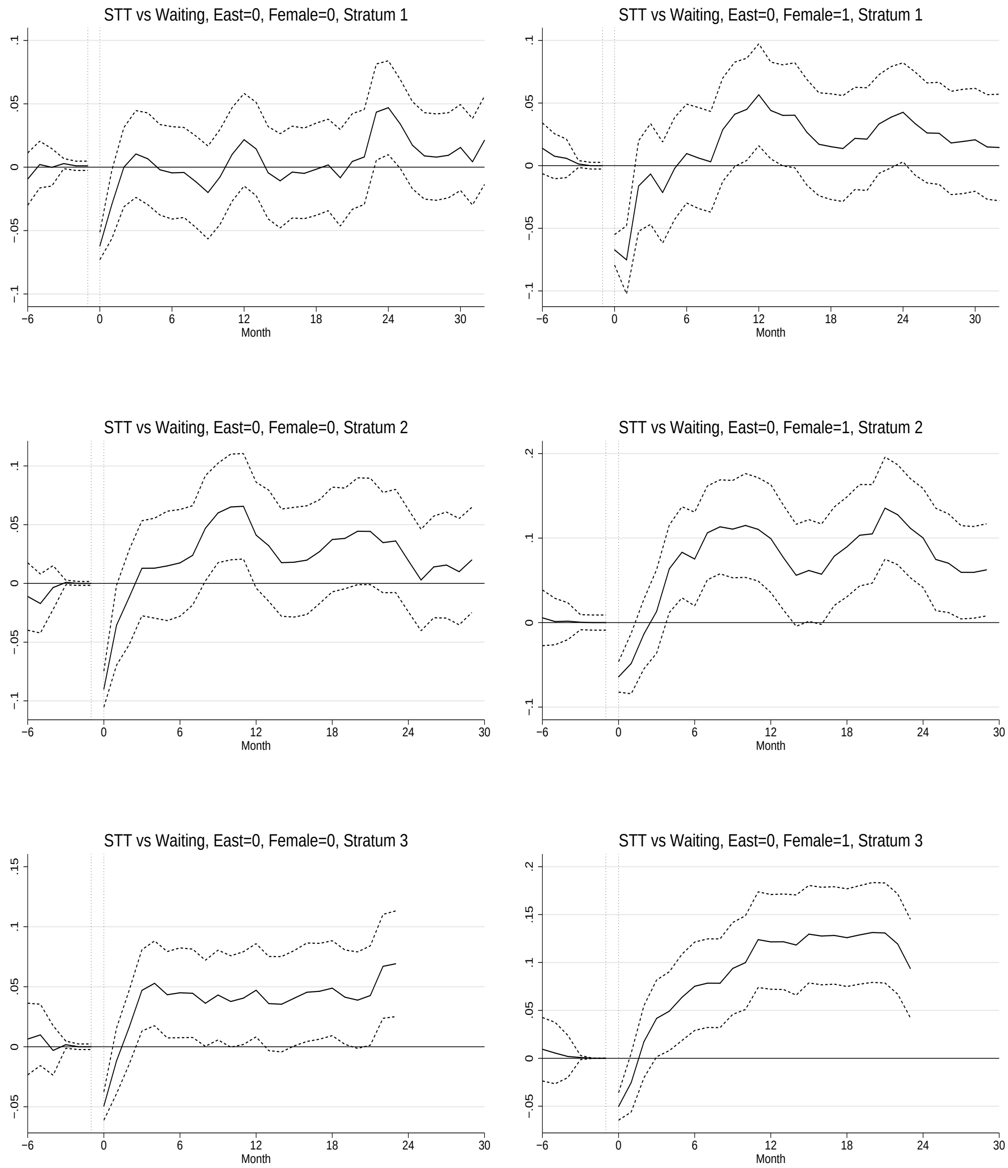
1.6.1 Training vs. 'Waiting'

The evaluation results for training vs. not participating in any measure of active labor market policy ('waiting') are shown in figures 1.2 to 1.7. Each graph displays the average treatment effect on the treated (ATT), i.e. the difference between the actual and the counterfactual employment outcome averaged over those individuals who participate in the program under consideration. More precisely, we compare the actual employment outcome of the treated to the employment outcome these individuals would have had, *had they not taken part in any other program in the respective time window of their unemployment spell.* As already mentioned, we distinguish between programs starting in three different time windows (strata) of elapsed unemployment: 0 to 3 months (stratum 1), 4 to 6 months (stratum 2), and 7 to 12 months (stratum 3). Due to the smaller number of treated individuals, we only consider one time window ranging from month 0 to 12 for participants in practical further training (PFT) and one ranging from month 0 to 3 for participants in retraining (RT).

We evaluate treatment effects at different points in time. On the time axis in our graphs, positive values denote months since the program start, while negative values represent pre-unemployment months. We omit the period between the start of unemployment and the start of the program where both control and treatment group are unemployed. The dashed lines around the estimated ATT are bootstrapped 95 percent confidence bands. Treatment effects for a particular month are statistically significant if zero is not contained in the confidence band.

Figure 1.2 shows estimated treatment effects for short-term training programs (STT) in West Germany. The results for men are given in the left column, while those for women are shown in the right column. The figures suggest short and not very pronounced lock-in effects of short-term training programs of minus five percentage points (i.e. during the program, participants had a five percentage points lower monthly employment probability than they would have had if they had not participated in the program). These lock-in effects do not last more than two or three months, which is not surprising given the average length of such programs. After the short lock-in period, the difference between actual and counterfactual employment outcomes of participants turn positive. However, results seem to depend strongly on elapsed unemployment duration. While there is no evidence for statistically significant treatment effects for individuals participating in the first three months of their unemployment spell (stratum 1), treatment effects for men starting a short-term training program in months 7 to 12 (stratum 3) of their unemployment spell, and women starting one in month 4 or later are positive and statistically significant (except for men in stratum 2). According to these estimates, the monthly employment probability of West German men participating in short-term training is increased by about 5 percentage points. At some 10 percentage points, this effect is larger for women.

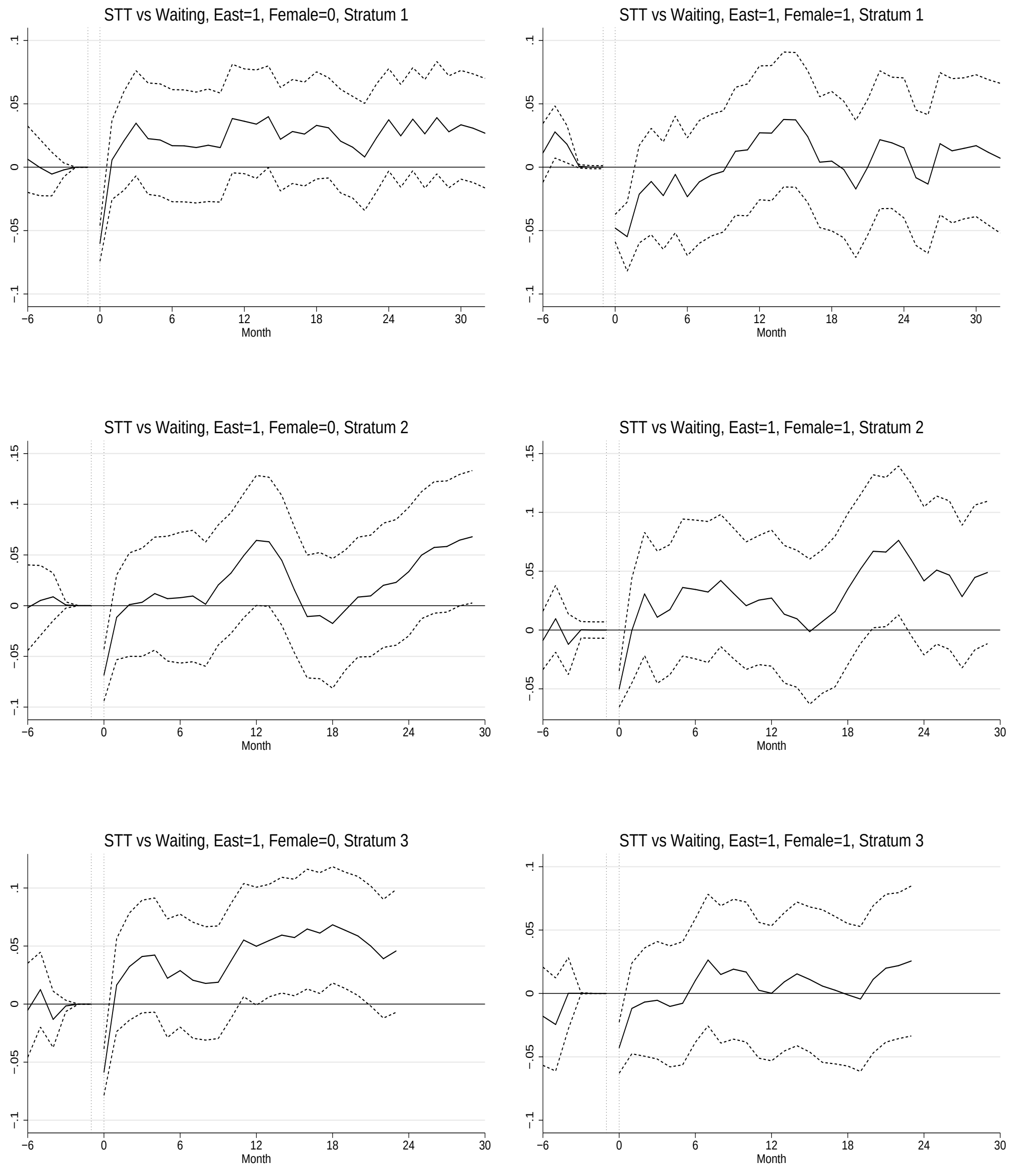
Figure 1.2: Treatment Effect STT vs. Waiting, West Germany



Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Stratum 1 denotes entry into program during months 0 to 3 of unemployment, stratum 2 during months 4 to 6, and stratum 3 during months 7 to 12.

Figure 1.3 presents the corresponding results for East Germany. They suggest that short-term training measures in East Germany generally do not have any positive effects on the employment probability of their participants. Measured average treatment effects are mostly small and statistically insignificant. The only exception are men who receive treatment in months 7 to 12 (stratum 3) of their unemployment spell. For these individuals, participating in short-term training increases their long-

Figure 1.3: Treatment Effect STT vs. Waiting, East Germany

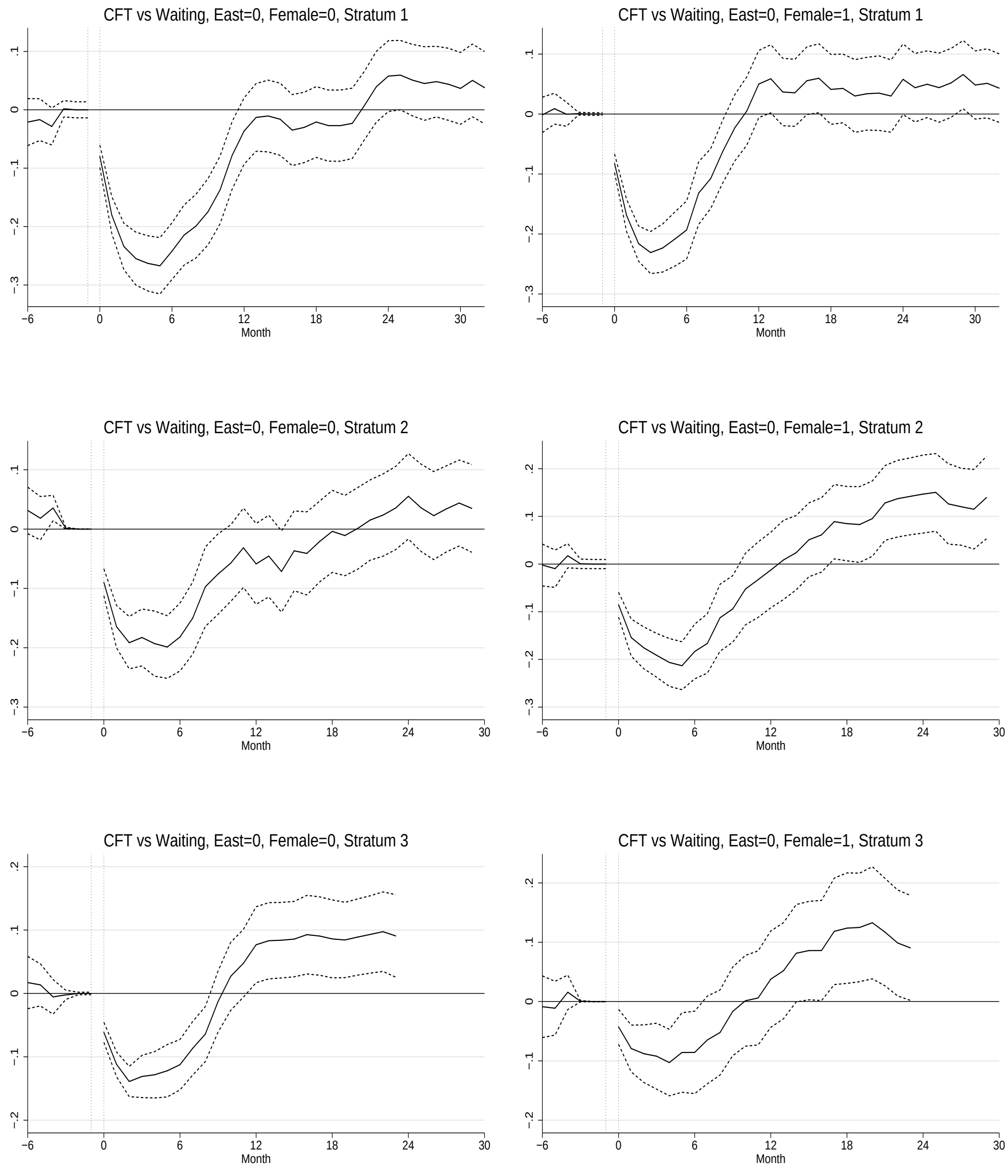


Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Stratum 1 denotes entry into program during months 0 to 3 of unemployment, stratum 2 during months 4 to 6, and stratum 3 during months 7 to 12.

term employment probability by about 5 percentage points. However, this effect is only marginally statistically significant and does not seem to last in the long run.

Results for the more substantive classroom further training measures (CFT) are given in figures 1.4 and 1.5. The most conspicuous difference between these results and those for short-term training programs is the long and pronounced lock-in effect. During the first months of their participation in the program, participants have

Figure 1.4: Treatment Effect CFT vs. Waiting, West Germany



Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Stratum 1 denotes entry into program during months 0 to 3 of unemployment, stratum 2 during months 4 to 6, and stratum 3 during months 7 to 12.

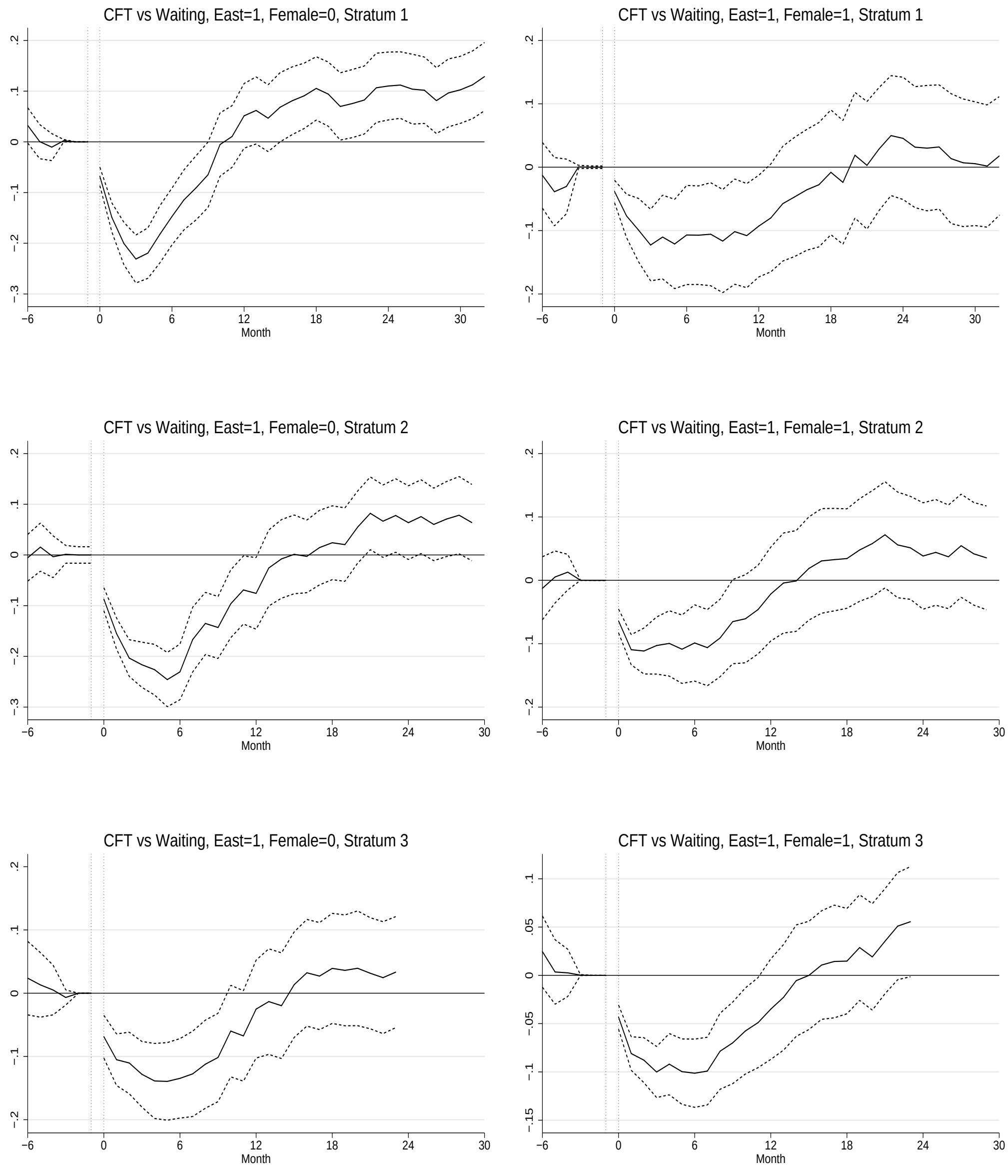
an employment rate that is up to 25 percentage points lower than it would have been if they had not taken part in the program. The lock-in period lasts up to 12 months for individuals who take up their treatment during the first 6 months of their unemployment spell. Interestingly, lock-in effects are less deep and shorter for individuals that have been unemployed for more than 6 months (stratum 3).

There are several possible reasons for this finding. First, it might be that individuals with a longer elapsed unemployment duration are assigned to shorter measures within the group of CFT programs. Second, it is possible that such individuals drop out of the program more often or earlier. A third reason may be that a large number of those just having become unemployed easily find new jobs if they do not take part in a training program. If these individuals are assigned to CFT measures anyway, they will be 'locked-in', while many of their counterparts in the control group have already found employment. This would imply that some of the short-term unemployed receive training even though they do not need it to overcome unemployment. In addition, there may be a tendency towards finding less pronounced lock-in effects for the late program starts if many of the long-term unemployed in the control group abandon their job search and move out of labor force. Hence, an additional channel through which training programs work may consist in keeping the long-term unemployed in the labor force.

While there is little evidence for statistically significant employment effects for West German men starting classroom further training in months 0 to 6 of their unemployment spell (strata 1 and 2) or West German women starting it in the first 3 months of unemployment (stratum 1), treatment effects for longer-term unemployed men (stratum 3) and medium to longer-term unemployed women (strata 2 and 3) are large and statistically significant. After the initial lock-in phase, they amount to some 8 percentage points for men and to some 10 percentage points for women. The corresponding results for classroom further training measures in East Germany are given in figure 1.5. As in West Germany, there are long and deep lock-in effects of up to twenty five percentage points in the first 12 to 15 months after treatment start. With the exception of men starting their program relatively early in their unemployment spell (stratum 1), there is no evidence for positive treatment effects after initial lock-in.

In contrast to pure classroom further training, practical further training (PFT) also includes practical elements such as internships or working in a practice firm. Evaluation results for these measures are given in figure 1.6. The results for West Germany shown in the first row of figure 1.6 suggest considerable positive employment effects of about 10 percentage points for women after a lock-in period of up to 8 months. There are no such effects for men. A reason for this finding could be that particularly in practice-related jobs, men and women select themselves into different occupations. If women more often participate in training for occupations in the service sector, where employment chances are generally better than in manufactur-

Figure 1.5: Treatment Effect CFT vs. Waiting, East Germany



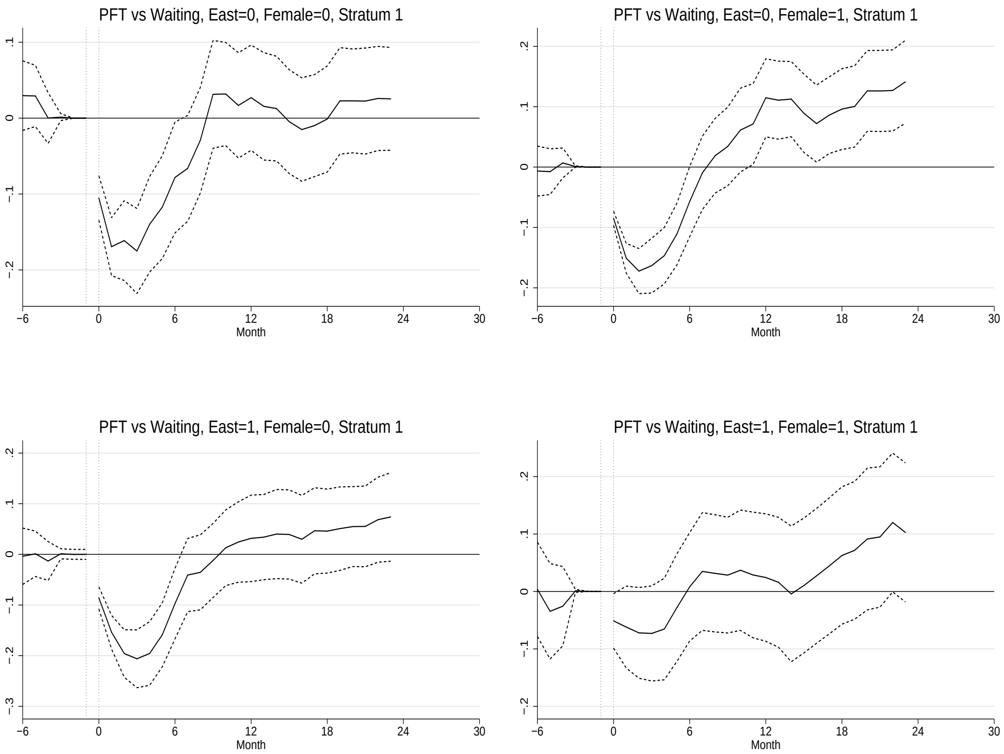
Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Stratum 1 denotes entry into program during months 0 to 3 of unemployment, stratum 2 during months 4 to 6, and stratum 3 during months 7 to 12.

ing or construction jobs, this will lead to more positive effects of practical training measures for women.[18]

Similar as for other types of training, there are no employment effects for participants in PFT in East Germany (second row of figure 1.6). The negative picture our results

[18]See Lechner et al. (2005b) or Fitzenberger and Völter (2007) who consider gender specific target professions for public sector sponsored training in East Germany in the 1990s.

Figure 1.6: Treatment Effect PFT vs. Waiting, West and East Germany

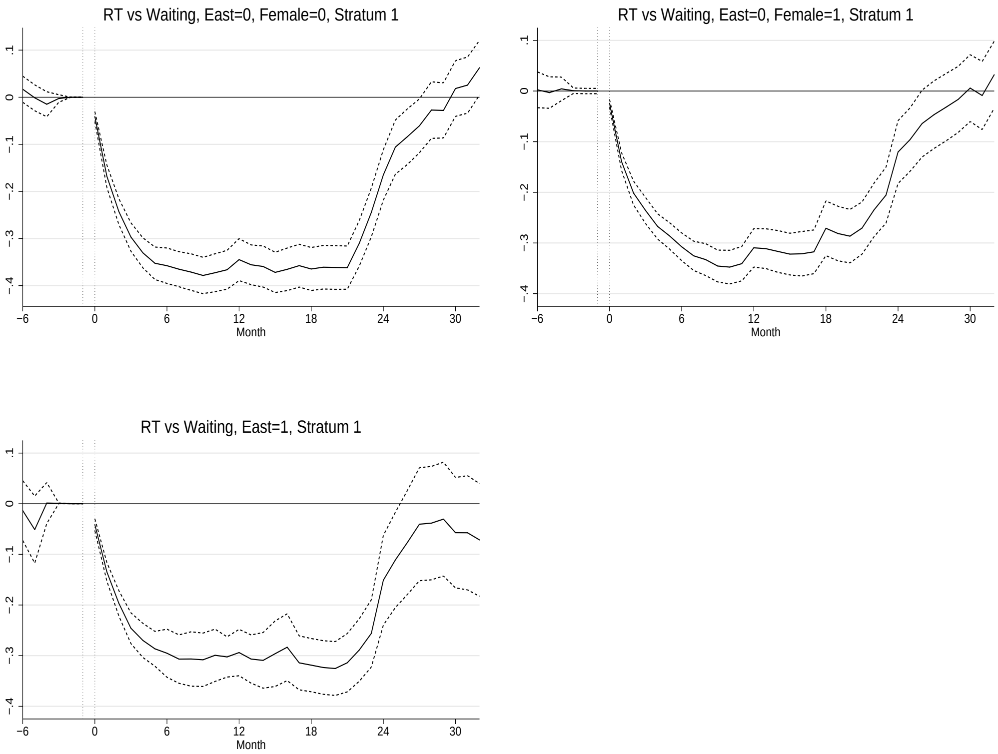


Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Entry into program during months 0 to 12 of unemployment spell.

draw for East Germany probably reflects the difficult labor market situation in large parts of East Germany. In districts where open jobs are extremely rare in all sectors, the potential employment effects of training programs may be very limited. In addition to this, it is likely that the group of participants in East Germany differs to some extent from that in West Germany. In regions with very high unemployment rates, training programs may to a certain extent be used to reduce the frustration of those who want to work, but have no employment prospects. In fact, differences in selection into treatments may induce differences in treatment effects, if treatment effects are heterogeneous.

Finally, figure 1.7 shows estimated treatment effects for the very long retraining measures. Although the large majority of these programs do not last longer than two years (see figure 1.16), no statistically positive employment effects can be observed up to thirty months after program start. On the contrary, retraining measures cause a grave lock-in effect of minus forty percentage points during most of the program's

Figure 1.7: Treatment Effect RT vs. Waiting, West and East Germany



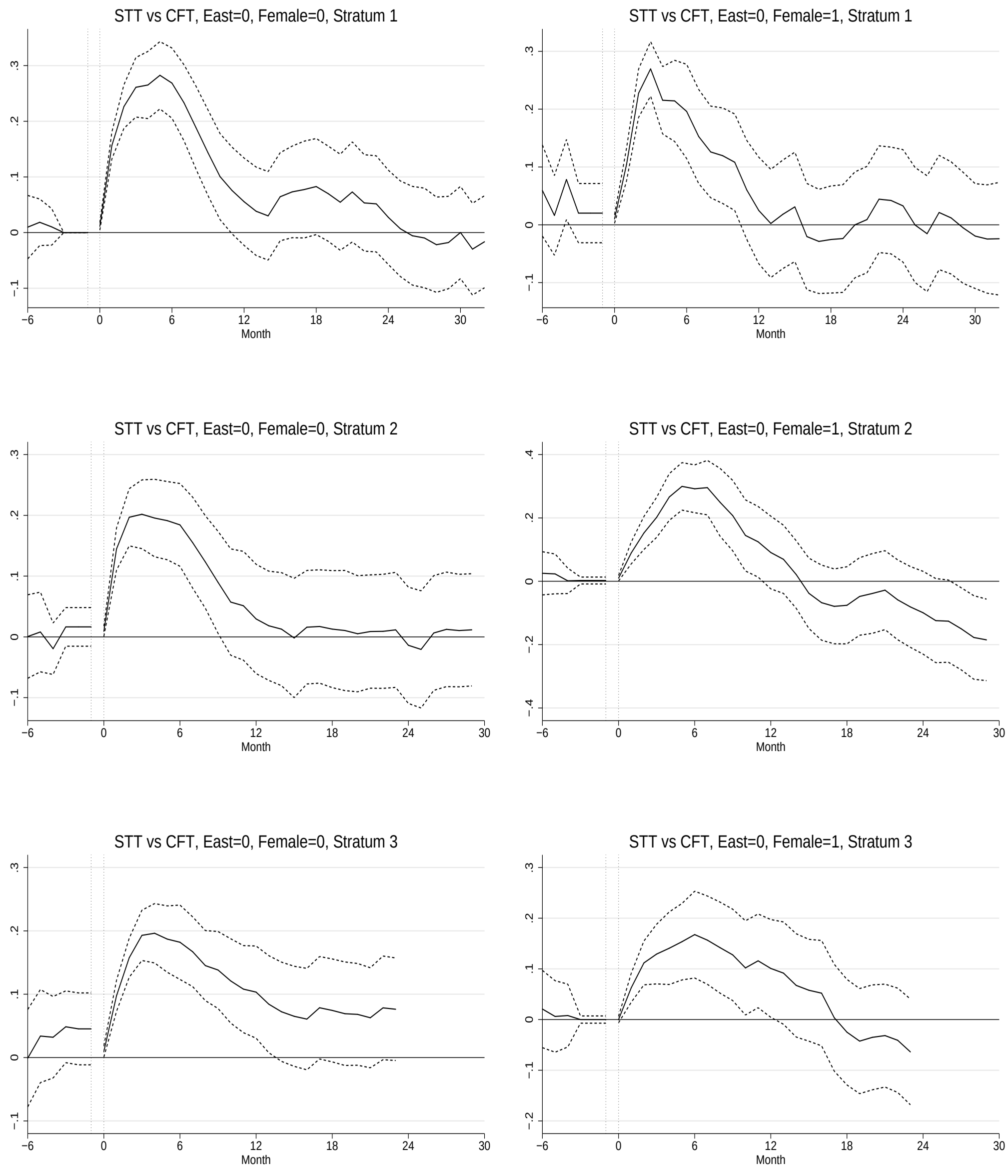
Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Entry into program during months 0 to 12 of unemployment spell. East Germany: pooled sample containing male and female participants.

duration. Just in order to break even, employment gains after completion of the program have to be very large and long-enduring given the large loss in employment probability caused by the participation in the program (see discussion of cumulated effects below).

1.6.2 Pairwise Evaluation of Training Programs

Given that in many cases, especially in West Germany, training programs may have considerable employment effects when compared to attending no program, the question arises which of the different training programs is the most effective for a given subpopulation. The results presented in the previous section already suggest that short-term training may have similar positive effects as classroom further training or practical further training when each of the programs is compared to attending no program. However, this does not necessarily mean that participants in short-term training could not have improved their employment chances by attending classroom or practical further training instead, or that participants in classroom or practical further training would not have lost from taking part in short-term training instead. This is the question we address next.

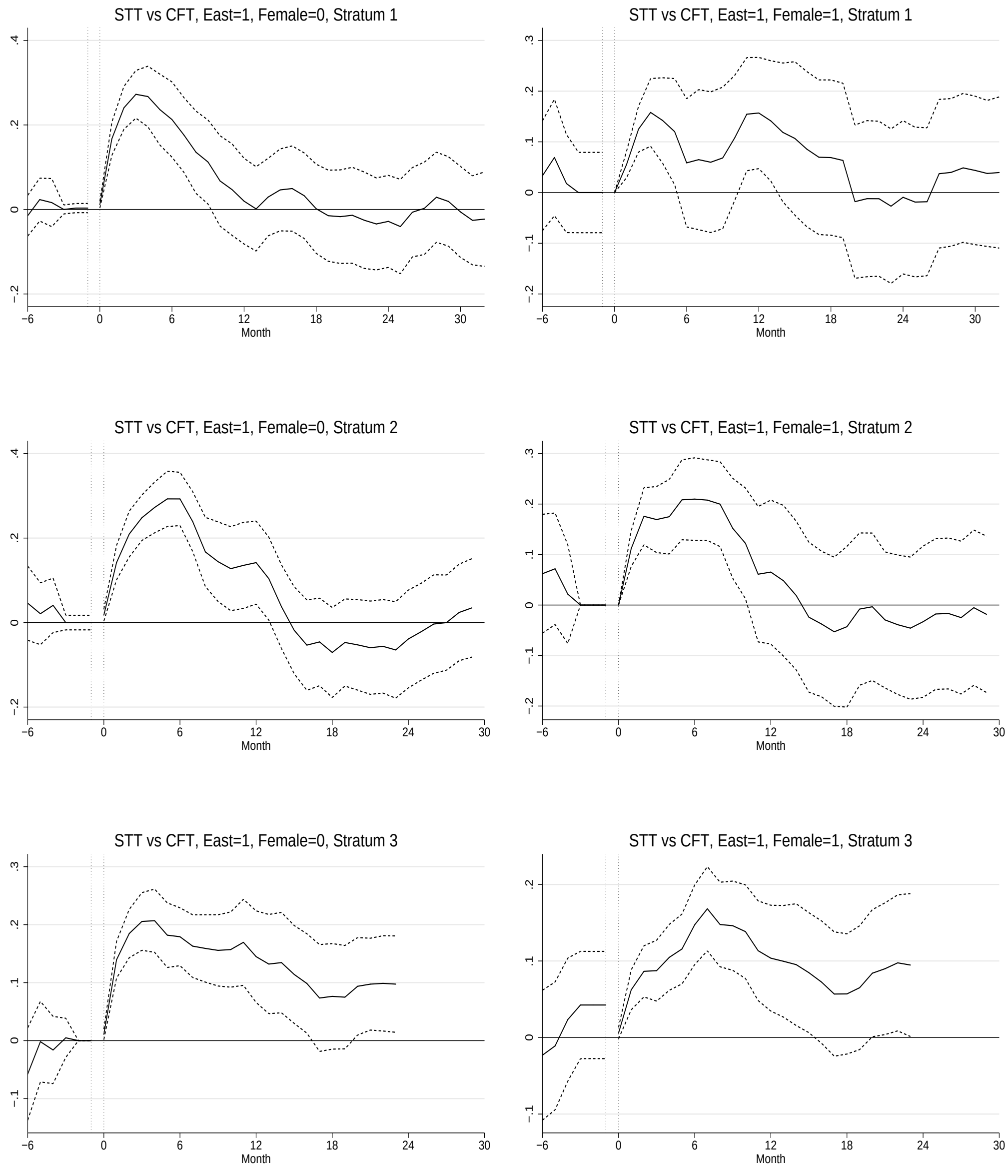
Figure 1.8: Treatment Effect STT vs. CFT, West Germany



Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Stratum 1 denotes entry into program during months 0 to 3 of unemployment, stratum 2 during months 4 to 6, and stratum 3 during months 7 to 12.

Figures 1.8 and 1.9 show that participants in short-term training generally would *not* have improved their employment chances by attending class room further training,

Figure 1.9: Treatment Effect STT vs. CFT, East Germany



Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Stratum 1 denotes entry into program during months 0 to 3 of unemployment, stratum 2 during months 4 to 6, and stratum 3 during months 7 to 12.

neither in East nor in West Germany.[19] On the contrary, the much shorter and less pronounced lock-in effect of short-term measures makes this form of training seem more effective than the longer-term classroom training. This applies especially to East German participants in the later strata who would have significantly lowered

[19]The evidence is not so clear for West German women who started the program after having been unemployed for more than 4 months.

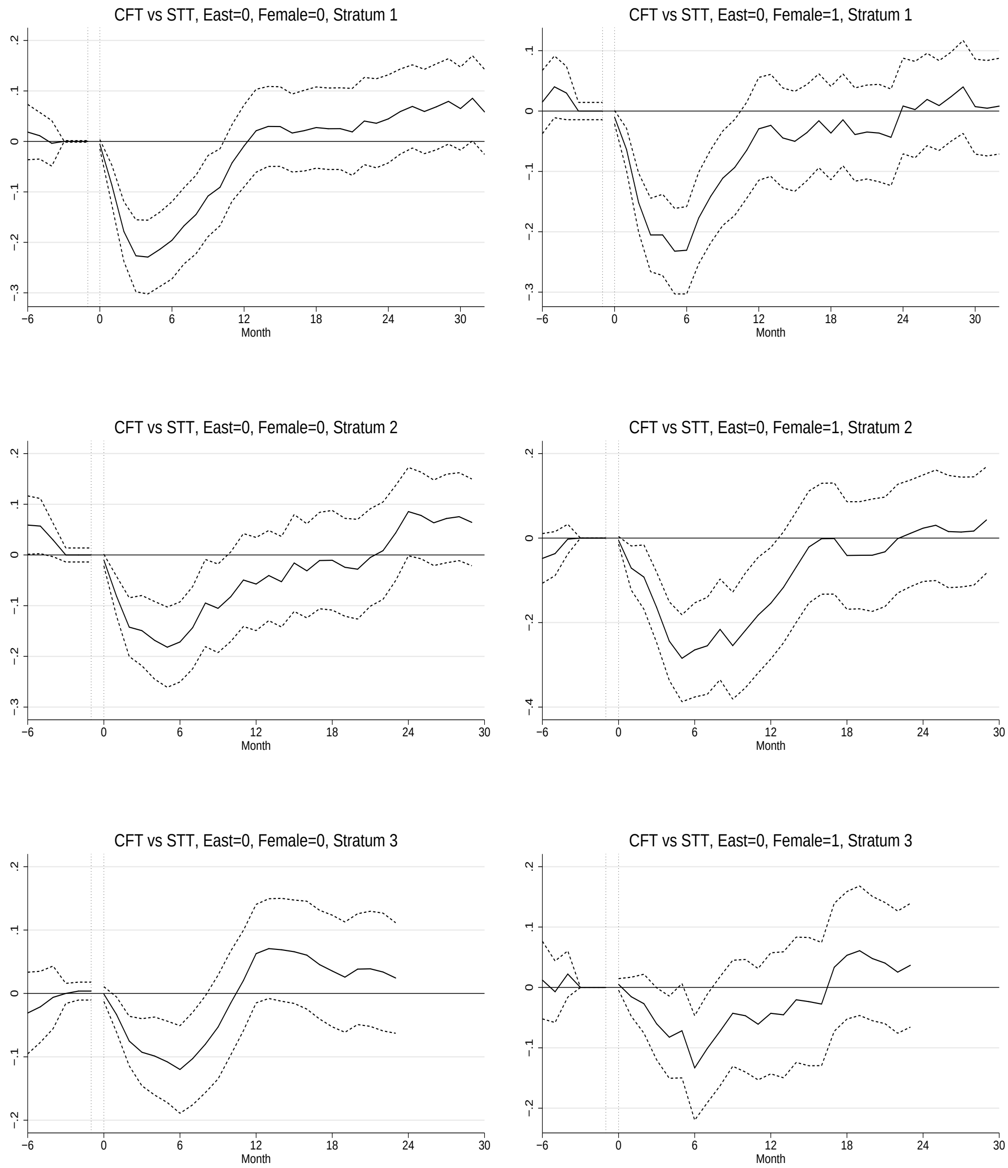
their employment probability even in the long-run if they had attended classroom further training instead of short-term training. Furthermore, figures 1.10 and 1.11 show that participants in classroom further training would *not* have lost from attending short-term training instead. This is remarkable since it means that these individuals could have been assigned to the much less expensive short-term measures without reducing their employment chances. Again, taking the shorter lock-in effect of short-term measures at program start into account, short-term training would have even been preferable for these individuals. Taken together, classroom further training is on balance not more effective than short-term training.

How does practical further training compare to short-term training? Figure 1.12 shows that individuals who took part in short-term training would not have gained from attending practical training instead. However, figure 1.13 indicates that practical further training was significantly *more effective* for West Germans taking part in this kind of training than short-term training would have been. This means that it would not necessarily have been advisable for these individuals to substitute the longer practical training courses by the shorter programs. It also means that, to a certain extent, participants were well allocated to courses, as neither individuals in short-term training nor individuals in practical further training would have gained by reallocating them to the other program. As the last row of figure 1.13 shows, this does not necessarily apply to East German participants in practical training whose employment chances would not have been significantly reduced if they had been reallocated to short-term training.

How do the practical training courses compare to the more theoretical classroom further training courses? Evidence on this comparison is given in figures 1.14 and 1.15. Figure 1.14 suggests that practical training was better for West German participants of practical training than classroom training would have been. This holds especially for female participants in West Germany whose employment probability was significantly higher than it would have been in classroom training even long after the programs ended. Note that in West Germany practical training programs also exhibited significantly smaller lock-in effects which is not surprising given their shorter length. The lower row of figure 1.14 shows that East German participants in practical training would neither have gained nor would they have lost from taking part in classroom further training instead.

We omit comparisons with the long retraining programs as these comparisons are

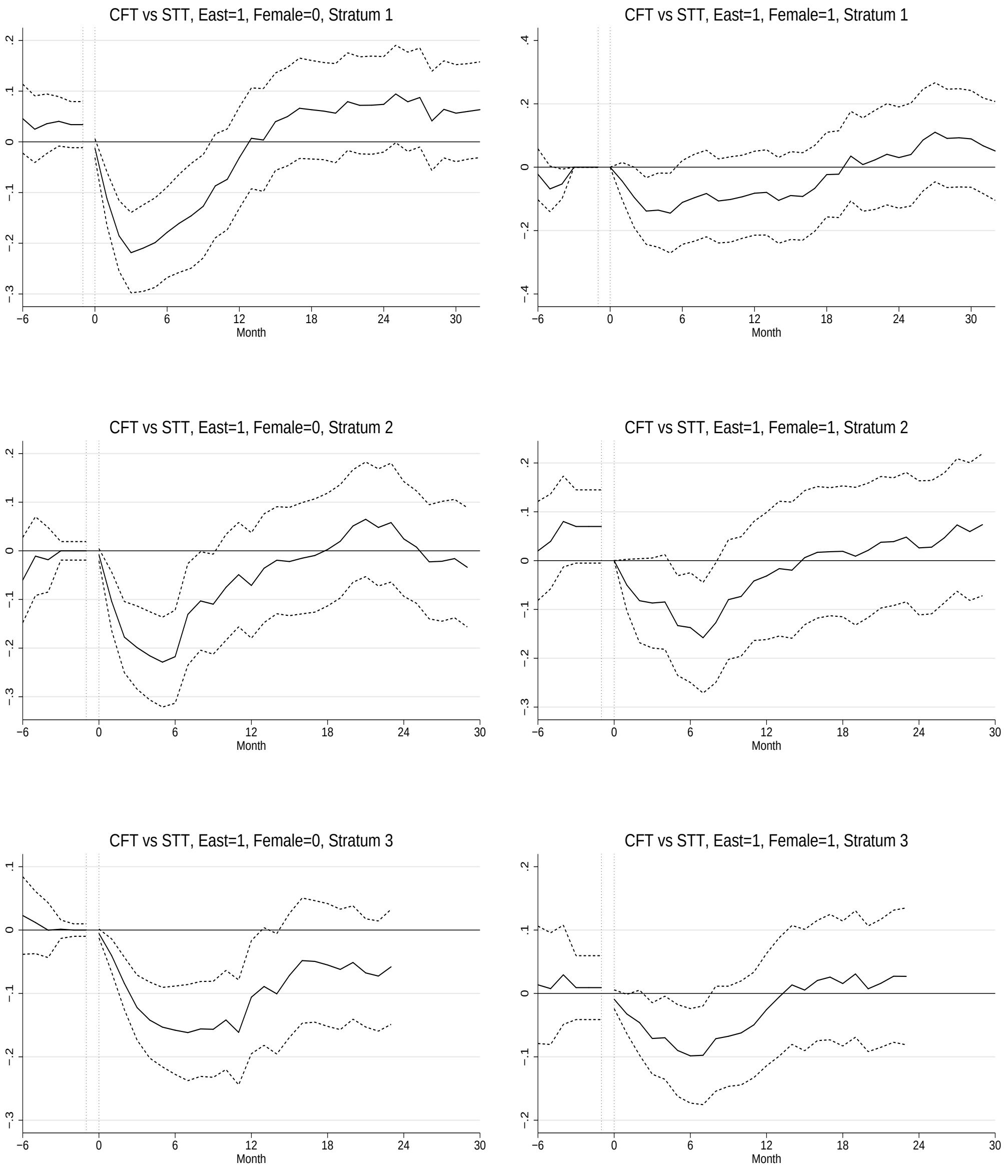
Figure 1.10: Treatment Effect CFT vs. STT, West Germany



Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Stratum 1 denotes entry into program during months 0 to 3 of unemployment, stratum 2 during months 4 to 6, and stratum 3 during months 7 to 12.

entirely determined by the extensive lock-in effect of retraining (see figure 1.7). The general conclusion is that shorter programs with positive employment effects outperform longer programs due to the difference in the length of the lock-in period.

Figure 1.11: Treatment Effect CFT vs. STT, East Germany

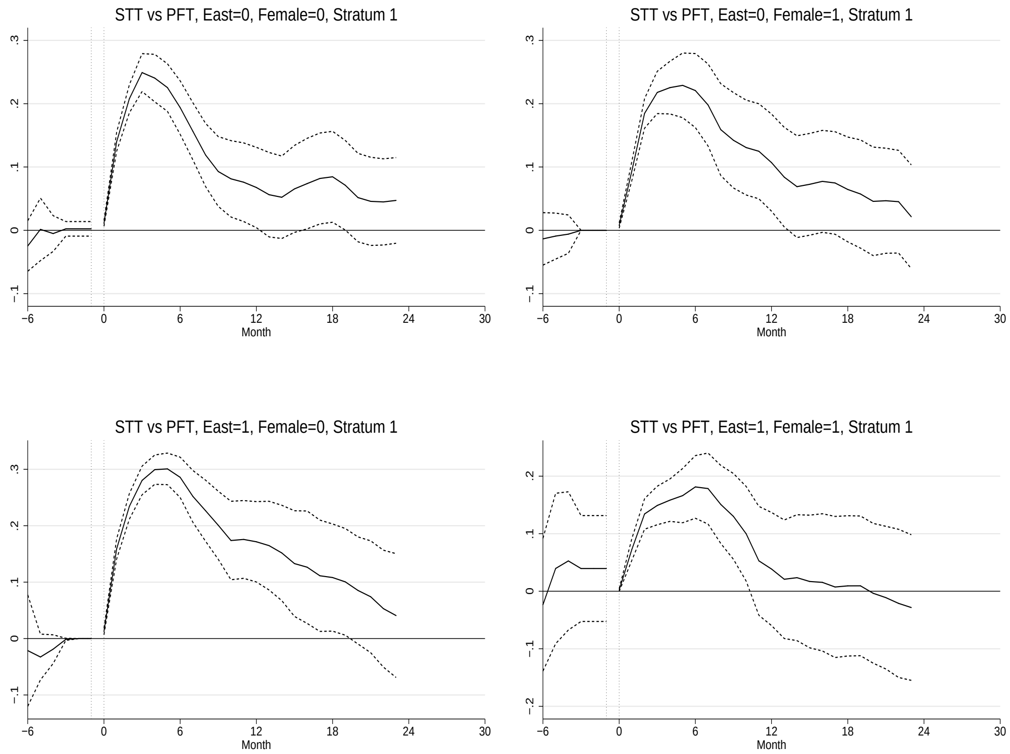


Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Stratum 1 denotes entry into program during months 0 to 3 of unemployment, stratum 2 during months 4 to 6, and stratum 3 during months 7 to 12.

1.6.3 Cumulated Effects

The higher effectiveness of shorter programs is confirmed in table 1.7 in the appendix which shows gains and losses in months employed for all pairwise comparisons cumulated over two years after program start. For example, for West German men participating in short-term training after having been unemployed for at least 7

Figure 1.12: Treatment Effect STT vs. PFT, West and East Germany

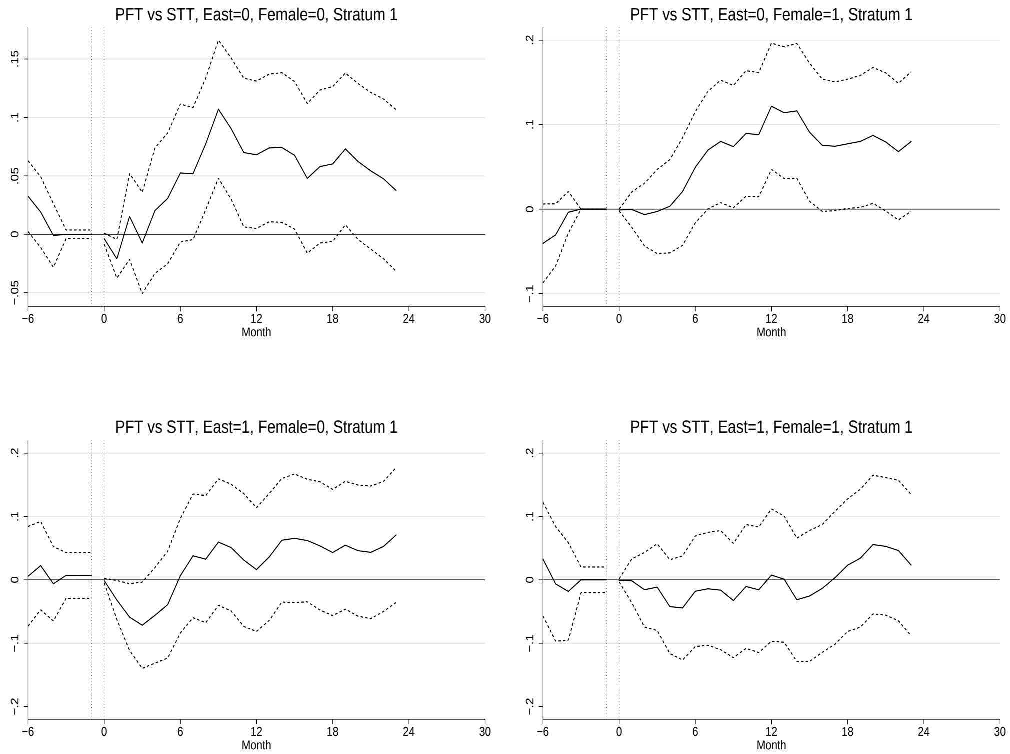


Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Entry into program during months 0 to 12 of unemployment spell.

months (stratum 3) the net gain of the program versus waiting is 0.903 employment months during the first 24 months after the program start. The effects are even stronger for West German women who gain on average 1.767 (stratum 2) and 2.122 (stratum 3) employment months. In East Germany, only long-term unemployed men gain from taking part in short-term training (plus 0.947 employment months, stratum 3). The results suggest that short-term training is the only form of training that has positive and statistically significant cumulated employment effects during the first two years, and this only for the individuals who are not treated too early after entering unemployment.

Table 1.7 also nicely summarizes the comparative effectiveness of the different programs. Rows 2, 3, 16 and 17 show that short-term training was better in terms of cumulated employment months for those participating in it than classroom further training or practical further training would have been (STT vs. CFT). On the other hand, rows 6 and 20 (CFT vs. STT) suggest that for participants in classroom fur-

Figure 1.13: Treatment Effect PFT vs. STT, West and East Germany

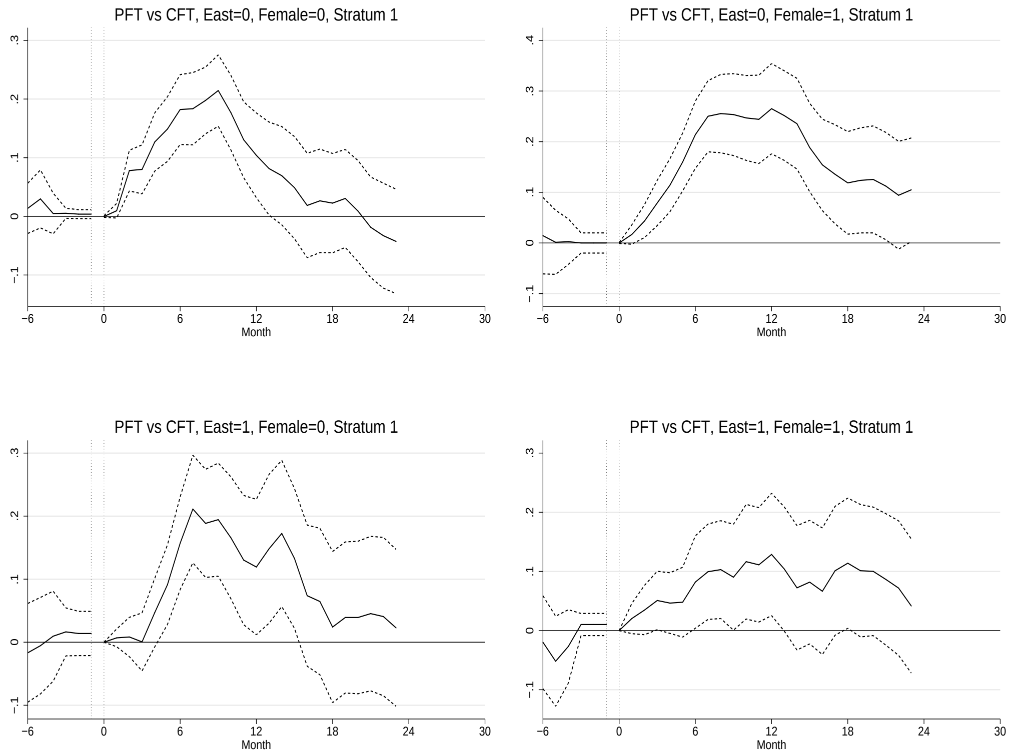


Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Entry into program during months 0 to 12 of unemployment spell.

ther training, short-term training would have been more effective. However, row 10 (PFT vs. STT) indicates that short-term training is not uniformly better than the longer further training programs as West German participants in *practical* further training would have lost if they had been assigned to short-term training measures instead.[20] On the one hand, the practical training programs stand out as quite effective, as e.g. participants of these courses fared significantly better than if they had taken part in classroom further training courses (rows 11 and 25). On the other hand, participants of classroom further training would not necessarily have gained from switching to more practical courses (rows 7 and 21).

[20]Of course, practical training courses are much more costly so that short-term training may still have been the better alternative.

Figure 1.14: Treatment Effect PFT vs. CFT, West and East Germany



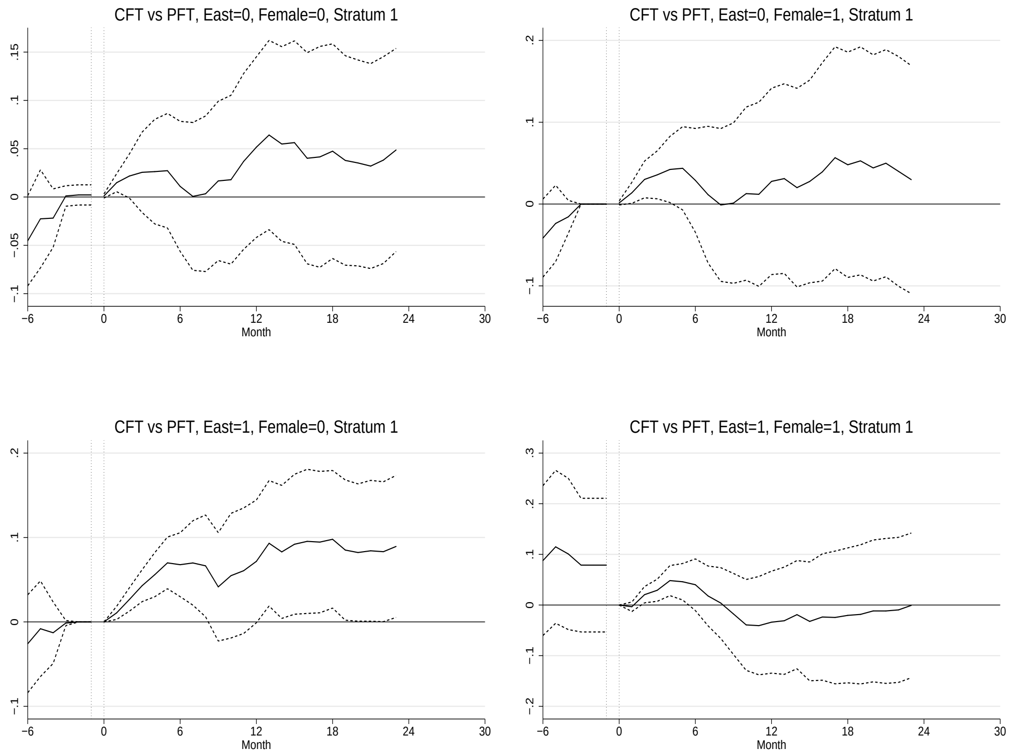
Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Entry into program during months 0 to 12 of unemployment spell.

1.6.4 Effect Heterogeneity

The results presented so far reveal considerable heterogeneity in the effects of the programs in different subpopulations. A consistent finding for West Germany seems to be that programs are only effective for those individuals who start a program after having been unemployed for some time. In East Germany this seems to be reversed (see first two graphs in figure 1.5), suggesting that selection of subpopulations into treatment may differ between East and West. Another finding is that training effects are generally larger for women than for men (see e.g. figures 1.2 and 1.4). Women also seem to benefit from practical training, while men do not (figure 1.6). These results show that treatment effects averaged over too broad subpopulations or heterogenous programs may hide statistically and economically significant effects for particular subprograms or subpopulations.

In order to investigate effect heterogeneity also *within* the subgroups defined above,

Figure 1.15: Treatment Effect CFT vs. PFT, West and East Germany



Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa. Entry into program during months 0 to 12 of unemployment spell.

we regressed cumulated individual treatment effects on a number of observed personal characteristics, see equation (1.4). Concretely, we considered individual treatment effects (i.e. the difference between the actual and the nonparametrically predicted counterfactual outcome) cumulated over the period after lock-in (STT: after month 8, PFT: after month 12, and CFT: after month 18) and divided by the number of months after lock-in. We are interested in how the treatment effects vary with personal characteristics. We also account for the mismatch of the treated individual and the control group with respect to particular individual characteristics. This allows us to compute 'mismatch corrected' treatment effects (by omitting the mismatch when calculating average treatment effects based on the regression of individual treatment effects on personal characteristics). If our matching approach works well, the mismatch-corrected average treatment effects have to coincide with uncorrected average treatment effects.

The results are given in tables 1.8 to 1.10 in the appendix. Generally, we only report specifications with significant covariates (we omit some covariates because of likely multicollinearity problems). Therefore, a case with no coefficients reflects a situation where no significant covariates was found. We find some evidence for effect heterogeneity as in many cases older participants benefited less or not at all from training programs. This holds specifically for East and West German participants in short-term training (see table 1.8), West German women and East German men and women participating in classroom further training (table 1.9) and West German women participating in practical further training (table 1.10). In some cases, treatment effects also vary with educational qualifications, especially in the case of West German men taking part in short-term training (top panel of table 1.8) and West German women taking part in classroom further training (middle panel of table 1.9). In both cases low educational qualifications harm the effect of training on employment outcomes. We also note that in all cases, the mismatch-corrected estimates of average treatment effects coincide with the uncorrected estimates which supports the validity of our matching procedure.

We also investigated effect heterogeneity in pairwise program comparisons, but the results were less clear-cut.[21] The general finding of both the comparisons of program vs. non-participation and the cross-program comparisons seems to be that - in addition to the heterogeneity found between the subgroups defined by gender, region, and unemployment duration - there is generally little heterogeneity along observed characteristics. However, when there is such heterogeneity, it can be very strong. This suggests that a better targeting of the programs could be achieved along these lines.

1.7 Conclusions

This paper analyzes and compares the employment effects of the four important types of public sector sponsored training in Germany in the early 2000s. These are short-term training (STT), classroom further training (CFT), practical further training (PFT), and retraining (RT). In light of recent policy reforms fostering shorter training programs, we are particularly interested in the question of how short-term training programs compare in terms of effectiveness to traditional medium-term further training schemes. Our econometric approach uses nonparametric kernel

[21]These results are available upon request.

matching methods in a dynamic, multiple treatment framework.

Our results suggest that the effectiveness of the different programs strongly depends on the personal characteristics of the participants and the circumstances of program participation. For West Germany, we find statistically significant positive employment effects for male and female participants in short-term training and classroom further training who started their training not too early during their unemployment spell. Moreover, West German women but not West German men benefited from practical further training measures. A closer look reveals that, within the time window permitted by our data set, employment effects of short-term training were of a similar magnitude as those of traditional medium-term measures, but, due to the shorter length, these positive effects materialized much earlier. According to our results, West German men taking part in short-term training or medium-term training may increase their medium-term employment rate by some 5 to 10 percentage points. The effect for women is even larger, leading to increases in employment probabilities of 10 percentage points or more.

The surprising effectiveness of short-term training when compared to the different forms of medium-term training is also confirmed in pairwise comparisons. According to our results, participants in short-term training would in general not have gained if they had taken part in medium-term further training, and participants in the latter programs would generally not have lost if they had been assigned to short-term courses instead. In particular, this holds for the comparison of short-term training and classroom further training. However, this is less clear in comparison to practically oriented further training, where it appeared that participants of practical further training courses may have reduced their employment chances if they had taken part in short-term training instead. As to the comparison of classroom vs. practical further training, our results suggest that practical training may have advantages over pure classroom training, a finding that is consistent with the international evidence (see Martin and Grubb (2001) and OECD (2005)), but not with evidence for Germany in the 1990's (Lechner et al. (2005a), Fitzenberger et al. (2006a), Fitzenberger and Völter (2007)).

We do not find any positive employment effects for the long retraining measures during the time window permitted by our data. Even more than six months after the completion of such a program, the employment rate of participants is below than or equal to the employment rate of a comparable control group of non-participants.

Given the strong lock-in effect of these programs – during participation participants have employment rates that may be 40 percentage points lower than those of non-participants – it seems extremely unlikely that these measures justify their large costs. Long-run evidence on retraining during the 1990's in Lechner et al. (2005a,b), Fitzenberger et al. (2006a), and Fitzenberger and Völter (2007) suggests that, although there may be positive long-run effects, cumulated employment effects are always lower than those of shorter programs.

The general ineffectiveness of long-term retraining and the only moderate effectiveness of medium-term training when compared to short-term training suggests that the policy shift that took place in 2002 and that massively substituted long-term and medium-term training courses by inexpensive short-term courses was justified. In fact, given our results, it may be well the case that further substitution of long- and medium-term programs by short-term training may be warranted. Although we lack detailed information on training costs, the results in table 1.7 in the appendix combined with the information on average costs in table 1.1 suggest that an average short-term training course costing only 627 Euros may lead to an employment gain of one month within twenty four months after program start, while an average further training or retraining course costing 6175 Euros leads to *no gain* or even a loss in months employed within the first twenty four months after program start.

While it is a common finding shared by studies such as Lechner et al. (2005a,b), Fitzenberger et al. (2006a) and Hujer et al. (2004) that, especially if lock-in effects are taken into account, shorter training programs may outperform longer ones,[22] it seems surprising that short-term programs lasting only two to twelve weeks may have employment effects at all. Given that it is hard to believe that such short programs lead to substantial increases in human capital, other aspects may be more relevant. Looking at the particular contents of short-term training analyzed here, it seems more plausible that these programs help to activate their participants who may otherwise not look as intensively for new jobs as they do when they are assigned short-term training programs that often comprise elements of profiling, job search assistance, or monitoring. Our results are therefore in line with a number of recent studies that focus on the positive effects of increased job search assistance and activation, see e.g. Blundell et al. (2004), Weber and Hofer (2004), Fougère et al. (2005), Hujer et al. (2005), Crépon et al. (2005), and Van den Berg and Van der

[22] However, note that neither of these studies consider the very short training programs analyzed here.

Klaauw (2006).

Furthermore, our results show that the effects of training programs may be very different across different subgroups. One result is that employment effects are usually larger for individuals who start their program at a later point during their unemployment spell. In fact, in many cases we do not find significant employment effects for individuals who start their treatment very early in their unemployment spell. It would be wrong to conclude from this that treatment is the more effective the later it is provided to the participants as individuals who are long-term unemployed may differ in observed and unobserved characteristics from those who are short-term unemployed. However, the result is remarkable because it suggests that in cases of long-term unemployment, training programs may help to restore the employment chances of their participants. We also find that training effects may be heterogenous with respect to gender, age, and qualification. In line with other results in the literature (see Bergemann and van den Berg (2006) for an overview) we find that employment effects of training are generally larger for women than for men. Moreover, it seems that older individuals benefit less from the training programs analyzed here. In some cases, this also applies to individuals with low educational qualifications. Based on these results, it seems advisable that the targetting in program assignment should be improved.

To analyze the overall welfare effect of these programs is far beyond this study's scope and would require much more information. First, to assess the program benefits, it would be necessary to follow the participants for a longer time period. But for relatively recent program starts it is not possible to estimate long run effects to see how long positive effects will last. However, the overall program effect would be crucial to calculate the program benefits concerning transfer payments and tax revenues. Second, this study only estimates the effects these programs have on their participants. To assess the general welfare effect knowledge on general equilibrium effects would be important. These are likely to be negative (for example there might be substitution effects), but could also be positive (risk averse individuals might be more likely to decide for an apprenticeship in an emerging industry if there are public sponsored retraining programs available in case their decision turns out to be wrong). Third, there may also exist direct program benefits which are difficult or even impossible to measure, like, for example, a positive effect on educational outcomes for children due to a human capital increase or improved economic and social situation of the parents, or a decline in criminal activity. With regard to the costs detailed information on direct and indirect costs of the programs would be required.

Finally, in contrast to the result of positive treatment effects in a number of cases for West Germany, we find only little evidence for positive treatment effects in East Germany. Apart from positive effects for East German men taking part in short- and medium-term training after having been unemployed for more than six months, and positive effects for men beginning classroom further training in the first three months of their unemployment spell, we see little benefits from short-, medium-, or long-term training in East Germany. In particular, we do not find any positive effects for women. Our results for East Germany reflect the generally difficult labor market situation in the East, especially for women. High unemployment rates seem to render both short and medium-term training programs ineffective to a large extent, showing that the effect of training may strongly depend on the specific circumstances of the labor market under consideration. The ineffectiveness of training in East Germany in the early 2000's is in line with results by Lechner and Wunsch (2006), although the latter paper takes a different methodological approach. However, the results are in contrast to the somewhat more positive findings for the 1990's (Lechner et al. (2005b), Fitzenberger and Speckesser (2007), Fitzenberger and Völter (2007)).

References

Bender, S., M. Biewen, B. Fitzenberger, M. Lechner, S. Lischke, R. Miquel, A. Osikominu, T. Wenzel, C. Wunsch (2004). "Die Beschäftigungswirkung der FbW-Maßnahmen 2000 - 2002 auf individueller Ebene: Eine Evaluation auf Basis der prozessproduzierten Daten des IAB - Vorläufiger Zwischenbericht Oktober 2004", Goethe University Frankfurt and SIAW St. Gallen.

Bender, S., M. Biewen, B. Fitzenberger, M. Lechner, R. Miquel, A. Osikominu, M. Waller, C. Wunsch (2005). "Die Beschäftigungswirkung der FbW-Maßnahmen 2000 - 2002 auf individueller Ebene: Eine Evaluation auf Basis der prozessproduzierten Daten des IAB - Zwischenbericht Oktober 2005", Goethe University Frankfurt and SIAW St. Gallen.

Bergemann, A. and G. J. van den Berg (2006). "Active Labor Market Policy Effects for Women in Europe: A Survey". IZA Discussion Paper No. 2365.

Biewen, M. and B. Fitzenberger (2004). "Neuausrichtung der Förderung beruflicher Weiterbildung." In: Hagen, T. and A. Spermann *Hartz-Gesetze - Methodische Ansätze zu einer Evaluierung*, ZEW Wirtschaftsanalysen, 74, Baden-Baden, 190-202.

Blundell, R., M. Costas Dias, C. Meghir, J. Van Reenen (2004). "Evaluating the Employment Impact of a Mandatory Job Search Program". *Journal of the European Economic Association* 2, 569-606.

Bundesagentur für Arbeit (2001, 2002a, 2003a, 2004a). *Daten zu den Eingliederungsbilanzen 2000, 2001, 2002, 2003*, Nürnberg (various issues).

Bundesagentur für Arbeit (2002a, 2003a, 2004a, 2005a). *Arbeitsmarkt 2001, 2002, 2003, 2004*, Nürnberg (various issues).

Bundesagentur für Arbeit (2005b). *Geschäftsbericht 2004*, Nürnberg.

Crépon, B., M. Dejemeppe and M. Gurgand (2005). "Conseling the unemployed: does it lower unemployment duration and recurrence?" IZA Discussion Paper 1796.

Dolton, P. and D. O'Neill (2002). "The Long-Run Effects of Unemployment Monitoring and Work-Search Programs: Experimental Evidence from the United Kingdom." *Journal of Labor Economics* 20, 381-403.

Dyke, A., C.J. Heinrich, P.R. Mueser, K.R. Troske, and K.-S. Jeon (2006): "The Effects of Welfare-to-Work Program Activities on Labor Market Outcomes", *Journal of Labor Economics* 24, 567-607.

Fitzenberger, B., A. Osikominu and R. Völter (2006a). "Get Training or Wait? Long Run Employment Effects of Training Programs for the Unemployed in Germany." ZEW Discussionpaper No. 06-039.

Fitzenberger, B., A. Osikominu, and R. Völter (2006b). "Imputation Rules to Improve the Education Variable in the IAB Employment Subsample." Forthcoming *Journal of Applied Social Science Studies.*

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

Fitzenberger, B., R. Völter (2007). “Long Run Effects of Training Programs for the Unemployed in East Germany.” Forthcoming *Labour Economics*.

Fougère, D., J. Pradel and M. Roger (2005). “Does Job Search Assistance Affect Search Effort and Outcomes?”. IZA Discussion Paper 1825.

Fredriksson, P. and P. Johansson (2003). “Program Evaluation and Random Program Starts.” Institute for Labour Market Policy Evaluation (IFAU), Uppsala, Working Paper, 2003:1.

Fredriksson, P. and P. Johansson (2004). “Dynamic Treatment Assignment - The Consequences for Evaluations Using Observational Data.” IZA Discussion Paper No. 1062.

Gerfin, M. and M. Lechner (2002). “Microeconometric Evaluation of the Active Labor Market Policy in Switzerland.” *Economic Journal*, 112 (482), 854-893.

Hardoy, I. (2005): ”Impact of Multiple Labour Market Programmes on Multiple Outcomes: The Case of Norwegian Youth Programmes.” *Labour* 19, 425-467.

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

Hujer, R., S.L. Thomsen, C. Zeiss (2004). “The Effects of Vocational Training Programmes on the Duration of Unemployment in Germany”. IZA Discussion Paper No. 1117.

Hujer, R., M. Caliendo and S. Thomsen (2005). “Mikroökonometrische Evaluation des Integrationserfolges”, *Beiträge zur Arbeitsmarkt- und Berufsforschung (BeitrAB)* 291, Nürnberg.

Hujer, R., S.L. Thomsen and C. Zeiss (2006). “The Effects of Short-Term Training Measures on the Individual Unemployment Duration in West Germany”. ZEW Discussion Paper No. 06-065.

Hummel, E., P. Jacobebbinghaus, A. Kohlmann, M. Oertel, C. Wübbeke, M. Ziegerer (2005). “Stichprobe der Integrierten Erwerbsbiographien, IEBS 1.0”, FDZ Datenreport 6/2005, Institut für Arbeitsmarkt- und Berufsmarktforschung (IAB), Nürnberg.

Imbens, G. (2000). "The Role of the Propensity Score in Estimating Dose-Response Functions" *Biometrika* 87, 706-710.

Jacobi, L. and J. Kluve (2006). "Before and After the Hartz Reforms: The Performance of Active Labour Market Policy in Germany". IZA Discussion Paper No. 2100.

Kluve, J. (2006). "The Effectiveness of European Active Labor Market Policy". IZA Discussion Paper No. 2018.

Kluve, J. and C.M. Schmidt (2002). "Can training and employment subsidies combat European unemployment?". *Economic Policy* 35, 409-448.

Kurtz, B. (2003). "Trainingsmaßnahmen - Was verbirgt sich dahinter?" IAB Werkstattbericht Nr. 8, Nürnberg.

Lechner, M. (1999). "Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification." Journal of Business and Economic 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 Labour Market Policies in Europe*, Heidelberg: Physica-Verlag.

Lechner, M. (2002). "Program heterogeneity and propensity score matching: An Application to the evaluation of active labor market policies." *Review of Economics and Statistics* 84, 205-220.

Lechner, M., R. Miquel, and C. Wunsch (2005a). "Long-Run Effects of Public Sector Sponsored Training in West Germany." IZA Discussion Paper No. 1443.

Lechner, M., R. Miquel, and C. Wunsch (2005b). "The Curse and Blessing of Training the Unemployed in a Changing Economy: The Case of East Germany after Unification." Discussion Paper, University of St. Gallen.

Lechner, M. and C. Wunsch (2006). "Active Labour Market Policy in East Germany: Waiting for the Economy to Take Off." IZA Discussion Paper No. 2363.

Martin, J.P. (2000). "What works among active labour market policies: Evidence from OECD countries' experiences". *OECD Economic Studies* No. 30, 79-113.

Martin, J.P. and D. Grubb (2001). "What works for whom: A review of OECD countries' experiences with active labour market policies". *Swedish Economic Policy Review* 8, 96.

OECD (2005). "Labour Market Programmes and Activation Strategies: Evaluating the Impacts." *OECD Employment Outlook*, Ch. 4.

Osikominu, A. (2005). "Eine Analyse der Teilnehmerselektion in die berufliche Weiterbildung auf Basis der Integrierten Erwerbsbiografien (IEB)." IAB Forschungsbericht Nr. 23/2005, Institut für Arbeitsmarkt- und Berufsmarktforschung (IAB), Nürnberg.

Rosenbaum, P.R. and D.B. Rubin (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70, 41-55.

Roy, A.D. (1951). "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3, 135-146.

Rubin, D.B. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66, 688-701.

Schneider, H., K. Brenke, D. Hess, L. Kaiser, J. Steinwedel, A. Uhlendorff (2006). "Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission - Modul 1b: Förderung beruflicher Weiterbildung und Transferleistungen." IZA Research Report No. 7, Bonn.

Sianesi, B. (2003). "Differential Effects of Swedish Active Labour Market Programmes for Unemployed Adults in the 1990s." Discussion Paper, Institute for Fiscal Studies, London.

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

Smith, J.A. and P. Todd (2005). "Rejoinder." *Journal of Econometrics* 125, 365-375.

Van den Berg, G.J. and B. van der Klaauw (2006). "Counseling and monitoring of unemployed workers: theory and evidence from a controlled social experiment." Forthcoming *International Economic Review*

Waller, M. (2007). “Do Reported End Dates of Treatments Matter for Evaluation Results? - An Investigation Based on the German Integrated Employment Biographies Sample”, FDZ Methodenreport No. 01-07, Institut für Arbeits- und Berufsforschung (IAB), Nürnberg.

Weber, A. and H. Hofer (2004). “Employment effects of early interventions on job search programs.” IZA Discussion Paper No. 1076.

Appendix

Table 1.2: Entries into Active Labor Market Programs in Germany from 2000 - 2004

	2000	2001	2002	2003	2004
Qualification schemes	1,153,720	1,069,409	1,457,047	1,502,166	1,548,439
– further/retraining	551,534	449,622	456,301	254,718	185,041
– short-term training	476,672	565,132	877,038	1,064,293	1,188,369
Employment subsidies	458,557	464,904	538,312	807,682	950,109
Placement and advisory services	601,281	742,065	947,098	1,460,170	2,566,780
Specific measures for young adults	445,823	457,724	447,265	388,810	408,168
Public Job Creation	314,291	246,084	219,626	193,999	170,107
Other	391,122	515,670	453,224	212,183	309,446
Total	3,364,794	3,495,856	4,062,572	4,565,010	5,953,049

Source: Bundesagentur für Arbeit, Arbeitsmarkt 2002b-2005b, own calculations.

Table 1.3: Sample Sizes

	East=0, Fem.=0	East=0, Fem.=1	East=1, Fem.=0	East=1, Fem.=1
Stratum 1 (0-3 Months)				
Waiting	29351	18409	15505	8538
STT	912	693	621	368
CFT	389	344	265	136
RT	263	262	86	
Stratum 2 (4-6 Months)				
Waiting	18529	12572	10270	6450
STT	547	409	339	286
CFT	251	194	218	143
Stratum 3 (7-12 Months)				
Waiting	10996	8421	5810	4277
STT	662	497	471	353
CFT	270	201	264	218
Aggregated Stratum 1 for PFT (0-12 Months)				
Waiting	25854	16060	12636	6614
STT	2120	1593	1432	1013
CFT	915	741	742	495
PFT	263	234	145	98

Table 1.4: Variable Definitions

Name	Definition
east	1 if place of residence is in East Germany (Berlin included), 0 otherwise
female	1 if female, 0 otherwise
agegroup	age in 6 groups
foreigner	1 if citizenship is not German, 0 otherwise
ethnicgerman	1 if ethnic German, i.e. returned settler from former German settlements, 0 otherwise
qualification	1 no degree, 2 vocational training degree, 3 university or technical college degree
schooling	1 no schooling degree, 2 Hauptschulabschluss or Mittlere Reife /Fachoberschule (degrees reached after completion of the 9th or 10th grade), 3 Fachhochschulreife or Abitur/Hochschulreife (degrees reached after completion of the 12th or 13th grade)
health	1 no health problems mentioned, 2 health problems, but considered without impact on placement, 3 health problems considered to have an impact on placement
pasthealth	same categories as health, but referring to the past two years before the beginning of the unemployment spell
disabled	1 if disabled, 0 otherwise
land	16 categories for the German Bundesländer
area	German Bundesländer aggregated into 6 categories. 1 SH, NI, HB, HH; 2 NW, 3 HE, RP, SL; 4 BY, BW; 5 MV, BB, BE; 6 SN, ST, TH
region	classification of the districts of residence according to local labor market conditions in 5 groups
family	1 missing, 2 living alone, 3 not married, but living together with at least one person, 4 single parent, 5 married
married	1 missing, 2 married, 3 not married
child	1 if at least one child, 0 otherwise
youngchild	1 if at least one child younger than 10 years, 0 otherwise

<continued on next page>

Variable Definitions <continued>

Name	Definition
occupation	occupation of last employment in 7 categories
industry	industry of last employment in 6 categories
occhange	1 missing, 2 if the person wishes to work in the same occupation as in the last employment, 3 otherwise
parttime	1 if the person worked less than full-time in the last employment, 0 otherwise
whitecollar	2 if the previous employment was a white-collar job, 3 if it was a blue-collar job, 1 missing
problemgroup	1 if participation in a program with a social work component within the last three years, 0 otherwise
onlyparttime	1 if information available that only part-time job is desired, 0 otherwise
endlastjob	2 termination of last occupation by employer, 3 by employee, 4 limited in time, 5 other and missing
quarter	quarter of the end of the last employment (from 1 to 9)
penalty	1 if the unemployed had a period of disqualification from benefits within the last three years, 0 otherwise
motivationlack	1 if within the last three years there is information, that the person did not appear regularly at the labor office, on lack of cooperation, availability or similar
pasttreatcancel	1 if abandonment of a program in the past according to the benefit data, 0 otherwise
pastincapacity	1 if incapapacity of work due to illness, parental leave, cure or therapy within the last three years
proposals	number of placement proposals divided by the days since the beginning of the unemployment spell and the start date of the spell from which the information is taken
dapp	1 if employed as apprentice within the last three years before the beginning of the unemployment spell, 0 otherwise

<continued on next page>

Variable Definitions <continued>

Name	Definition
countemp, countub, countua, countsp, countoos, countcontact	number of days within the last three years before the beginning of unemployment spent in regular employment, receiving unemployment benefits, unemployment assistance, subsistance payment, out of sample, in contact with the labor office, respectively
demp6, demp12, demp24, demp6_12, demp12_24	1 if in regular employment 6, 12, 24, 6 and 12 and 12 and 24 months, respectively, before the beginning of the unemployment spell
waged	daily wage in the last job(s) before the beginning of the unemployment spell
ddssec, ddcens, ddmarg	dummies if waged is censored: ddsec is 1 if earnings are within the social security thresholds, ddcens is 1 if earnings are above the social security threshold, ddmarg is 1 if earnings are below the social security threshold
lnwage, lnwagedsq	log(waged) and log(waged) squared interacted with ddssec
wage	total wage in the last year before the beginning of the unemployment spell
dssec, dcens, dmarg	censoring dummies referring to wage (see above)
lnwage, lnwagesq	log(wage) and log(wage) squared interacted with dssec
ur_yb, ur_qb, ur_qb3, ur_qb6, ur_qb12, ur_qb24	unemployment rate in the individual's home district in the calendar year before the beginning of unemployment, in the last month of the quarter before the beginning of unemployment, and in the last month of the quarter before the beginning of the stratum, respectively

Note: If not mentioned otherwise, variables are defined relative to the beginning of the time window of elapsed unemployment duration.

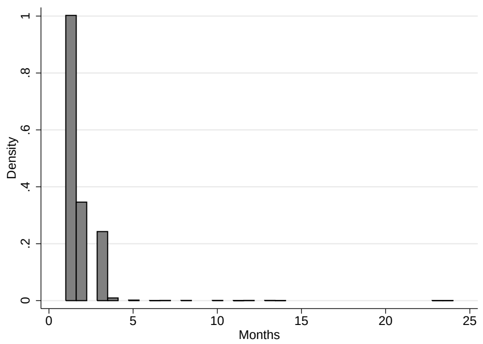
Table 1.5: Descriptive Statistics for Selected Variables

Variable	Mean	SD	Min	Median	Max
East=0, Female=0					
age	36.90053	7.718731	25	36	53
foreigner	.1736603	.378823	0	0	1
schooling1	.1237411	.3292911	0	0	1
schooling2	.7414522	.4378433	0	1	1
schooling3	.1348067	.3415223	0	0	1
qualification1	.3572672	.4792019	0	0	1
qualification2	.5908243	.4916894	0	1	1
qualification3	.0519085	.2218457	0	0	1
countemp	801.0523	287.0826	86	873	1096
East=0, Female=1					
age	37.94112	7.889209	25	37	53
foreigner	.1079787	.3103611	0	0	1
schooling1	.0723604	.25909	0	0	1
schooling2	.7323496	.4427451	0	1	1
schooling3	.19529	.3964335	0	0	1
qualification1	.3214931	.467061	0	0	1
qualification2	.6059022	.488668	0	1	1
qualification3	.0726047	.2594928	0	0	1
countemp	776.9238	312.8626	86	845	1096
East=1, Female=0					
age	38.3997	7.844472	25	38	53
foreigner	.0340613	.1813919	0	0	1
schooling1	.0546255	.2272544	0	0	1
schooling2	.8529804	.3541357	0	1	1
schooling3	.0923941	.2895899	0	0	1
qualification1	.1134218	.3171169	0	0	1
qualification2	.8383827	.3681101	0	1	1
qualification3	.0481956	.2141855	0	0	1
countemp	829.9092	272.4151	86	914	1096
East=1, Female=1					
age	39.21149	7.836369	25	39	53
foreigner	.0256919	.1582228	0	0	1
schooling1	.038329	.1919993	0	0	1
schooling2	.8185901	.3853776	0	1	1
schooling3	.1430809	.3501737	0	0	1
qualification1	.1095561	.312352	0	0	1
qualification2	.8129504	.3899717	0	1	1
qualification3	.0774935	.2673868	0	0	1
countemp	748.3454	309.1351	86	762	1096

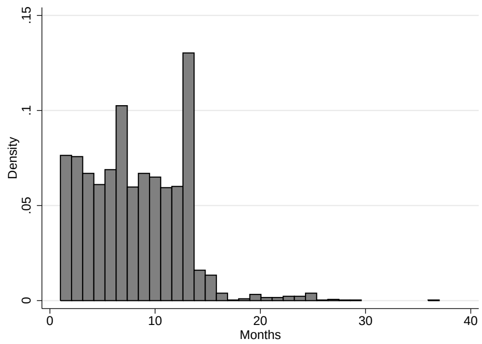
Table 1.6: Program Duration of Analyzed Participations in Months

	N	Min	Mean	Median	75th Perc.	95th Perc.	Max
STT	6158	1	1.6	1	2	3	24
CFT	2893	1	8.2	8	12	14	37
PFT	740	1	6.5	7	8	12	26
RT	1066	1	23.5	24	26	37	53
Total	10857	1	5.8	2	8	24	53

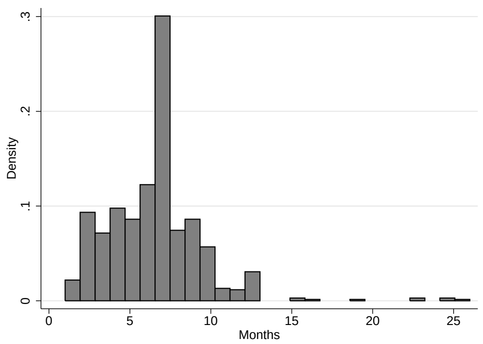
Figure 1.16: Densities of Program Duration



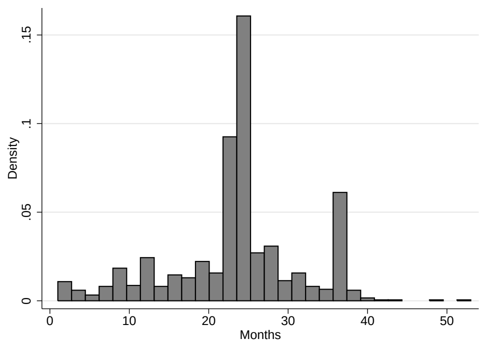
Short Term Training



Classroom Further Training



Practical Further Training



Retraining

Table 1.7: Cumulated Treatment Effects after 24 Months

	Men			Women		
	Stratum 1	Stratum 2	Stratum 3	Stratum 1	Stratum 2	Stratum 3
	West Germany					
STT vs Waiting	-0.054 (0.323)	0.573 (0.399)	0.903 (0.357)**	0.314 (0.368)	1.767 (0.532)***	2.122 (0.429)***
STT vs CFT	2.937 (0.638)***	1.750 (0.716)**	2.597 (0.573)***	1.878 (0.683)***	2.000 (0.877)**	1.545 (0.849)*
STT vs PFT	2.536 (0.518)***	.	.	2.700 (0.643)***	.	.
STT vs RT	9.173 (0.391)***	.	.	8.983 (0.429)***	.	.
CFT vs Waiting	-2.518 (0.488)***	-1.827 (0.535)***	0.159 (0.446)	-1.134 (0.446)**	-0.776 (0.608)	0.448 (0.687)
CFT vs STT	-1.407 (0.693)**	-1.604 (0.780)**	-0.189 (0.684)	-2.090 (0.665)***	-2.761 (1.069)***	-0.571 (0.821)
CFT vs PFT	0.752 (0.822)	.	.	0.699 (1.003)	.	.
CFT vs RT	7.205 (0.471)***	.	.	7.519 (0.487)***	.	.
PFT vs Waiting	-0.817 (0.588)	.	.	0.593 (0.553)	.	.
PFT vs STT	1.208 (0.520)**	.	.	1.530 (0.657)**	.	.
PFT vs CFT	1.845 (0.596)***	.	.	3.788 (0.727)***	.	.
RT vs Waiting	-7.800 (0.378)***	.	.	-6.601 (0.351)***	.	.
RT vs STT	-6.781 (0.912)***	.	.	-9.229 (1.406)***	.	.
RT vs CFT	-5.780 (0.819)***	.	.	-9.198 (1.174)***	.	.
	East Germany					
STT vs Waiting	0.484 (0.365)	0.270 (0.517)	0.947 (0.448)**	0.003 (0.453)	0.658 (0.547)	0.123 (0.508)
STT vs CFT	2.020 (0.795)**	2.104 (0.670)***	3.150 (0.613)***	1.857 (1.111)*	1.643 (1.053)	2.324 (0.594)***
STT vs PFT	3.922 (0.627)***	.	.	1.554 (0.873)*	.	.
STT vs RT	7.607 (0.705)***	.	.	.	.	.
CFT vs Waiting	-0.528 (0.497)	-1.744 (0.540)***	-1.075 (0.665)	-1.487 (0.784)*	-0.690 (0.609)	-0.793 (0.390)**
CFT vs STT	-1.170 (0.843)	-1.546 (1.027)	-2.314 (0.720)***	-1.603 (1.257)	-0.906 (1.150)	-0.607 (0.776)
CFT vs PFT	1.619 (0.579)***	.	.	-0.133 (0.875)	.	.
CFT vs RT	6.140 (0.607)***	.	.	.	.	.
PFT vs Waiting	-0.573 (0.668)	.	.	0.478 (0.996)	.	.
PFT vs STT	0.567 (0.843)	.	.	-0.047 (0.871)	.	.
PFT vs CFT	2.123 (0.822)***	.	.	1.872 (0.792)**	.	.
RT vs Waiting	-6.623 (0.446)***	.	.	-6.623 (0.446)***	.	.
RT vs STT	-7.559 (1.077)***	.	.	-7.559 (1.077)***	.	.
RT vs CFT	-4.901 (0.800)***	.	.	-4.901 (0.800)***	.	.

*** = statistically significant at 1 %, ** = at 5 %, * = at 10 %, bootstrapped standard errors

Table 1.8: Effect Heterogeneity, STT vs. Waiting

East=0, Sex=0			
	Stratum 1	Stratum 2	Stratum 3
agegroup6	-0.107 (0.051)**		
qualification1			-0.083 (0.031)***
_cons	0.015 (0.016)		0.082 (0.022)***
N	908	547	662
comparable uncorr. ATT	0.008	0.033	0.047

East=0, Sex=1			
	Stratum 1	Stratum 2	Stratum 3
agegroup5			-0.139 (0.058)**
agegroup6			-0.124 (0.063)*
qualification1	-0.059 (0.159)		
mismqualification1	-0.079 (0.152)		
_cons	0.045 (0.053)		0.151 (0.026)***
N	693	409	492
corrected ATT with CI	0.026 [-0.006, 0.059]		
comparable uncorr. ATT	0.028	0.090	0.113

East=1, Sex=0			
	Stratum 1	Stratum 2	Stratum 3
agegroup6	-0.208 (0.057)***	-0.282 (0.140)**	-0.133 (0.132)
family5		0.127 (0.075)*	0.160 (0.053)***
mismagegroup6		0.077 (0.124)	-0.041 (0.113)
mismfamily5		-0.014 (0.061)	-0.055 (0.038)
mismqualification3			0.117 (0.068)*
_cons	0.045 (0.017)***	-0.001 (0.048)	-0.005 (0.035)
N	619	332	470
corrected ATT with CI		0.026 [-0.024, 0.075]	0.055 [0.013, 0.097]
comparable uncorr. ATT	0.027	0.027	0.053

East=1, Sex=1			
	Stratum 1	Stratum 2	Stratum 3
agegroup6			-0.200 (0.105)*
mismagegroup1			-0.091 (0.043)**
mismagegroup6			0.020 (0.087)
_cons			0.045 (0.030)
N	368	286	350
corrected ATT with CI			0.018 [-0.031, 0.067]
comparable uncorr. ATT	0.011	0.038	0.014

Empty entries indicate that the regressor is not included in the final specification due to lack of significance or due to a multicollinearity problem. Specifications without significant covariates are not reported (columns in which no coefficients are reported).
*** = statistically significant at 1 %, ** = at 5 %, * = at 10 %, bootstrapped standard errors

Table 1.9: Effect Heterogeneity, CFT vs. Waiting

East=0, Sex=0			
	Stratum 1	Stratum 2	Stratum 3
occupation1		0.178 (0.094)*	
_cons		-0.003 (0.033)	
N	385	249	265
comparable uncorr. ATT	0.023	0.021	0.095

East=0, Sex=1			
	Stratum 1	Stratum 2	Stratum 3
agegroup4			-0.209 (0.110)*
agegroup5			-0.166 (0.245)
agegroup6			-0.210 (0.243)
qualification1		-0.168 (0.095)*	
mismagegroup5			0.025 (0.201)
mismagegroup6			0.025 (0.216)
_cons		0.154 (0.044)**	0.187 (0.073)**
N	344	192	199
corrected ATT with CI			0.099 [0.016, 0.181]
comparable uncorr. ATT	0.045	0.121	0.098

East=1, Sex=0			
	Stratum 1	Stratum 2	Stratum 3
agegroup6		-0.300 (0.093)***	-0.219 (0.121)*
mismagegroup6			-0.008 (0.069)
_cons		0.085 (0.033)**	0.078 (0.045)*
N	265	218	252
corrected ATT with CI			0.052 [-0.028, 0.131]
comparable uncorr. ATT	0.096	0.060	0.046

East=1, Sex=1			
	Stratum 1	Stratum 2	Stratum 3
agegroup6			-0.187 (0.101)*
_cons			0.035 (0.032)
N	136	143	217
comparable uncorr. ATT	0.014	0.045	0.020

Empty entries indicate that the regressor is not included in the final specification due to lack of significance or due to a multicollinearity problem. Specifications without significant covariates are not reported (columns in which no coefficients are reported).

*** = statistically significant at 1 %, ** = at 5 %, * = at 10 %, bootstrapped standard errors

Table 1.10: Effect Heterogeneity, PFT vs. Waiting

East=0, Sex=0	
N	258
comparable uncorr. ATT	0.023
East=0, Sex=1	
agegroup6	-0.279 (0.111)**
_cons	0.121 (0.031)***
N	233
comparable uncorr. ATT	0.094
East=1, Sex=0	
agegroup6	-0.196 (0.244)
mismagegroup6	-0.009 (0.237)
_cons	0.072 (0.057)
N	145
corrected ATT with CI	0.045 [-0.030, 0.119]
comparable uncorr. ATT	0.045
East=1, Sex=1	
N	97
comparable uncorr. ATT	0.068

Empty entries indicate that the regressor is not included in the final specification due to lack of significance or due to a multicollinearity problem. Specifications without significant covariates are not reported (columns in which no coefficients are reported).

*** = statistically significant at 1 %, ** = at 5 %, * = at 10 %, bootstrapped standard errors

Chapter 2

On the Importance of Correcting Reported End Dates of Labor Market Programs

2.1 Introduction

Large administrative data sets are becoming increasingly available for empirical research. Therefore the quality of crucial variables of process generated data is of growing interest to researchers. This paper investigates a sensitive part of a new German data set: the reported end dates of labor market programs in the Integrated Employment Biographies Sample (IEBS). The IEBS covers about 1.4 million individuals and rich, daily information on employment, job search, transfer payments and active labor market programs. It has therefore become a very important data set for microeconometric labor market policy evaluation in Germany. It is the basis for the ongoing government conducted evaluation of recent years' labor market reforms. The data are considered highly reliable, but end dates of labor market measures are an exception to this: a considerable part of reported end dates is later than the end of actual participation. The impact of this measurement error on evaluation results is analyzed in this paper for the example of further training.

Because measurement error in end dates may influence evaluation results through several channels, it is difficult to predict ex ante how results will be affected. But the IEBS has the advantage that due to its special feature of including data from different administrative processes, it is possible to correct almost all relevant end dates of further training programs. This study introduces three approaches to deal with error-prone end dates, a “naive” approach, a standard approach and a mechanism to explicitly correct end dates. These three approaches are used to study through which channels and to what degree upward measurement error in end dates influences results. For this objective employment rates in a framework with a simple treatment variable (as typical for matching studies) and a duration model with time-varying treatment variables are estimated. A setting with typical features of evaluation studies like employment as the outcome of interest, an evaluation period starting with the start of the program, and the consideration of program effects as opposed to pure threat effects is chosen. There are two aims of this exercise. The first is to learn more about the quality of a sensitive variable in the IEBS and to gain knowledge on how to handle the problem in future studies using the IEBS. The second is to get insights on how measurement error in end dates of treatments influences evaluation results in empirical studies in general. This might be helpful for studies using other administrative data sets, which are supposed to suffer from measurement errors in end dates that cannot be corrected. To the best of my knowledge, there is no guidance in the literature on this problem.

The remainder of this paper is structured as follows: section two presents the data set and discusses the relevance of treatment end dates and why they are a sensitive part of the IEBS. Section three discusses possible corrections and introduces three procedures to deal with error-prone end dates of further training programs, and presents their impact on the sample used for the empirical analysis. Sections four and five investigate the sensitivity of evaluation results in two different frameworks. Section six concludes and links the conclusions of this paper to the validity of existing studies on further training using the IEBS.

2.2 End Dates of Labor Market Programs in the IEBS

2.2.1 Data Set

The IEBS consists of a 2.2% random sample of individuals data drawn from the universe of data records collected in four different administrative processes: the IAB Employment History (*Beschäftigten-Historik*), the IAB Benefit Recipient History (*Leistungsempfänger-Historik*), the Data on Job Search Originating from the Applicants Pool Database (*Bewerberangebot*), and the Particpants-in-measures Data (*Maßnahme-Teilnehmer-Gesamtdatenbank*).[1] This study uses version 2.05 of the IEBS and focusses on unemployment periods beginning in between February 2000 and January 2002.[2] The data contains detailed daily information on employment subject to social security contributions, receipt of transfer payments during unemployment, job search, and participation in different programs of active labor market policy. Thus, the IEBS is particularly useful to evaluate different parts of German active labor market policies in detail. It is the data set that is mainly used for the evaluations of the so-called *Hartz-Reformen*, several major labor market reforms of

[1] For detailed information on the IEBS see Hummel et al. (2005) and Bender et al. (2005). Information in English can be found on the website of the Research Data Center (FDZ) of the Federal Employment Office (BA) (http://fdz.iab.de/en), in particular the documentation "The German Integrated Employment Biographies Sample IEBS" by P. Jacobebbinghaus and S. Seth. The website also describes the conditions under which researchers may use the IEBS and the process to get the permission.

[2] The data used here has been supplemented with some additional information compared to the standard version.

recent years.[3] In addition, the IEBS has already been used for several further evaluation studies, for example Biewen et al. (2006), Biewen et. al. (2007), Boockmann et al. (2007), Jaenichen and Stephan (2007), Lechner and Wunsch (2006), Pfeiffer and Winterhager (2006), Schneider and Uhlendorff (2006). Certainly further studies will follow as the data set is unique in Germany concerning its largeness and richness in detailed information on employment biographies and as it will be updated in the future to always include recent years.

The first of the four administrative data sources, the IAB Employment History, consists of social insurance register data for employees subject to contributions to the public social security system. It covers the time period from 1990 to 2004. The main feature of these data is detailed daily information on the employment status of each recorded individual. In evaluation studies this information can be used to account for the labor market history of individuals as well as to measure employment outcomes. For each employment spell, in addition to start and end dates, data from the Employment History contains information on personal as well as job and firm characteristics such as wage, industry or occupation.

The IAB Benefit Recipient History, the second data source, includes daily spells of unemployment benefit, unemployment assistance and subsistence allowance payments the individuals received between January 1990 and June 2004. In addition to the sort of the payment and the start and end dates of periods of transfer receipt the spells contain further information like sanctions, periods of disqualification from benefit receipt and personal characteristics. The Benefit Recipient History is important as it provides information on the periods during which individuals were out of employment and therefore not covered by the Employment History.

The third data source included in the IEBS is the so-called Data on Job Search Originating from the Applicants Pool Database, which contains rich information on individuals searching for jobs covering the period January 1997 to June 2004. The spells include detailed information concerning job search, regional information and personal characteristics, in particular on educational qualifications, nationality, and marital status. They also provide information on whether the applicant wishes to change occupation, how many job proposals he or she already got, and about health problems that might influence employment chances.

[3] Compare the report of the federal government (Bericht 2006 der Bundesregierung zur Wirksamkeit moderner Dienstleistungen am Arbeitsmarkt) for an overview of the results.

The Participants-in-measures Data, the fourth data source, contains diverse information on participation in public sector sponsored labor market programs for example training programs, job-creation measures, integration subsidies, business start-up allowances covering the period January 2000 to July 2004. Similar to the other sources, information comes in the form of spells indicating the start and end dates at the daily level, the type of the program as well as additional information on the program such as the planned end date, whether the participant entered the program with a delay, and whether the program was successfully completed.

2.2.2 Relevance of End Dates

There exist several studies on measurement error in the treatment variable. Molinari (2005) develops limits for treatment effects in the case that the treatment variable has missings in survey data. Battistin and Sianesi (2006) characterize the bias if treatment status is mismeasured and provide bounds. Lewbel (2004) develops GMM estimators for the scenario that the treatment variable is measured with error and an instrument that influences the probability of treatment but is conditional independent of the misclassification probabilities and the average treatment effect is available. For the case where no such instrument is available bounds are developed. The problem analyzed in this paper is different in two respects. First, the problem itself is more complicated, because it is not the treatment indicator which is mismeasured but the program end dates. Measurement error in end dates can affect the treatment indicator but it may also affect the results through other channels as discussed below. But second, using the IEBS data it is possible to correct the end dates. Therefore the approach of this paper is to develop procedures to correct the end dates and then to analyze through which channels and to what extent wrong end dates influence results.

Through which channels upward measurement error of end dates potentially influences employment effects depends on the evaluation design. Using descriptive analysis of employment rates or matching with a simple treatment variable, program end dates have no direct effect on the results but may bias them indirectly through outcome measurement and through the treatment indicator. First, if the outcome is measured as regular employment or non-employment (including every other status including program participation), too late end dates of programs lead to a contradiction: the researcher observes program spells and regular employment spells in parallel for some time. A decision whether to count this time as employment or program participation (and thus non-employment) is necessary and will influence employment rates and treatment effects. Second, end dates define the actual length of program participation, which can be relevant for the decision if a program has been attended long enough to be counted for evaluation. Too late end dates can lead to measurement error in the treatment indicator: it may indicate participation, although it should indicate non-participation, as in reality the participant did not attend long enough.

Measurement error in the end dates influences the results more directly in estimation designs in which it is of importance if a participant is in a program at a certain point in time or in which it is relevant whether a program has been completed or not, thus in frameworks with a time-varying treatment variable. An example for this is a duration analysis approach in which attending an uncompleted program and having attended a program in the past are considered separately.[4] In conclusion, there exist different channels through which measurement errors in program end dates may bias evaluation results and it is therefore difficult to predict the direction and magnitude of a potential bias.

2.2.3 Error-Proneness of End Dates for Labor Market Programs

The reliability of the IEBS data was checked carefully by Bender et al. (2004, 2005). Concerning calendar dates, their conclusion is that start and end dates in the employment and benefit data are very highly reliable. Calendar dates seem to be less reliable in the Particpants-in-measures Data and the Data on Job Search Originating from the Applicants Pool.[5] Bender et al. (2005) point out that the end dates of further training and retraining programs are error-prone. It is possible that end dates of other programs in the Participants-in-measures Data like short-term training or job creation measures suffer from similar measurement error that leads to comparable biases. But note that data constellations pointing at wrong end dates vary for different programs. Job-creation measures for example are expected

[4]Measuring treatments by dose (for instance using days of treatment as a treatment variable) is another framework in which the end date is of direct importance, but only if dose is measured in realized duration and not in planned duration.

[5]For job search data the measurement error seems to be quite severe, but it is possible to circumvent this problem by defining the labor market status using benefit and employment data.

to have an employment spell in parallel, whereas this is implausible for further training programs as discussed in section 2.3.1. The analysis of passive policies like the receipt of unemployment benefits and unemployment assistance is not supposed to suffer from similar biases, because the information on benefit receipt originates from the IAB Benefit Recipient History and not from the Participants-in-measures Data.

There are two aspects which determine the reliability of administrative data. One is how the information is registered during the administrative process itself. The other is what rules the providers of the data use to define which piece of information of the administrative data bases will finally appear in the scientific data set.[6] The reason why end dates for labor market programs are a sensitive part of the IEBS seems to be the concurrence of at least two problems: First, the correct reporting of end dates of individual program participation is not always directly relevant for payments (Bernhard et al. 2006, p. 5), mainly because for part of the measures the employment office pays per measure and not per person. This is in contrast to for instance the dates of benefit spells, which are directly and even technically linked to payments and thus much more reliable in the data. Second, the end of program participation often changes after the date is first registered. This can be due to drop-out of the program, non-attendance, change of course or shift of the course. If then the registered date is not corrected or if the correction does not reach the data set provided to the researcher, the end date of participation in the IEBS will be incorrect.[7]

[6]Jaenichen et al. (2005) analyze inconsistencies of the participation data that are related to the end date problem. One of their conclusions is that both aspects are relevant, but the problems in the registering of the data themselves might be the major problem. Kruppe and Oertel (2003) provide detailed information on aspects of the data creation. The rules to create the Participants-in-Measure Data have been changed between the IEB versions 3 and 4 (see Waller (2007)).

[7]Start dates are more reliable than end dates, probably because drop-outs are irrelevant and because they lie in the nearer future, so that fewer changes occur. In case of non-attendance start and end dates are per definition incorrect. In this case the correction of the end date leads to non-participation in a program and thereby also to a correction of the start date.

2.3 Empirical Approach: The Example of Further Training

Concerning data checks and corrections, the IEBS has a great advantage: the fact that it includes data of different administrative processes can be exploited to check plausibility and correct implausible information. It is thus possible to correct end dates and to analyze if and how errors in treatment end dates lead to biased estimation results. The impact of upward measurement error in treatment end dates will be investigated in this paper for the example of further training programs. Further training is an important part of active labor market policy in Germany.[8] It has already been evaluated using the IEBS several times (see e.g. Biewen et al. (2007), Lechner and Wunsch (2006), Schneider and Uhlendorff (2006) and Schneider et al. (2007)). For investigating the error-proneness of end dates, further training has the advantage that participants receive subsistence allowance while they are in the program and this information helps to correct end dates.

2.3.1 Plausibility Checks

This section discusses which information in the data indicates a wrong end date. A constellation that is a clear contradiction is a regular employment spell that starts while the participant of further training is still attending the program.[9] This is a contradiction (Bernhard (2006), p. 25), because once a person is regularly employed, he or she cannot continue the program.[10] As calendar dates in the IAB Employment History are much more reliable than in the Particpants-in-measures Data, the employment information indicates that the correct end date of program participation is at the latest one day before regular employment starts. A second major possibility for corrections is provided by subsistence allowance (*Unterhaltsgeld*) spells. Subsistence allowance are payments of the labor agency to cover living costs of the participants of further training programs. They are a subsidy to unemployment benefit or unemployment assistance for the time of the program. With

[8]See Bundesagentur für Arbeit (2006).

[9]In this study regular employment is defined as non-minor unsubsidized employment on the first labor market with a minimum length of two weeks.

[10]A participant of an active labor market program to whom a job is offered must decide whether to take the job *or* to continue the program. Usually participants are encouraged to take the job because of the rule of priority to job placement (*Vorrang der Vermittlung*), SGB III § 4.

very few exceptions all participants of further training receive subsistence allowance for the complete time of the program, a fact that proves true in the data. Dates of subsistence allowance spells are very reliable. Thus, if a subsistence allowance spell that has started in parallel to a further training spell finishes before the training spell, one can conclude that the end date of the program spell is wrong. Furthermore two variables of the Participants-in-measures Data can be used: a variable indicating that someone did never attend (*Maßnahmeerfolg: Nichtantritt)* and a variable indicating the date a participant notifies a dropout *(FbW Abmeldedatum)*. These two variables have many missings, but used with a lot of caution they can help to correct end dates in some cases.

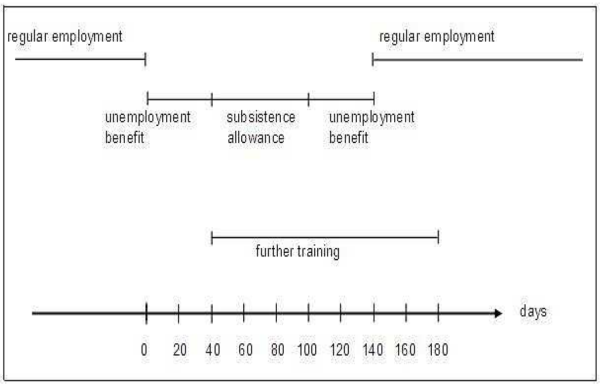
There is other information in the data one might be tempted to use, but which would lead to false corrections in some cases. First, one should not use the length of program spells. The law provides rules for the length of certain programs, but despite of this in practice there exist - though rarely - much longer programs. Therefore one should not change end dates in the data just because a spell is surprisingly long. Second, while regular employment parallel to training programs is a contradiction, employment of a few hours only is possible (Bernhard et al. 2006, p. 24) and is no hint for a wrong end date.[11]

2.3.2 Three Procedures to Deal with Error-prone End Dates

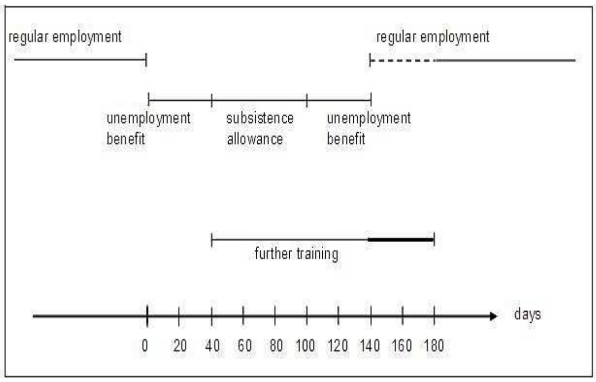
This section introduces three ways to deal with the error-prone end dates. They are illustrated using the fictitious example shown in the upper left diagram of figure 2.1: the individual in the example becomes unemployed (out of regular employment) at day zero and receives unemployment benefit. He or she starts a further training program at day 40 of his or her unemployment period and receives subsistence allowance in parallel. The receipt of subsistence allowance ends on day 100 and the individual takes up regular employment on day 140. The reported end date of the further training spell is on day 180, but according to the argumentation of the last section the correct end of program participation would be day 100, because this is when the subsistence allowance spell ends.

[11]For rarer data constellations that indicate or do not indicate a wrong end date and for some aspects to be cautious about when relying on subsistence allowance spells see Waller (2007).

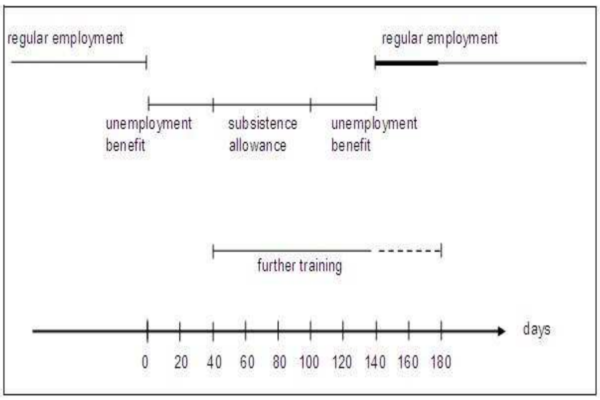
Figure 2.1: Illustration of the Three Procedures using a Fictitious Example



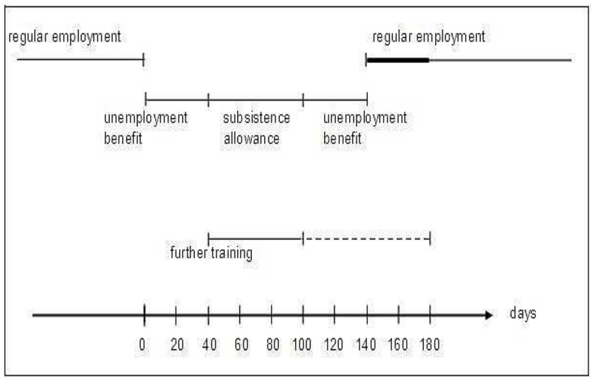
The Data



Procedure 1



Procedure 2



Procedure 3

Procedure 1

The underlying idea of procedure 1 is that program participation is the most important information in a data set mainly created for evaluation studies. Therefore participation spells are taken as they are in the data. But if a participation spell conflicts with a regular employment spell for some time, the researcher is forced to take a decision.[12] The rule of procedure 1 is to always give priority to the program spell. There are two important situations where program and regular employment spells may conflict and the rule applies. The first is shown in the upper right diagram of figure 2.1: in between day 140 and day 180 the regular employment and the further training spell are in parallel in the data. To measure the outcome, the status of the person must be defined (employed or not employed, e.g. in a program).

[12] All procedures assume that it is impossible to start regular employment but continue a further training program, as argued above. If one assumed that this was possible, one would explicitly allow for this situation when generating treatment and outcome variables. Results should then be similar to the results of procedure 2.

The rule of procedure 1 gives priority to the program spell, thus the status of the person is defined as attending a program (and thus not employed).

The second situation where the rule applies is the same data constellation (regular employment and a continuing program spell) but not for a program within the unemployment spell to be evaluated but for a former program in a former unemployment period. The rule of procedure 1 gives priority to the program spell and thus the concerned days until the program spell ends are not counted as employment. If an evaluation study focusses only on unemployment periods before which the individuals have been employed a certain amount of days, the rule of procedure 1 will in some cases prevent an unemployment spell to be used. This is because the criterion of sufficient pre-employment is not fulfilled when not counting the concerned days as employment.

In short, the basic assumption of Procedure 1 is that program spells are more reliable than employment spells. To get the results of procedure 1 this assumption does not necessarily have to be implemented explicitly. The results of procedure 1 will just appear when one first checks if there is program participation at a certain day and second if there is employment, without having checked for contradictions before. Procedure 1 is called the "naive" procedure, because a close look at examples in the data reveals that dates of employment spells are more reliable than end dates of program spells.[13]

Procedure 2

The basic assumption underlying procedure 2 is that regular employment spells are more valid than programm spells. The rationale is that employment dates in the IEBS are very reliable, because the length of the spells is directly relevant for pension payment. Thus in procedure 2 regular employment spells are given priority in case they conflict with program spells. The rule to give priority to employment information is always applied when the researcher is forced to take a decision, thus in the two situations described above (measurement of the outcome as well as measurement of labor market status before the unemployment period in

[13]One hint for this is for instance, that subsistence allowance and employment spells almost never conflict, whereas it occurs quite often that the end of program spells does not fit to the end of subsistence allowance. Another hint are examples in which several annual employment spells follow each other in a regular way, while the program spell is still continuing in parallel. Furthermore, Bernhard et al. (2006, p. 46) advise the researcher to give priority to employment spells.

focus). The lower left diagram of figure 2.1 illustrates that procedure 2 gives priority to the employment spell for the time employment and program spells are parallel.

But note that no ex ante correction of the program end dates is implemented. Procedure 2 involves no explicit data correction. Only when the employment status of the individual on a certain day is to be defined, the researcher in the first step checks for regular employment and if there is none he or she secondly checks for program participation. Thus in case of conflict regular employment will be counted. Procedure 1 and procedure 2 have in common that there is no explicit data correction at the beginning of data preparation. They just use one rule whenever a conflict appears. And the rule of procedure 2 (priority to regular employment spells) is the contrary of the rule of procedure 1 (priority to program spells). Because procedure 2 does not implement an explicit data correction, but uses its rule only when the employment status must be defined, the wrong end dates themselves are not changed in the data and when for instance the length of a program is calculated the wrong end date is used. Procedure 2 is called the standard procedure, because it seems to be the best choice if one does not want to implement an explicit correction mechanism.

Procedure 3

Procedure 3 uses the same rule for the definition of the employment status as procedure 2, but in addition a mechanism to correct end dates of further training programs is implemented at the beginning of the data preparation: First, an end date of a further training spell is changed if the subsistence allowance spell ends before the further training spell. The end date of the program is set to the last day of receipt of subsistence allowance. This is illustrated in the lower right diagram of figure 2.1: the program end date is set from the date of day 180 to the date of day 100. Second, in the rare cases where there is no subsistence allowance spell, and only then, other correction possibilities like the variable indicating that someone did never attend (*Maßnahmeerfolg: Nichtantritt)* or the variable indicating the date a participant notifies a dropout *(FbW Abmeldedatum)* are used if they are filled and indicate that a correction is necessary. Third, if a regular employment spell starts before the end of the program, the day before the start of employment is set as the corrected program end date (if no stronger correction has been implemented before).[14]

[14]Waller (2007) implements a fourth procedure because procedure 3 could be biased as using employment spells for changing end dates (and thus possibly not counting the program anymore, because participation becomes too short) leads to counting programs of those who find employment

When implementing the corrections, several technical particularities of the IEBS as well as some special regulations (in particular concerning programs partly sponsored through the European Social Funds) must be taken into account. See Waller (2007) for more details on the corrections and a description of how the corrections may be implemented. Procedure 3 relies on all assumptions of the corrections, in particular on the reliability of subsistence allowance spells. In conclusion, procedure 3 is the only procedures presented involving explicitly changing information in the data, that means setting some end dates to different dates at the beginning of data preparation.

2.3.3 Treatment and Sample

Further training (FT) programs are defined in this study as those measures that train profession skills and last typically several months up to a year. Other training programs, like the longer retraining (*Umschulung)*, which leads to a new degree within the German vocational training system or short-term training (*Trainingsmaßnahme)* are not analyzed here.[15] The effect of participating in a program (as opposed to a possible threat effect of the announcement to be assigned to a program) shall be evaluated and therefore programs are counted only if the unemployed has participated a minimal amount of days. The limit has been set considering program aims and the distribution of planned program durations to 28 days.[16]

The sample chosen for the empirical analysis consists of unemployment periods of women (aged between 25 and 53 years) living in West Germany which start in between February 2000 and the end of January 2002 after continuous regular employment of at least three months. Entering unemployment is defined as quitting regular employment and subsequently being in contact with the labor agency (not necessarily immediately) either through benefit receipt, program participation or a job search spell.[17] Only program participation out of such an unemployment period

less often. It turns out that this bias is negligible.

[15]See Waller (2007) for an investigation of the end dates of retraining.

[16]If one person has several participation spells within one unemployment spell, the spells are connected if there are at most 14 days in between two spells. If a person participated in several programs within one unemployment period with an interruption of more than two weeks, the first program is evaluated.

[17]Note that this implies that the same individual may appear more than once in the evaluation sample. Approximately ten percent of the individuals are represented by more than one unemployment spell according to the above definition.

is counted.

It turns out that it is possible to check almost all end dates of relevant treatments using procedure 3. For the sample described above only 2.5% of the end dates of FT programs in focus can neither be confirmed nor corrected. For some of them there are hints in the data that the end date is correct or not, but the information seems not reliable enough to be used for corrections.

Table 2.1: Programs in the Different Procedures

Procedure	1	2	3
Valid unemployment spells	20165	20439	20435
Valid FT treatments	879	896	884
Average program duration	214.66	215.05	202.54
... for programs valid in all procedures	214.67	212.90	200.87
...and until validated employment starts only	214.67	202.84	200.04

Table 2.1 gives the number of valid unemployment spells, valid FT treatments and the duration of the programs for each procedure. There are less valid employment spells (and as a consequence also less valid treatments) using procedure 1 due to the condition of entering unemployment out of three months of employment. This condition is met a little less often in procedure 1, because also participation spells of earlier programs dominate earlier employment spells (compare section 2.3.2). Less programs are valid in procedure 3 than in procedure 2, because due to the corrections more program spells are affected by the minimal attendance criterion of 28 days. If the durations of the program spells are compared considering only those that are valid in every procedure (second last line), the average length is considerably shortest for procedure 3, where the end dates are explicitly corrected. Considering the average length of those programs valid in the respective procedure, but not necessarily in all procedures (the row in the middle of the table), sample differences make this picture less clear. The last line in table 2.1 reflects the outcome measurement: it gives the length of the program spell, but cut if a validated regular employment spell starts. This length is on average only a bit longer for procedure 2 than for procedure 3, but considerably longer for procedure 1. This is because for 7.3% of the treatments valid in all procedures regular employment starts on average 128 days before the end of the original program spell.

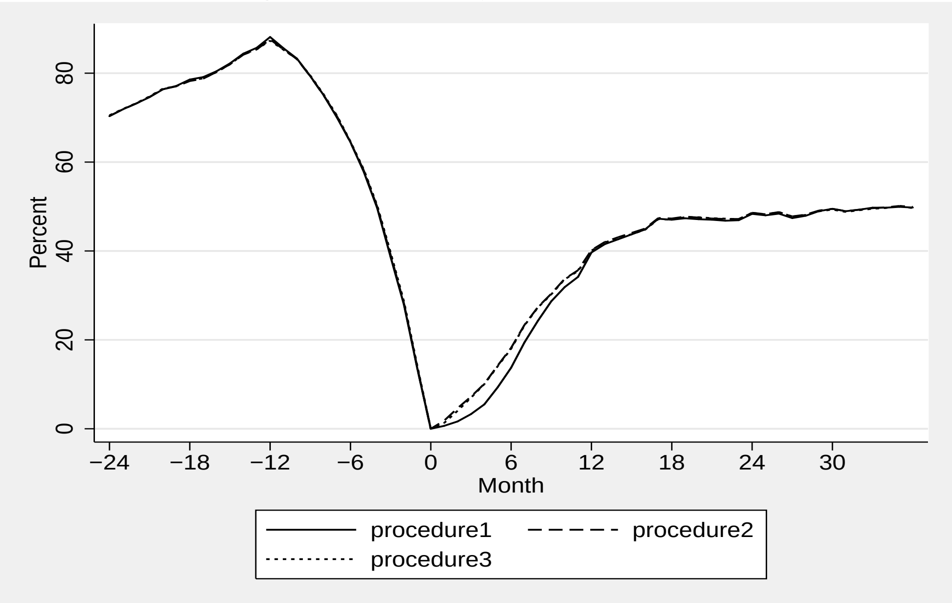
2.4 Sensitivity Analysis I: Frameworks with a Simple Treatment Variable

This section investigates the sensitivity of the use of the different procedures and thus the impact of measurement error in program end dates for frameworks with a simple time constant treatment variable. In these frameworks individuals who start a program within some time window and attend it for a minimal amount of days are counted as participants (in a multiple framework of one of the measures). Individuals who do not attend a program (or who do not attend it long enough) are counted as nonparticipants.

2.4.1 Impact on Employment Rates of Participants

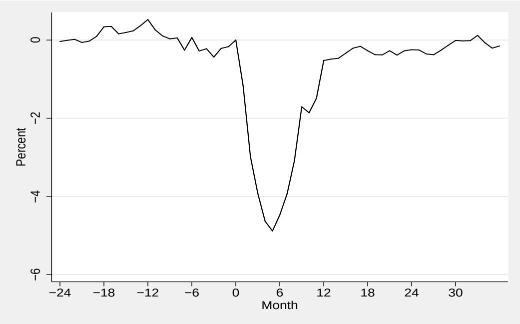
The channels through which measurement error in end dates influences results in frameworks with a simple treatment variable can be analyzed in studying employment rates estimated using the three procedures. Figure 2.2 shows the employment rate of FT participants for each month before and after the beginning of treatment (month zero) for each of the three procedures.[18] Figure 2.3 is just another way of presenting the results by showing the differences of the graphs of figure 2.2.

Figure 2.2: Employment Rate of FT Participants

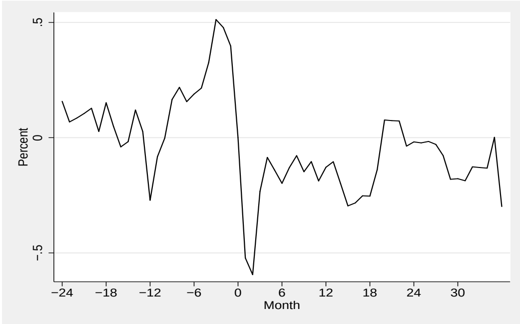


[18]For later months participation rates might be a little underestimated, because employment data of year 2004 is not yet complete.

Figure 2.3: Differences (Employment Rate of FT Participants)



Procedure 1 minus Procedure 2



Procedure 3 minus Procedure 2

Procedure 1 underestimates the employment rate up to almost 5 percentage points as compared to procedure 2. This is because when measuring the outcome, program participation spells are given priority to regular employment spells. If a researcher uses procedure 1 instead of procedure 2 he or she would misinterpret the size of the employment rate.[19] Differences between procedure 1 and procedure 2 occur almost exclusively in between the start and the planned ends of programs, because the two procedures differ with respect to the priority in case of conflict between employment and participation. These conflicts appear within the planned duration of the programs.[20] Thus using procedure 1 or procedure 2 is only relevant for the time where part of the participants is "locked" into programs.

Explicitly correcting end dates (procedure 3) results in a very slightly smaller employment rate than the standard procedure (procedure 2). The measurement of the outcome is the same for procedure 2 and 3, thus differences in the employment rate can only be due to differences in the validation of programs. Some program spells which are wrongly classified as long enough without corrections are too shortly attended to be evaluated or not attended when using corrections. In other words, the treatment indicator will in some cases indicate participation using procedure 2 and non-participation using procedure 3. If the individuals that are counted wrongly as participants in procedure 2 have on average a higher employment rate, procedure 2 will overestimate the employment rate. This seems to be the case (after the start of the programs) but the impact is very small (see figure 2.3). The reason becomes

[19]This is true independently of significance of the difference (because the researcher would only have the results of procedure 1 and interpret these). But to get an idea on significance, figure 2.4 in the appendix gives confidence intervals of the estimated employment rates. But note that they do not represent a correct test of the difference between the procedures, because the results are dependent.

[20]Because of later programs and because of sample differences, there occur minor differences at other times.

clear looking at the numbers of corrections: While end dates change quite often due to corrections (out of the 896 valid treatments in procedure 2 and 3, 13.1% (117) have an earlier end date due to correction) and the corrections are often quite severe (on average 90 days, 59 days is the median, 10% have corrections less than 2 days and 5% more than 266 days), only very few corrections influence the sample and can thus influence the employment rates.[21] Due to the corrections only 15 treatments valid in procedure 2 are not valid in procedure 3 (10 due to correction based on subsistence allowance, 4 due to a very early deregistration date and 1 due to reported non-attendance).

Differences between procedures 2 and 3 last longer than differences between procedures 2 and procedure 1, because the former are due to selection effects and the latter are due to outcome measurement. In conclusion, a considerable amount of end dates is corrected using procedure 3, but this correction has few implications, because the end dates influence the results only through the minimal length criterion, which is rarely concerned by the corrections. This result suggests that in approaches with a simple treatment variable and employment as the outcome variable, it does not make much of a difference if one uses procedure 2 or the more involved procedure 3. But using procedure 1 leads to a downward bias.

2.4.2 Impact on Treatment Effects Using Matching

When estimating employment effects using matching, one typically compares the employment status of participants with a fitted employment status of matched non-participants. This is essentially a comparison of the employment rates of participants and weighted non-participants. As the choice of procedure does (almost) not influence the employment rates of nonparticipants, the impact of the choice of procedure on treatment effects using matching is very similar to the impact on employment rates investigated in section 2.4.1. Therefore the sensitivity on matching results will be discussed only very shortly in this paper.

As shown in section 2.4.1, Procedure 1 underestimates the employment rate of participants for the time of planned program durations as compared to procedure 2, because of different measurement of the employment status in case of wrong end dates. When using matching this underestimation of the employment status of par-

[21] The overall sum of corrections in the data is much higher. Here only treatments in focus are counted.

ticipants will lead to overestimating a negative treatment effect (lock-in-effect) in the first months after program begin. Waller (2007) shows that indeed procedure 1 overestimates the lock-in-effect up to 5.28 percentage points (in month six after program start at a treatment effect of -17.23% for procedure 2) as compared to procedure 2.[22] The difference vanishes about one year after program start. It may be of importance when a cumulative employment effect is calculated for example for cost-benefit analysis. In this case the researcher should refrain from using procedure 1, because this would lead to a considerable bias in estimation results.

A difference between the matching results using procedure 2 or procedure 3 can only evolve (as for the employment rates in section 2.4.1) through the minimal attendance criterion. Using procedure 2 some non-participants will be wrongly classified as participants and this influences the samples of participants and non-participants. This may have an impact on matching results, if those misclassified are on average more or less successful than the other participants. In addition, influence through the estimation of the propensity score or the non-participants available for matching are possible. But as the minimal length criterion is very rarely concerned by the corrections (as shown in section 2.4.1), there are very few changes of the treatment indicator. Thus treatment effects differ only negligibly (up to one percentage point according to Waller (2007)). Therefore in a matching framework with the features described above, an explicit correction of end dates as implemented in procedure 3 does not seem to be necessary unless the exact magnitude of the treatment effect is needed.

2.5 Sensitivity Analysis II: Framework with Time-varying Treatment Variables

Apart from matching methods, duration models are very popular for the estimation of program effects.[23] A simple approach is to compare the Kaplan-Meier survivor functions in unemployment of participants and non-participants. Figure 2.5 in the

[22] For women in West Germany the employment effect of taking an FT program within the first three months of an unemployment period against not taking a program at least until then is estimated. The outcome variable is the probability of regular employment in each month after the start of the program.

[23] Recent examples of studies estimating employment effects of further training programs with German administrative data using duration analysis are Hujer et al. (2006), Schneider and Uhlendorf (2006) and Schneider et al. (2007).

appendix shows the survivor function for FT participants (program start within the first year of unemployment) and non-participants. Day zero is the first day of an unemployment period and the hazard is defined as leaving this unemployment period towards regular employment.[24] Comparing the slopes of the survivor functions of participants and non-participants, the slope of the survivor function of participants is much less steep in the beginning, reflecting the lock-in-effect. But later on, when participants have finished the program and intensify their search for employment again, the slope becomes steeper. The Kaplan-Meier estimates show this effect only very roughly, because participants start and leave the programs at different points in time (and because no observable characteristics are controlled for). But parametric duration analysis (e.g. a proportional hazard model) allows to separate these two phases of the program effects by including treatment variables that change over time. In such a framework the impact of the measurement error is likely to be different from the frameworks with a simple treatment variable because of the importance of the program end date when defining the treatment variables over time.

To investigate the sensitivity of reported end dates a proportional hazard model is estimated. A Weibull distribution is chosen to allow for duration dependence. The duration is again defined as starting with the beginning of a valid unemployment spell and ending with a new regular employment spell. In addition to personal and regional characteristics and information on the individual's labor market history that are supposed to influence the hazard rate (see appendix for the final specification), three time-varying covariates are included in the estimation.[25] The day an individual enters the program under consideration the dummy variable "lock" changes to one. Once she leaves a completed program (defined as having participated for least 80% of the planned duration), the "lock" dummy changes to zero again and a second dummy ("treatfin") is set to one, indicating that this individual has finished a program.[26] In case the individual leaves an uncompleted program, "lock" is also set to zero and a third dummy ("postdrop") is set to one, indicating that the individual has dropped

[24]Sample, minimal length criterion, unemployment and regular employment defined as before. The differences between the procedures are similar to the differences in employment rates discussed in section 2.4.1 for participants and not visible for non-participants (therefore only the results for procedure 2 are shown).

[25]A time-varying covariate is interpreted as a measure of the effect of a one unit change in the covariate at time t on the log hazard (see Lancaster (1990)).

[26]The last day of a completed program is already considered as "treatfin" (if the individual leaves directly to employment), because regarding the effect of a finished program, starting a job directly after a completed program or having days of unemployment in between is considered the same given the length of the whole unemployment duration.

out of a program in the past. These three time-varying dummies allow to study separately how programs bind the unemployed on the one hand and the time after a completed program on the other hand.[27] The distinction between having completed a program or having dropped out is implemented, because these two situations represent different conditions for finding employment. The coefficients may not be interpreted as treatment effects, they just describe some aspects of the complex process that is going on. Particular problems preventing a causal interpretation are the potential endogeneity of the program end date and the relation between the dummies "lock" and "treatfin".

Table 2.2: Hazard Ratios for Time-varying Dummies

Procedure	1	2	3
lock	0.086 [0.059; 0.126]	0.129 [0.094; 0.177]	0.291 [0.235; 0.360]
	(879)	(896)	(884)
	(27)	(39)	(86)
treatfin	1.829 [1.671; 2.001]	1.831 [1.674; 2.004]	1.669 [1.516; 1.836]
	(811)	(814)	(731)
postdrop	0.711 [0.452; 1.116]	0.794 [0.517; 1.220]	0.745 [0.530; 1.053]
	(41)	(43)	(67)

Extract of the results of the PH model estimating the hazard to regular employment. In squared brackets the 95% confidence intervals of the hazard-rates are given. The numbers in parentheses are the numbers of individuals that are in the respective state for at least one day of their duration. The second parentheses of "lock" give the numbers of individuals who do only reach "lock", that is leave to employment (or are censored) out of an unfinished treatment. The whole number of individuals varies between 16773 and 18772.

In this framework the influence of the program end date on the results is still indirect, as the end date itself is neither regressor nor outcome variable. But measurement error in the end date may lead to measurement error in the covariates, the coefficients of which shall be interpreted. Table 2.2 shows the hazard ratios (the exponentiated coefficients) for the dummies "lock", "treatfin" and "postdrop" for FT programs for the three procedures. For the coefficients, including those of the additional covariates and standard errors, see appendix. A hazard ratio of 0.291 for "lock" means that the hazard rate for those being currently in an unfinished program is just 29.1% of the hazard rate of those not being in a program. As one would expect, "lock" has a negative and highly significant effect: attending a non finished program comes

[27]This framework is inspired by Schneider and Uhlendorf (2006) and Schneider et al. (2007), who distinguish between a lock-in-effect and a post program effect.

along with a drastic reduction in leaving unemployment. This is also visible from the numbers in the brackets. Whereas 884 women enter an FT program (procedure 3), only 86 end their duration out of the uncompleted program. Using the procedures with no explicit corrections, much less individuals are assessed to end their duration out of an unfinished program. This influences the hazard ratios of "lock": they differ 4.3 percentage points between procedures 1 and 2 and 16.2 percentage points between procedures 2 and 3. Thus the difference between procedures 2 and 3 is more important than between procedures 2 and 1.[28]

The large majority of those assessed to take a program finish it and "treatfin" has a significant positive effect on the hazard rate. As discussed above, this is not to be interpreted as a positive treatment effect, it just says that individuals having finished a program leave unemployment more often than others. The hazard ratios are similar for procedures 1 and 2 but differ 16.2 percentage points between procedures 2 and 3. The reason for this difference is that a procedure without an explicit correction of end dates misclassifies individuals to have finished a program, while they should be classified as being unemployed after an unfinished program or leaving to employment out of an unfinished program (as one can also see from the numbers in brackets). A second effect is that in procedure 2 too many individuals are assessed as leaving directly out of an unfinished program, while in reality they have left the program even before and should be classified as "postdrop" equal to one and "lock" equal to zero. This effect leads c.p. to a too high hazard ratio of "lock" and a too low hazard ratio for "postdrop" using procedure 2. The coefficients of "postdrop" are not significant.

In sum, the results show that in a framework with time-varying treatment variables it can be of importance to explicitly correct end dates when preparing the data as done in procedure 3. In the above duration framework the end date affects the results more directly than in the frameworks of section 2.4, because it is of importance if the end date of the program lies in the past when the individual starts employment and if the end date of the program lies considerably before the planned end date (dropout). But also in this framework, measurement error in end dates changes only

[28]The confidence intervals of the estimates for procedure 2 and 3 do not overlap. This is a hint that the results might be significantly different. But it is not possible to tell for certain, because the estimates are not independent. The question if the results are significantly different or not is of minor importance for the analysis in this study: A researcher using the IEBS would interpret the results he or she gets using for instance procedure 2, provided the confidence interval is not too large. Had he or she used procedure 3, he or she would use these results and the conclusions about the size of the effect would be different.

the magnitude but not the direction of the results.

2.6 Conclusion

Program end dates are a sensitive part of the German Integrated Employment Biographies Sample. Mainly due to early drop-out not corrected in the data, a considerable part of end dates of further training programs are later than the actual end of participation. In this paper three procedures how to deal with the error-prone end dates are presented, a “naive” procedure, a standard procedure and a procedure that explicitly corrects the data before the analysis. The influence of the different procedures on evaluation results is studied in a framework with a simple treatment variable (like typically used in matching) and in a duration framework with time-varying treatment variables. In conclusion, for typical matching studies it does not seem necessary to explicitly correct the data before using it. But especially if there is interest in the size of the lock-in-effect, one should refrain from using the “naive” procedure (giving priority to program data when measuring the outcome), because it considerably overestimates the lock-in-effect. This may be a particular problem when treatment effects are averaged over time to get one number for program comparison or for cost-benefit analysis. In frameworks with time-varying treatment variables, like in the duration model investigated in section 2.5, reported end dates are more important for the generation of the treatment variables. Therefore, if one is interested in the size of the coefficients of the treatment variables, it may be necessary to correct the reported end dates before the analysis.

Concerning the studies that have already evaluated further training programs using the IEBS, the measurement error in program end dates should not be a problem for the conclusions Lechner and Wunsch (2006) draw. They calculate cumulated effects using their matching results to compare programs, but do not interpret the exact magnitude of these effects. Thus in case they gave priority to program data, this might have biased the size of the cumulated effects a little, but would not have changed their conclusions. For the results of the microeconometric analysis of further training in the context of the so called *Hartz-Reformen* (Schneider et al. 2007) the error-prone end dates of the IEBS might be a (very small) problem. From the description on how the outcome variable is generated (Schneider et al. (2007), p.104) it seems as if program spells have been given priority to employment spells (procedure 1). Provided the authors have not done some end dates corrections not mentioned in the report before the generation of the outcome variable, the results should suffer from the problems which occur when using procedure 1. The authors use matching results for a cost-benefit analysis. They calculate a cost-benefit effect cumulated over the time from program start until the end of the observation period (p.144), thus they include results for the lock-in-period for which procedure 1 overestimates the negative effect. Therefore the estimate of the cost-benefit relation in the report might be a bit too negative. In addition Schneider et al. (2007) estimate a duration model with time-varying treatment variables. The model is somewhat different from the one estimated above, but the way measurement error in end dates influences the results should be similar. Thus, if Schneider et al. (2007) have not corrected the end dates before their analysis, the size of the results might be a little biased. But this is not a problem for their conclusions, because the authors do not interpret the coefficients themselves but focus on the reform effect. In conclusion, the overall small effect of error-prone end dates on evaluation results is good news for researchers using the IEBS and also for those using different administrative data sets in which reported program end dates cannot be corrected but may nevertheless be error-prone.

References

Battistin, E. and B. Sianesi (2006). "Misreported Schooling and Returns to Education: Evidence from the UK." Cemmap Working Papers, CWP 07/06.

Bender, S., M. Biewen, B. Fitzenberger, M. Lechner, S. Lischke, R. Miquel, A. Osikominu, T. Wenzel and C. Wunsch (2004). "Die Beschäftigungswirkung der FbW-Maßnahmen 2000 - 2002 auf individueller Ebene: Eine Evaluation auf Basis der prozessproduzierten Daten des IAB - Vorläufiger Zwischenbericht Oktober 2004." Goethe University Frankfurt and SIAW St. Gallen.

Bender, S., M. Biewen, B. Fitzenberger, M. Lechner, R. Miquel, A. Osikominu, M. Waller and C. Wunsch (2005). "Die Beschäftigungswirkung der FbW-Maßnahmen 2000 - 2002 auf individueller Ebene: Eine Evaluation auf Basis der prozessproduzierten Daten des IAB - Zwischenbericht Oktober 2005." Goethe University Frankfurt and SIAW St. Gallen.

Bericht 2006 der Bundesregierung zur Wirksamkeit moderner Dienstleistungen am Arbeitsmarkt (2006). Deutscher Bundestag, Drucksache 16/3982.

Bernhard, S., C. Dressel, B. Fitzenberger, D. Schnitzlein and G. Stephan (2006). "Überschneidungen in der IEBS: Deskriptive Auswertung und Interpretation." FDZ Methodenbericht 04/06, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Biewen, M., B. Fitzenberger, A. Osikominu, M. Waller and R. Völter (2006). "Beschäftigungseffekte ausgewählter Maßnahmen der beruflichen Weiterbildung in Deutschland - eine Bestandsaufnahme." *Journal for Labor Market Research* 39, 365-390.

Biewen, M., B. Fitzenberger, A. Osikominu and M. Waller (2007). "Which program for whom? Evidence on the comparative effectiveness of public sponsored training programs in Germany." IZA Discussion Paper No. 2885.

Boockmann, B., T. Zwick, A. Ammermüller and M. Maier (2007). "Do Hiring Subsidies Reduce Unemployment Among the Elderly? Evidence From Two Natural Experiments." ZEW Discussion Paper No. 07-001, Mannheim.

Bundesagentur für Arbeit (2006). "Geschäftsbericht 2005." Nürnberg.

Hujer, R., S. Thomsen and C. Zeiss (2006). "The Effects of Vocational Training Programmes on the Duration of Unemployment in Eastern Germany." *Allgemeines Statistisches Archiv* 90, 299-321.

Hummel, E., P. Jacobebbinghaus, A. Kohlmann, M. Oertel, C. Wübbeke and M. Ziegerer (2005). "Stichprobe der Integrierten Erwerbsbiographien, IEBS 1.0, FDZ Datenreport 6/2005." Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Jaenichen, U., T. Kruppe, G. Stephan, B. Ulrich and F. Wießner (2005). "You Can Split it if You Really Want: Korrekturvorschläge für ausgewählte Inkonsistenzen in der MTG und IEB, FDZ Datenreport 4/2005." Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Jaenichen, U. and G. Stephan (2007). "The Effectiveness of Targeted Wage Subsidies for Hard-to-place Workers." IAB Discussion Paper 16/2007, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Kruppe., M and T. Oertel, (2003). “Von Verwaltungsdaten zu Forschungsdaten.” IAB Werkstattbericht 10/2003, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Lancaster, T. (1990). “The Econometric Analysis of Transition Data.” Cambridge.

Lechner, M. and C. Wunsch (2006). “Active Labour Market Policy in East Germany: Waiting for the Economy to Take Off.” IZA Discussion Paper No. 2363.

Lewbel, A. (2004). “Estimation of Average Treatment Effect with Missclassification.” Boston College.

Molinari, F. (2005). “Missing Treatments.” CAE Working Paper, 05-11.

Pfeiffer, F. and H. Winterhager (2006). “Vermittlungsgutscheine und Beauftragungen Dritter im Vergleich.” *Journal for Labour Market Research* 39, 425-445.

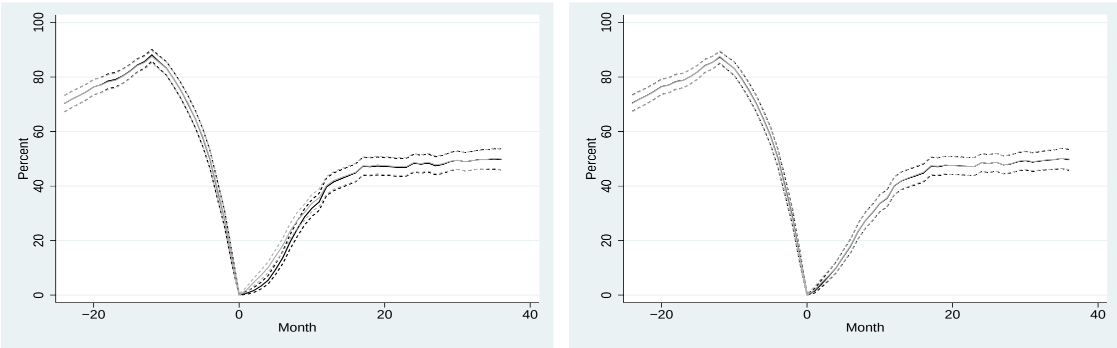
Schneider, H. and A. Uhlendorff (2006). “Die Wirkung der Hartz-Reform im Bereich der beruflichen Weiterbildung.” *Journal for Labor Market Research* 39, 477-490.

Schneider, H. , K. Brenke, B. Jesske, L. Kaiser, U. Rinne, M. Schneider, J. Steinwede and A. Uhlendorff (2007). “Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission - Modul 1b: Förderung beruflicher Weiterbildung und Transferleistungen.” IZA Research Report No. 10, Bonn.

Waller, M. (2007). “Do Reported End Dates of Treatments Matter for Evaluation Results? - An Investigation Based on the German Integrated Employment Biographies Sample.” FDZ Methodenreport 01/07, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Appendix

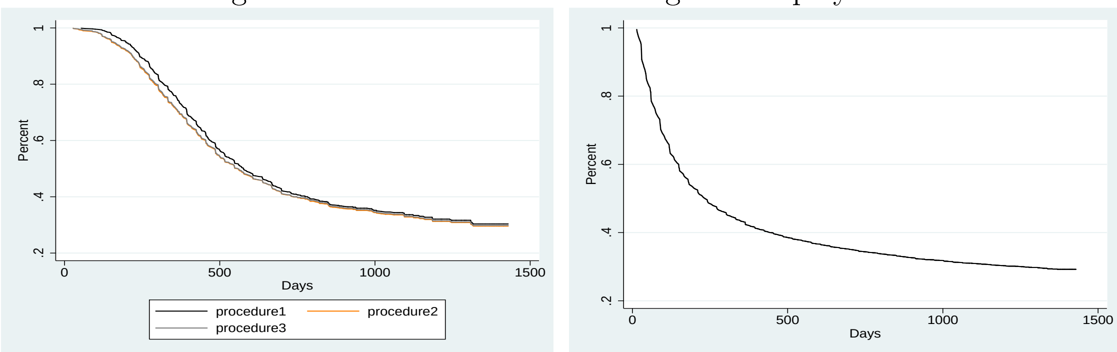
Figure 2.4: Employment Rate of FT Participants with Confidence Intervals



Procedure 1 in black, Procedure 2 in grey

Procedure 3 in black, Procedure 2 in grey

Figure 2.5: Survival until New Regular Employment



FT Participants

Non-participants

Table 2.3: Coefficients of PH Model

Covariate (exact definition upon request)	Procedure 1	Procedure 2	Procedure 3
lock (currently attending the program)	-2.451 (0.193)***	-2.048 (0.161)***	-1.235 (0.108)***
treatfin (has attended and completed)	0.604 (0.046)***	0.605 (0.046)***	0.512 (0.049)***
postdrop (has dropped out)	-0.342 (0.230)	-0.231 (0.219)	0.291 (0.175)*
age 25 - 29	0.169 (0.031)***	0.170 (0.031)***	0.169 (0.031)***
age 30 - 34	0.156 (0.030)***	0.161 (0.030)***	0.158 (0.030)***
age 35 - 39	0.198 (0.030)***	0.203 (0.029)***	0.200 (0.029)***
age 40 -44	0.280 (0.030)***	0.290 (0.030)***	0.285 (0.030)***
days in employment last 3 years	0 (0.000)***	0 (0.000)***	0.000 (0.000)***
log last daily censored wage	0.051 (0.015)***	0.050 (0.015)***	0.050 (0.015)***
unemployment benefit last 3 years	0.328 (0.023)***	0.324 (0.023)***	0.322 (0.023)***
unemployment assistance last 3 years	-0.152 (0.034)***	-0.146 (0.033)***	-0.140 (0.033)***

	Procedure 1	Procedure 2	Procedure 3
days out of sample last 3 years	0.00009 (0.000)	0.00009 (0.000)	0.0008 (0.000)
days subsistence allowance last 3 years	0.0002 (0.000)	0.0002 (0.000)	0.0002 (0.000)
unemployment rate home district	-1.175 (0.262)***	-1.570 (0.262)***	-1.752 (0.262)***
foreigner	-0.109 (0.033)***	-0.110 (0.033)***	-0.108 (0.033)***
region3 (IAB classification)	0.038 (0.027)	0.029 (0.027)	0.029 (0.027)
region4 (IAB classification)	0.153 (0.034)***	0.139 (0.034)***	0.134 (0.034)***
region5 (IAB classification)	0.205 (0.029)***	0.185 (0.029)***	0.184 (0.029)***
health problem (no impact)	-0.298 (0.047)***	-0.291 (0.046)***	-0.294 (0.046)***
health problem (impact on placement)	-0.438 (0.053)***	-0.435 (0.052)***	-0.439 (0.052)***
no degree	-0.038 (0.048)	-0.041 (0.048)	-0.041 (0.048)
vocational training degree	-0.018 (0.044)	-0.023 (0.043)	-0.022 (0.043)
9 or 10 years of schooling degree	0.036 (0.040)	0.038 (0.040)	0.041 (0.040)
12 or 13 years of schooling degree	0.115 (0.048)**	0.118 (0.048)**	0.120 (0.048)
living alone	0.843 (0.032)***	0.833 (0.032)***	0.831 (0.032)***
not living alone but not married	0.709 (0.055)***	0.694 (0.055)***	0.691 (0.055)***
single parent	0.608 (0.044)***	0.596 (0.044)***	0.591 (0.044)***
married	0.546 (0.032)***	0.538 (0.031)***	0.543 (0.031)***
at least one child	0.188 (0.025)***	0.184 (0.025)***	0.186 (0.025)***
last job less than full time	-0.113 (0.024)***	-0.106 (0.024)***	-0.107 (0.024)***
last job in agriculture	0.251 (0.073)***	0.253 (0.073)***	0.245 (0.073)***
last job in industry	-0.172 (0.030)***	-0.170 (0.030)***	-0.171 (0.030)***
last job in commerce, traffic, hotel	0.078 (0.024)***	0.080 (0.024)***	-0.077 (0.024)***
last job in financial sector	0.015 (0.028)	0.015 (0.028)	0.015 (0.028)
last job whitecollar-job	0.083 (0.026)***	0.088 (0.026)***	0.090 (0.026)***
no wish to change occupation	0.208 (0.026)***	0.211 (0.025)***	0.212 (0.025)***
end of last job in 2 or 3/2000	-0.074 (0.043)*	-0.062 (0.043)***	-0.056 (0.043)
end of last job in 4/2000 - 6/2000	-0.148 (0.038)***	-0.142 (0.037)***	-0.141 (0.037)***
end of last job in 7/2000 - 9/2000	-0.129 (0.035)***	-0.122 (0.035)***	-0.119 (0.035)***
end of last job in 10/2000 - 12/2000	-0.053 (0.033)	-0.057 (0.032)*	-0.053 (0.032)
end of last job in 1/2001 - 3/2001	-0.073 (0.033)**	-0.065 (0.032)**	-0.060 (0.032)*
end of last job in 4/2001 - 6/2001	-0.116 (0.036)***	-0.112 (0.036)***	-0.108 (0.036)***
end of last job in 7/2001 - 9/2001	-0.128 (0.035)***	-0.127 (0.035)***	-0.126 (0.035)***
lack of cooperation	-0.044 (0.032)	-0.047 (0.032)	-0.049 (0.032)
program with social assistance in past	-0.166 (0.068)**	-0.123 (0.060)**	-0.124 (0.060)**
_cons	-5.568 (0.122)***	-5.518 (0.121)***	-5.510 (0.122)***

Chapter 3

Déjà Vu? Short-Term Training in Germany 1980-1992 and 2000-2003

3.1 Introduction

> *"… there is almost never a stable set of active programmes to evaluate. Countries are continuously chopping and changing the mix of programmes."*
>
> *Martin and Grubb (2001, p. 21)*

Over the past few years, active labor market policies (ALMP) have placed a greater emphasis on job search assistance, monitoring and testing work availability, as well as limited training to activate the unemployed in the short run (OECD, 2007). There has been a greater focus to activate the unemployed to find unsubsidized jobs instead of placing unemployed in longer traditional training programs or public employment schemes. Short-term programs are replacing longer programs in order to prevent long lock-in effects.

Public sector sponsored training has traditionally been a main part of ALMP in many countries including Germany, see the surveys in Fay (1996), Martin and Grubb (2001), and Kluve (2006). Although there were many pessimistic assessments regarding the usefulness of such programs, these surveys point out that small-scale training programs, which are well targeted to specific groups and which involve a strong on-the-job component, can show positive employment effects. Little is known in the literature on the medium- and long-run effects of activation strategies which combine training, job search assistance, and monitoring.

In Germany, the focus on activation strategies is reflected in the recent shift away from traditional longer further training programs, typically lasting a couple of months up to two years, to short-term training programs (*Trainingsmaßnahmen*, henceforth denoted by 'ST00'≡short-term training in the 2000s) lasting at most twelve weeks. In fact, ST00 have become the largest programs in Germany regarding the number of participants with 1.07 million individuals entering such a program in 2007 (Bundesagentur für Arbeit, 2007, pp. 54, 57). In contrast, only 356 thousand individuals entered longer further training programs in 2007.[1] Longer further training programs used to be the largest programs in Germany but have been replaced to a large extent by ST00. In light of the recent evidence that long training programs mostly show positive long-run employment effects (Fitzenberger et al., 2008; Lechner et al., 2004), one might be concerned that a focus on activation strategies comes

[1] In 2000, there was a reverse ranking with 552 thousand individuals entering longer further training programs and 477 thousand individuals entering ST00 (see table 3.2 in the appendix).

at the expense of pushing the unemployed into instable jobs which do not result in permanently better employment prospects. There are different types of ST00 programs. In this paper, we distinguish programs which focus on skill provision and programs which focus on testing and monitoring search effort.

Between 1980 and 1992, short-term training programs similar in nature to ST00 were in place in West Germany (and since 1990 also in East Germany). These were the 'programs according to §41a Employment Promotion Act' (*Maßnahmen nach §41a Arbeitsförderungsgesetz*, henceforth denoted by 'ST8092'≡short-term training between 1980 and 1992). Due to budgetary reasons these programs were abolished in 1992. In 1998, short-term training in form of ST00 was reintroduced into the Social Code III (*Sozialgesetzbuch III*) that nowadays regulates labor market policy. While activation and monitoring is a major goal of ST00, the older ST8092 focus solely on job search assistance, limited training, and guidance towards future participation in longer training programs. Furthermore, the ST8092 programs were targeted to low-skilled and hard-to-place unemployed. The common features of the two programs are provision of short-term training, assessment of the unemployed (e.g. regarding future assignment to longer labor market programs), and job search assistance.

There have been a number of studies which evaluate effects of short-term training since 2000 (ST00) using different program evaluation estimators (Hujer et al., 2006; Lechner and Wunsch, 2007; Biewen et al., 2007; Stephan et al., 2006; Büttner, 2007; Osikominu, 2008). We are not aware of any study which uses modern approaches to estimate treatment effects for the older ST8092 programs. In the following, we summarize the evidence for ST00 and other short further training programs in West Germany. Lechner et al. (2004) analyze shorter further training programs in the 1990s that last longer than ST00 or ST8092 programs and provide more sizeable investments into professional skills. This study finds that the cumulated long-run employment effects of shorter training are higher than for longer training programs. Lechner and Wunsch (2006) show that the effect of longer training programs differs over the business cycle such that these programs show better employment effects when unemployment is high. This suggests that the activation effect of these programs on the unemployed is higher when unemployment is high. To our knowledge, no comparable evidence exists for short-term training, which has a stronger focus on activation compared to longer training programs.

Hujer et al. (2006) and Osikominu (2008) apply duration methods to evaluate ST00.

They find evidence that ST00 reduce the duration of unemployment by increasing the hazard rate for exits from unemployment to employment in the short run. Osikominu (2008) finds no long-run effects on the hazard rate from employment back to unemployment. According to Hujer et al. (2006) men tend to benefit more strongly than women. Osikominu (2008), in contrast, does not find significant gender differences. Using different versions of matching estimators, Biewen et al. (2007) and Lechner and Wunsch (2007) show that for the time period of the early 2000s ST00 tend to perform better than longer further training programs regarding their employment effects. Biewen et al. (2007) find some significantly positive employment effects for ST00 in West Germany, whereas Lechner and Wunsch (2007) find no significantly positive treatment effects.

The studies reviewed so far do not distinguish different types of ST00. Stephan et al. (2006) consider participation in different versions of ST00 in the second half of the year 2002. The study uses a matching estimator and finds differing results depending on the type of ST00. The monitoring and testing version of ST00 does not show positive results, whereas the training versions show significantly positive or negative results depending upon whether the training takes place in a firm. Büttner (2007) uses a small experimental data set for 2005 in one region of West Germany and investigates the effect of sending an invitation to participate in a ST00 program which involves monitoring and testing. Out of 189 unemployed receiving an invitation 77 actually participate. The focus is on distinguishing the effects of announcement ('threat') of treatment from the effect of actual treatment. The study finds differences between the announcement effect and the treatment effect. In fact, the announcement results in earlier exits from unemployment, whereas the actual treatment shows no such effects. The exits from unemployment, however, do not translate into significantly higher exits to employment.

The literature review reveals that most studies did not distinguish different types of ST00 programs and that estimates of the long-run effects of short-term training are missing. This paper estimates the effects of short-term training programs in West Germany both for the time period 1980 to 1992 (ST8092) and 2000 to 2003 (ST00) regarding both employment and participation in longer-term training programs. This paper is the first to use state-of-the-art estimators of treatment effects for the short-term training programs in the 1980s and early 1990s. We investigate in particular whether there are lasting positive effects on employment outcomes and whether participation in these programs leads to higher participation in longer training programs afterwards. Furthermore, we analyze whether treatment effects differ over calendar time. We provide similar estimates for participation in short-term training in the early 2000s. Because the ST8092 programs were not intended to test and monitor the unemployed, we distinguish two versions of ST00, namely the training variant which focuses on skill provision (QST00) and the checking variant which focuses on testing and monitoring search effort (MST00). We argue that the ST8092 programs are to be compared to the QST00 version of ST00.

Methodologically, this paper follows Sianesi (2004) and estimates the effects of treatment starting after some given unemployment experience against the alternative of not starting treatment at this point in time and waiting longer. To be able to compare the results for the 1980s, 1990s, and early 2000s, we use the same methodological approach in all cases. Most evaluation studies in the past used a static approach evaluating the effects of receiving treatment during a certain period of time against the alternative of not receiving treatment during this period of time.[2] In a dynamic setting, the timing of events becomes important, see Abbring and van den Berg (2003), Fredriksson and Johansson (2003), and Sianesi (2003, 2004). Static treatment evaluations implicitly condition on future outcomes leading to possibly biased treatment effects. The nontreated individuals in the data might be observed as nontreated because their treatment starts after the end of the observation period or because they exit unemployment before treatment starts (Fredriksson and Johansson, 2003).

Appropriate data for a long-term evaluation of public sector sponsored training programs were not available for a long time. This is the first paper using administrative data covering such a long time period, namely 18 years in the 1980s and 1990s and four years in the early 2000s to study the medium-term and, for the earlier time period, also the long-term employment effects of short-term training. The comparison between the earlier and the more recent time period is interesting because of the similarities between the two programs. In addition to employment, we also consider the effects on future participation in longer further training programs. This is important because one stated goal of short-term training in Germany is to assess the unemployed's need to participate in longer-term training programs. However, with an increasing focus on short-run activation strategies this goal may have lost in importance over time.

[2]Hujer et al. (2006), Biewen et al. (2007), and Osikominu (2008) are exceptions.

According to our results, short-term training shows mostly persistently positive and often significant employment effects. The effects are particularly strong when participation starts during months seven to twelve of the unemployment spell. The effects for short-term training starting during the second year of the unemployment spell tend to be smaller. The monitoring variant MST00 shows slightly smaller effects compared to the pure training variant QST00. The lock-in periods last longer for ST8092 compared to ST00 and the employment effects tend to be smaller for the earlier time period compared to QST00 but not compared to MST00. Short-term training results in higher future participation in longer further training programs and this effect is much stronger for the earlier time period.

The remainder of this paper is structured as follows: Section 3.2 discusses the institutional aspects of short-term training in Germany. Section 3.3 focuses on the data used. Section 3.4 describes the methodological approach to estimate the treatment effects. The empirical results are discussed in section 3.5. Section 3.6 concludes. The appendix provides detailed empirical results.

3.2 Institutional Background

In Germany, training is traditionally a very important part of active labor market policy that aims at permanently reintegrating unemployed individuals into the labor market.[3] Among the different types of training programs offered, medium- and long-term further training programs with a duration of up to two years used to play the most important role since their introduction in 1969. During the 1980s and since 1999, short-term training programs have been used at a large scale, too. Table 3.1 displays the entries into different types of active labor market programs in West Germany in the period 1979 to 1992. It can be seen that entries into short-term training rose steadily until 1987, remained at a lower level in 1988 and 1989 and peaked again in 1990. Table 3.2 shows the participation numbers in Germany as well as West Germany for the more recent period since 1999. During the recent years participation in short-term training rose considerably. Since 2001, short-term training has become the most important type of training regarding the number of participants.

[3] Other important policy instruments are for instance employment subsidies, job creation in the public sector and measures to promote self-employment.

Modern short-term training programs (ST00) have two main goals. First, they are supposed to enhance reintegration of the participants into employment through guidance and qualification. This may comprise training job search skills through activities such as job-application training, simulation of job interviews or general counseling on job search methods. It may also involve the provision of specific skills (like limited computer skills or some technical tasks) that are necessary to improve the job seeker's labor market prospects.[4] The second aim of short-term training is to assess the job seekers' labor market opportunities and their suitability for different jobs but also their availability and willingness to work. This may entail the preparation of detailed work plans to reintegrate the job seeker into the labor market, which can include participation in a longer training program.[5] The availability of the unemployed is checked by pledging him or her to attend the fulltime training program. In our empirical analysis we therefore distinguish short-term training programs for which the objective of qualifying the job seeker dominates from programs that put more emphasis on testing the availability to work and assessing the job seekers' opportunities using the information on the program codes in the data. Such a distinction can only be an approximation, as the same program can serve both purposes, even for the same participant. However, this distinction is also useful for the comparison with short-term training in the period 1980 to 1992, where testing work availability was no (official) goal.

ST00 programs last between two and twelve weeks (with median duration around four weeks). Therefore, they are relatively cheap compared to the longer further training programs. In fact, a one-month short-term training course costs on average €590 per participant, whereas participation costs for a further training course lasting nine months amount to about €5850, see Biewen et al. (2007, table 1).

In the 1980s and 1990s, there existed short-term training programs (ST8092) that were very similar to those described above. The law governing active labor market policy at that time, the Employment Promotion Act *(Arbeitsförderungsgesetz)*, included a paragraph on "measures to improve the employment chances for the unemployed" (*Maßnahmen zur Verbesserung der Vermittlungsaussichten für Arbeitslose*). The number of this paragraph gave the programs their name: 'measures according to §41a'. These programs where introduced in 1979 after the German

[4]For more details on the contents of short-term training see Kurtz (2003).

[5]One element of the law called Job-AQTIV Gesetz introduced in 2002 is to assess the job seeker soon after becoming unemployed. This may be done through a short-term training program (Kurtz, 2003).

labor market conditions had worsened in the 1970s and the number of long-term unemployed had risen considerably.

ST8092 programs were particularly targeted at individuals with lower reemployment chances as women, individuals without formal qualification and long-term unemployed. Hard-to-place and low-skilled individuals were under-represented in the existing medium- and long-term training programs. Short-term training was intended to counsel job seekers about their employment chances and the possibilities of participating in medium- or long-term training programs on the one hand and to teach limited skills helpful for either employment or participation in a longer training program on the other hand (Dobischat and Wassmann, 1981). Similar to ST00, ST8092 programs mostly consisted of fulltime classroom training. The curriculum covered e.g. job counseling, information on public sponsored further training programs and on the general labor market situation, application and communication training, visiting firms and exercises with the intention to stabilize the personality of the participants. The maximal length was in general six weeks and there was no exam at the end of the course (Schneider, 1981).

At the end of 1992, ST8092 programs were abolished in order to reduce the costs of active labor market policy in a time of narrow budget. More intensive and completely sponsored short-term training programs only reappeared in 1997 (in the first years with a small number of participants only) and became important again from 1999 onwards (Kurtz, 2003).

When becoming unemployed individuals have to personally register at the local labor office. This involves a first counseling interview with the caseworker. Further interviews may follow from time to time. Based on these interviews in general the case workers decides whether to assign an unemployed to a program. Besides being registered as unemployed or as a job seeker at risk of becoming unemployed, candidates for short-term training do not have to fulfil any additional eligibility criteria. Depending on regional and local circumstances, caseworkers exercise a considerable amount of discretion when allocating unemployed to the different programs. Suitable programs are chosen from a pool of certified public or private providers.

The employment office pays all direct training costs for short-term training programs. In addition, ST00 participants continue to receive unemployment benefits or means-tested unemployment assistance, if they are eligible for such transfer payments. Thus, in the early 2000s, there exist no pure financial incentives for unemployed individuals to participate in ST00, in contrast to the situation in Germany before 1998. In the 1980s, short-term training was treated in the same way as longer further training programs. This means that participants who fulfilled certain eligibility criteria (mainly 720 days of employment subject to social security contributions within the last three years) received an income maintenance allowance which was more generous than the usual unemployment compensation. Those who where not eligible to receive income maintenance allowance continued to receive the means-tested unemployment assistance (Bender et al., 2005).

3.3 Data

3.3.1 Administrative Data Sets Used

This study uses large administrative data sets for both time periods under investigation. For the 2000s, the empirical analysis is based on the so-called Integrated Employment Biographies Sample (IEBS), a data set which has recently been made available by the Federal Employment Office of Germany.[6] The IEBS consists of a 2.2% random sample of individual data drawn from the universe of data records collected in four different administrative processes: the Employment History (*Beschäftigten-Historik*), the Benefit Recipient History (*Leistungsempfänger-Historik*), the Data on Job Search Originating from the Applicants Pool Database (*Bewerberangebot*), and the Participants-in-Measures Data (*Maßnahme-Teilnehmer-Gesamtdatenbank*).[7]

The Employment History is based on social insurance register data comprising employment information for employees subject to contributions to the public social security system. It covers the time period from 1990 to 2004. The main feature of these data is detailed daily information on the employment of each recorded individual. We use this information to account for the labor market history of individuals as well as to measure employment outcomes. For each employment spell, in addition to start and end dates, data from the Employment History contains information on personal as well as job and firm characteristics such as wage, industry, or occupation.

[6]For detailed information on the IEBS see Hummel et al. (2005) and Bender, Biewen et al. (2005). Information in English can be found in Jacobebbinghaus and Seth (2007) or on the website of the Research Data Center (FDZ) of the German Federal Employment Office (http://fdz.iab.de/en.aspx).

[7]The data used here have been supplemented with some additional information that are not available in the standard version.

The Benefit Recipient History, the second data source, includes daily spells of unemployment benefit, unemployment assistance and income maintenance allowance payments individuals received between January 1990 and June 2005. The Benefit Recipient History provides information on the periods in which individuals were out of employment and therefore not covered by the Employment History. Moreover, we use additional information contained in the Benefit Recipients History involving sanctions and periods of exclusion from benefit receipt that may serve as indicators for a lack of motivation. Based on the information in the Employment and the Benefit Recipient History we calculate the individual entitlement periods to unemployment benefits.[8]

The third data source included in the IEBS is the so-called Data on Job Search Originating from the Applicants Pool Database, which contains rich information on individuals searching for jobs covering the period January 2000 to June 2005. The spells include detailed information concerning job search and personal characteristics, in particular on educational qualifications, nationality, and marital status. They also provide information on whether the applicant wishes to change occupation, how many job proposals he or she already got, and about health problems that might influence employment chances. Finally, the data on applicants include regional and local identifiers, which we use to link regional and local information, for example unemployment rates at the district level.

The Participants-in-Measures Data, the fourth data source, contains detailed information on participation in public sector sponsored labor market programs covering the period January 2000 to June 2005. The data consist of spells indicating the start and end dates at a daily level, the type of the program as well as additional information. The Data Base of Program Participants allows us not only to identify participation in short-term training, but also in other programs such as employment subsidies. This is useful, as it enables us to distinguish between regular and subsidized employment when evaluating employment outcomes.

For the earlier time period covering the 1980s and 1990s, we use administrative individual data from three different sources. These data were assembled for the purpose of evaluating public sector sponsored training programs, see Bender, Bergemann et al. (2005) for a detailed description. The first data source is the IAB Employment Subsample (*IAB Beschäftigtenstichprobe*, IABS) of the Institute for Employment

[8]For this purpose we rely on Plaßmann (2002) who summarizes the regulations regarding entitlements to unemployment benefits.

Research (IAB), see Bender et al. (2000) and Bender, Bergemann et al. (2005, chapter 2.1). The IABS is a 1% random sample of all employment records subject to social insurance contribution in the period 1975-1997. It also contains some information on periods of transfer payments from the unemployment insurance. The second data source is the Benefit Payment Register (*Leistungsempfängerdatei*, LED) of the Federal Employment Office, see Bender, Bergemann et al. (2005, chapter 2.2). These data consist of spells on periods of transfer payments granted to unemployed and program participants in the period 1975-1997. They include very detailed information on income maintenance payments which allows to identify participation in different training programs, including the ST8092 programs investigated here. These benefit data contain more detailed information than the benefit data available in the IABS. The two data sources were merged to the so-called IABS-LED data set, see Bender, Bergemann et al. (2005) for details. Based on the IABS-LED data we calculate the individual entitlement periods to unemployment benefits.

As a third data source, we use an administrative survey on training participation, the so called FuU-data, see Bender, Bergemann et al. (2005, chapter 2.3). The Federal Employment Office collected these data for all participants in further training, retraining, and other training programs for internal monitoring and statistical purposes. For every participant, the FuU-data contain detailed information on the program and the participant.

The FuU-data were merged with the combined IABS-LED data by social insurance number and additional covariates. Numerous corrections were implemented in order to improve the quality of the data, see Bender, Bergemann et al. (2005, chapters 3-4) and Fitzenberger et al. (2008) for details. While the IABS provides information on personal characteristics and employment histories, the combination of the transfer payment data and the training participation data is used to identify the participation in different types of training programs.

3.3.2 Sample Selection

In this study, we analyze inflow samples into unemployment consisting of individuals living in West Germany who became unemployed after having been continuously employed for at least three months. The beginning of an unemployment spell is defined as the transition from regular (not marginal) employment to nonemployment and subsequently being in contact with the employment office (not necessarily immediately), either through benefit receipt, program participation or a job search spell.[9] This way, we focus on individuals closely attached to the labor market, which allows to construct a control group that exhibits a similar employment history as the treated individuals. Furthermore, the beginning of unemployment defines a natural time scale to align treated and nontreated individuals. In order to exclude individuals in formal education or vocational training and individuals eligible for early retirement schemes, we only consider persons aged between 25 and 53 years at the beginning of their unemployment spell. Our evaluation focuses on participation in short-term training as the first training program that is attended over the course of an unemployment spell. Later participation in other active labor market programs is regarded as an outcome. Individuals in our control group may participate in another training program as a first program.

For the evaluation of ST00, we focus on an inflow sample into unemployment between the beginning of January 2000 and the end of June 2001. The analysis of ST8092 is based on an inflow sample into unemployment from January 1980 to January 1991.[10] We consider participation in short-term training within the first two years of an unemployment spell. Thus, we evaluate ST8092 programs starting from January 1980 until their abolishment in December 1992 and ST00 programs starting between January 2000 and June 2003. For the earlier time period, the data allow us to follow all individuals until the end of 1997. Therefore, we are able to estimate long-term effects of the ST8092 programs for all participants in our data. We follow the individuals in the more recent sample until the end of 2004.

In the sample covering the early 2000s, we distinguish two types of short-term training programs: the first one puts more emphasis on qualifying the job seeker (QST00), while the second one focuses on monitoring and testing the availability for work (MST00). We argue that the QST00 variant of ST00 is more similar to the ST8092 programs. For both time periods, we distinguish between treatment starting during months 0 to 6 of the unemployment spell (stratum 1), treatment starting during months 7 to 12 (stratum 2), and treatment starting during months 13 to 24 (stratum 3). We consider two outcome variables: the monthly employment status and

[9]In the IEBS we can identify subsidized employment and thus exclude this from our definition of regular employment. This is unfortunately not possible for the 1980s and 1990s.

[10]This implies that the same individual may appear more than once in our evaluation sample. We take account of multiple inclusions of the same individual in the sample when calculating the standard errors.

participation in a longer-term training program later in the unemployment spell. The propensity scores and the treatment effects are estimated separately for the different program types, strata, and men and women. The number of participants and the size of the control group for each specification are depicted in table 3.3 in the appendix.

3.4 Evaluation Approach

Our goal is to analyze the effect of short-term training programs on two outcome variables, namely the individual monthly employment dummy and the individual participation in a longer-term training program.[11] The treatment we evaluate is participating in a short-term training program as a first training program over the course of an unemployment spell against the alternative of not participating in a short-term training program as a first training program. This alternative includes the case of participating in a longer training program as first training program or no participation in any training program. We estimate the average treatment effect on the treated (ATT) of short-term training as first treatment against this alternative. Extending the static treatment approach to a dynamic setting, we follow Sianesi (2004) and apply the standard static treatment approach recursively depending on the elapsed unemployment duration. The implementation builds upon the approach developed in Fitzenberger and Speckesser (2007), Biewen et al. (2007), and Fitzenberger et al. (2008).

Our empirical analysis is based upon the potential-outcome-approach to causality, see Roy (1951), Rubin (1974), and the survey of Heckman et al. (1999). Let the two potential outcomes be $\{Y^0, Y^1\}$, where Y^1 represents the outcome associated with participation in a short-term training program and Y^0 is the outcome when the individual does not participate in a short-term training program. For each individual, only one of the two potential outcomes is observed and the other outcome is counterfactual. We focus on the average treatment effect on the treated (ATT) of participating in a short-term training program against nonparticipation in a short-term training program at some given elapsed unemployment duration (treatment versus waiting).

[11] The individual participation is measured as a dummy variable which is equal to one when the individual participates in a longer-term training program at some time in the future within the same unemployment spell.

Fredriksson and Johansson (2003) argue that a static evaluation approach, which assigns unemployed individuals to a treatment group and a nontreatment group based on the treatment information observed in the data within a fixed time window, yields biased treatment effects. This is because the definition of the control group conditions on future outcomes or future treatment. For Sweden, Sianesi (2004) argues that all unemployed individuals are potential future participants in active labor market programs, a view which is particularly plausible for countries with comprehensive systems of active labor market policies (like Germany). In Germany, active labor market programs are implemented at a fairly large scale in international comparison. While unemployed, job seekers are continuously at risk of being assigned to an active labor market program. This discussion implies that a purely static evaluation of the different training programs is not warranted. Following Sianesi (2003, 2004), we analyze the effects of the first participation in a short-term training program during the unemployment spell considered *conditional on the starting date of the treatment.*

We analyze treatment conditional upon the unemployment spell lasting at least until the start of the treatment k and this being the first treatment during the unemployment spell considered. Therefore, the ATT parameter (comparing treatments k and l) of interest is

$$\begin{aligned}(3.1)\qquad \theta(u,\tau) = {} & E(Y^1(u,\tau)|T_u = 1, U \geq u-1, T_1 = ... = T_{u-1} = 0) \\ & -E(Y^0(u,\tau)|T_u = 1, U \geq u-1, T_1 = ... = T_{u-1} = 0)\ ,\end{aligned}$$

where T_u is the dummy variable indicating the start of treatment starting in month u of the unemployment spell. $Y^1(u,\tau)$, $Y^0(u,\tau)$ are the potential outcomes for treatments and nontreatment, respectively, in periods $u+\tau$, where treatment starts in period u and $\tau = 0, 1, 2, ...$, counts the months since the beginning of treatment. We compare treatment versus waiting (nonparticipation in the stratum). Note that the potential outcomes $Y^1(u,\tau)$, $Y^0(u,\tau)$ differ by the month u when treatment starts. The outcomes condition upon being unemployed at least until month u. Nonparticipation involves the possibility of treatment in a later stratum which implies that $Y^0(u,\tau)$ may correspond after a while to a post treatment outcome.

The treatment parameter we actually estimate is the average within a stratum

$$\theta(\tau) = \sum_u g_u \theta(u,\tau)\ ,$$

where the average is taken with respect to the distribution g_u of starting dates u within the stratum.

Our estimated treatment parameter (3.1) mirrors the decision problem of the case worker and the unemployed who recurrently during the unemployment spell decide whether to start any of the programs now or to postpone participation to the future.

We evaluate the effects of treatment assuming the following dynamic version of the conditional mean independence assumption (DCIA)

$$E(Y^0(u,\tau)|T_u = 1, U \geq u-1, T_1 = ... = T_{u-1} = 0, X) \tag{3.2}$$

$$= E(Y^0(u,\tau)|T_{\tilde{u}} = 0 \text{ for } u \leq \tilde{u} \leq \bar{u}, U \geq u-1, T_1 = ... = T_{u-1} = 0, X)\ ,$$

where X are observed characteristics that are time-invariant within an unemployment stratum and $\tau \geq 0$, see equation (3.1) above and the analogous discussion in Sianesi (2004, p. 137). $T_{\tilde{u}} = 0$ indicates nonparticipation between u and $\bar{u}$ ($\bar{u}$ is the end of the stratum of elapsed unemployment considered). We effectively assume that conditional on X, conditional on being unemployed at least until period $u-1$, and conditional on not receiving any treatment before the end of the stratum considered, $\bar{u}$, individuals are comparable in their nonparticipation outcome.

In our application, we apply propensity score matching building on Rosenbaum and Rubin's (1983) result on the balancing property of the propensity score in the case of a binary treatment. To account for the dynamic treatment assignment, we estimate the probability of treatment given that unemployment lasts long enough to make an individual 'eligible'. For treatment starting during months 1 to 6 (stratum 1), we take the total inflow sample of unemployed, and estimate a Probit model for treatment during stratum 1. The nonparticipation group includes those unemployed who either never participate in the treatment or who start a treatment after month 6. For treatment during strata 2 and 3, the basic sample consists of those unemployed who are still unemployed in the last month of the previous stratum. Implicitly, we assume that the actual beginning of treatment within a stratum is random conditional on X.

We implement a stratified local linear matching approach by imposing that the matching partners for a treated individual are still unemployed in the month before treatment starts, i.e. we exactly align treated and nontreated individuals by elapsed unemployment duration in months. In addition, we exactly align treated and controls by the calendar month in which the unemployment spell began. The expected counterfactual outcome for nonparticipation is obtained by means of a local linear regression on the propensity score. We use a crossvalidation procedure to obtain the bandwidth minimizing the squared prediction error for the average of the nonparticipation-outcome for the nearest neighbors of the treated individuals.[12] An estimate for the variance of the estimated treatment effects is obtained through bootstrapping based on 250 resamples. We resample individuals. This way, we take account of the sampling variability in the estimated propensity score and we obtain standard errors which are clustered at the individual level.

As a balancing test, we use the regression test suggested in Smith and Todd (2005) to investigate whether the covariates are balanced sufficiently by matching on the estimated propensity score. For this purpose, each regressor in a given propensity score specification is regressed on a flexible polynomial of the predicted propensity score and interactions of this polynomial with the treatment dummy. We then determine the number of covariates in each specification for which the balancing test passes, i.e. the zero hypothesis that the polynomial of the propensity score interacted with the treatment dummy equals zero is not rejected. Furthermore, we investigate whether treated and matched nontreated individuals differ significantly in their outcomes before the beginning of the unemployment spell. We estimate these differences in the same way as the treatment effects after the beginning of the program. By construction, treated individuals and their matched counterparts exhibit the same unemployment duration until the beginning of treatment.

We also investigate effect heterogeneity of the ATT over calendar time. For all treated individuals, we calculate the cumulated individual treatment effects by summing the individual monthly effects over the post-treatment time period. We then run a linear regression of these individual effects on dummy variables for the different calendar years.

Finally, we need to discuss the plausibility of the DCIA (3.2) for our application. As Sianesi (2004), we argue that the participation probability depends upon the variables determining re-employment prospects once unemployment begins. Consequently, all individuals are considered as matching partners who have left employment during the same time as the treated individuals (i.e. unemployment started in

[12]This method is an extension of the crossvalidation procedure suggested in Bergemann et al. (2008).

the same calendar month) and who have experienced the same elapsed unemployment duration before program participation. Furthermore, we included a rich set of individual characteristics and detailed information on previous employment experience in the propensity score estimation. For example, we consider skill information, regional information, occupational status, industry as well as information on the remaining entitlement period to unemployment benefits. We use detailed information on past employment and unemployment spells to proxy for 'soft factors' that may influence participation such as the ability or motivation of the unemployed. As participation occurred at a fairly large scale, we argue further that assignment was not very targeted and driven by the regional supply of programs. Moreover, caseworkers had little guidance on 'what works for whom'. Supporting our point of view, Schneider et al. (2006) suggest that until the end of 2002 assignment to training was strongly driven by the supply of available courses.[13]

3.5 Empirical Results

3.5.1 Estimation of the Propensity Scores

We fitted the propensity scores separately for each of the 18 groups. In each case, we run an extensive specification search. The final specification was chosen based on economic considerations, statistical significance of the variables included, and the balancing tests described above.[14] The final specifications include 15 to 31 covariates. The Smith and Todd (2005) balancing test is passed in almost all cases at a 1% significance level, except for one specification where we reject the null hypothesis for one regressor when using the quartic of the propensity score. Even regarding the 5% level we still pass 895 of 928 tests (both cubic and quartic regressions counted).

A closer look at the estimation results for the propensity scores reveals that the following information is particularly relevant: region, age, schooling degree, professional qualification, family status, children, foreign or German nationality, time

[13]For the evaluation of the employment effects of job creation schemes in 1999/2000 based on administrative data for Germany, Caliendo et al. (2004) were able to use a survey asking about the motivation of participants (such information is not available for our data). It turned out that both using administrative data and controlling for these motivational variables did not result in noticeably different estimated program effects compared to using administrative data only. This evidence also supports our point of view.

[14]Detailed estimation results are available in the additional appendix to this paper.

spent in different labor market states during the last three years, remaining claim on unemployment benefits, industry of last employment, last occupation, last wage, reason for the end of last employment, year or quarter the person became unemployed in, health status, past health problems, information on whether a program was canceled within the last three years, penalties and disqualification from benefits within the last three years, participation in a program with a social work component, indication of lack of motivation within the last three years.[15]

3.5.2 Estimated Treatment Effects

The evaluation results for short-term training as first training program vs. waiting are shown in figures 3.1 to 3.3. Each graph displays the average treatment effect on the treated (ATT), i.e. the difference between the actual and the counterfactual employment outcome averaged over those individuals who participate in the program under consideration. More precisely, we compare the actual employment outcome of the treated to the employment outcome these individuals would have had, *had they not taken part in short-term training as a first training program in the respective time window of their unemployment spell.* We distinguish between programs starting in three different time windows (strata) of elapsed unemployment: 0 to 6 months (stratum 1), 7 to 12 months (stratum 2), and 13 to 24 months (stratum 3). We evaluate treatment effects at different points in time. On the time axis in our graphs, positive values denote months since the program start, while negative values represent pre-unemployment months. We omit the period between the start of unemployment and the start of the program where both control and treatment group are unemployed. The dashed lines around the estimated ATT are bootstrapped 95 percent confidence bands. Treatment effects for a particular month are statistically significant if zero is not contained in the confidence band.

Figure 3.1 shows the estimated treatment effects for the short-term training programs in the early 2000s with a strong focus on qualification (QST00). The results for men are given in the left column, while those for women are shown in the right column. During the program and in the first time following the end of the program,

[15]The variables family status, reason for the end of last employment, health status, past health problems, penalties and disqualification from benefits within the last three years, participation in a program with a social work component, indication of lack of motivation within the last three years cannot be generated in the older data. We instead hope to capture the information by using detailed variables on the individuals' labor market history.

participants typically have a lower monthly employment probability than they would have had if they had not participated in the program. This is the so-called lock-in effect. Figure 3.1 suggests relatively short (1 to 4 months) and not very pronounced lock-in effects. These lock-in effects are a bit deeper for stratum 1 (about 7 percentage points) than for the later strata (2 to 4 percentage points). After the short lock-in period, the difference between actual and counterfactual employment outcomes of participants turns positive. We find significant positive effects for men in the second and third stratum (i.e. those men who have been unemployed for at least half a year before entering a program) and for women in the first and second stratum, but not in the third stratum. However, the point estimates for the latter are positive after six months. The largest employment effects occur between month 12 and month 18. In the four groups with significant effects, the size of the effects reaches 9 to 17 percentage points. After 18 months, the effects tend to decline a little, but positive ATTs of 7 to 12 percentage points persist until the end of the observation period (18 to 36 months after program start depending on the stratum). These long-lasting effects are quite remarkable given that the programs last only a few weeks. As we do not exclude participants who attend a second training program after short-term training, but regard the second program as an outcome, it could well be that the long-term effects are to some degree due to longer training programs which have been started as a result of the short-term training program. This would imply that short-term training serves as a bridge into a more intensive training program and this combination eventually leads to positive employment effects.

Figure 3.2 presents the corresponding results for the short-term training programs in the early 2000s we classified as having a strong focus on testing the availability and willingness to work as well as the skills of job seekers (MST00). The graphs suggest that, while the point estimates of the monthly average treatment effects are mostly positive, they generally fail to be clearly significant. The only exception are women who receive treatment in months 7 to 12 (stratum 2) of their unemployment spell. After a small and very short lock-in-effect, we first observe a small and insignificant positive effect. Rising steadily over more than two years, it turns significant after 9 months and eventually reaches 16 percentage points. This picture fits into the scenario that part of the participants attend a second program as a result of the MST00 program and this combination of programs eventually may lead to positive employment effects. The trend of the treatment effect is similar for men and women in stratum 1, but in these cases the level is much lower and the effect is insignificant. The lock-in-effects show a similar picture for MST00 as for QST00. In sum, participants of MST00 seem to have benefited less from their program than participants of QST00 from the program they were assigned to.

Results for short-term programs in the period 1980-1992 (ST8092) are given in figure 3.3. The estimated monthly average treatment effects are positive after an initial lock-in-effect. Remarkably, the lock-in period is typically longer for ST8092 than for ST00. Also, the monthly ATTs of ST8092 are mostly smaller than those of QST00 and statistically insignificant. Only for women unemployed for more than one year (stratum 3), the results show significantly positive treatment effects of 7 to 10 percentage points between month 6 and month 20 after program start. For the other groups the effects are - though always being positive for the time after the lock-in period - insignificant. Interestingly, for most groups the employment effect increases between month 18 and month 26 after treatment start. As discussed before, this is likely due to participation in another training program as a result of participation in short-term training. In sum, ST8092 programs were less successful in bringing people back to employment compared to ST00, in particular to QST00.

Table 3.5 shows averages of the monthly ATTs from month 6 after program start until the end of the observation period as a way of condensing employment effects after the end of the lock-in period. In four cases for QST00 and two cases for MST00, the figures reported in table 3.5 suggest highly significant employment effects between 6 and 14 percentage points, for the other groups the effects are smaller and not (or only slightly) significant. The results for the ST8092 programs suggest significantly positive ATTs for women in stratum 2 and 3 and for men in stratum 2 in the range from 6.2 to 7.4 percentage points despite the mostly insignificant point estimates in figure 3.3. The effects for the three other cases are smaller in size and not significant.

Table 3.4 shows gains and losses in months employed cumulated over up to two years (four years for the ST8092 programs, respectively) after program start as a way of condensing the graphical results in figures 3.1 to 3.3. This measures by how much a participation in short-term training increases the time spent in employment in a given time period, when initial negative and subsequent positive employment effects are weighed against each other. While the gains are very small or not even positive over the first 6 months, they increase for most groups over a longer period. The positive employment effects of QST00 are confirmed again. The effects of QST00 cumulated over 24 months are in general larger than those of MST00 and ST8092. In the cases where we find significant effects after 24 months, these lie in a range between one and two and a half months. For example, men and women participating in QST00 after having been unemployed between 7 and 12 months (stratum 2) gain 2.4 months in employment during the first 24 months after the program start. Women participating in MST00 in stratum 2 gain 2.6 months in employment in two years. For the ST8092 programs, there are surprisingly high gains for women who were long-term unemployed before the program. After 48 months, we find significantly positive cumulated employment effects of ST8092 programs (in one case significance is given only at the 10% level) for women in stratum 2 and 3 and for men in stratum 2.

Next, table 3.6 reports the estimated treatment effects on the participation rates. This means that instead of the employment effect we estimate the average effect of the short-term training program on the probability to participate in a longer-term further training program at least once during the remaining unemployment spell after the start of the short-term training program. All effects are positive and, with the exception of QST00 in stratum 2, they are all significant. Incidently, the employment effects for QST00 in stratum 2 are the highest among all QST00 cases while the participation effects are the lowest. In most cases, the participation effects are higher for women than for men and the effects are mostly higher for the ST8092 programs compared to short-term training in the 2000s. Furthermore, the effects are much higher for MST00 compared to QST00. This is in line with what one would expect: an important goal of MST00 is to define a path back into employment, including for some job seekers participation in a more intensive training program. A limited skill upgrade to directly enhance placement is a strong focus of QST00, as a result future program participation is a bit less of an issue. For the ST8092 programs, guiding needy job seekers into a long-term training program was an official goal. The estimated effects of participation rates reflect this goal, in particular for female participants who show an 18 to 25 percentage points higher probability to participate in a longer training program.

This study investigates program effects of ST8092 programs over 13 calendar years. Given this very long period, one could suspect that the employment effects differ over calendar time. Possibly, the activation effect of such programs is higher for the hard-to-place when unemployment is low or the programs give the unemployed an additional edge when unemployment is high. In order to investigate this type of potential effect heterogeneity, we regress the average individual treatment effects after the lock-in period (summarized in table 3.5) on an intercept, year dummies, and the individual elapsed unemployment duration to investigate whether the effects differ between years. Bootstrap standard errors are calculated based on the resamples which are also used to bootstrap the standard errors of the effect estimates. According to the results of these regressions, the ATTs do not differ over time: a chi-square test for joint significance of the year dummies does not suggest any effect heterogeneity (see table 3.7). Thus, there is no evidence for the business cycle affecting the employment effects of short-term training, a finding which is in contrast to the results for longer training programs in Lechner and Wunsch (2006).

3.6 Conclusions

Recently, there has been a greater emphasis on job search assistance, monitoring and testing work availability, as well as limited training to activate the unemployed (OECD 2007). In Germany, the focus on activation strategies is reflected in the recent shift away from traditional longer further training programs typically lasting a couple of months up to 2 years to short-term training programs (ST00) lasting at most 12 weeks. In fact, ST00 have become the largest program in Germany regarding the number of participants with 1.07 million individuals entering such a program in 2007 (Bundesagentur für Arbeit, 2007, pp. 54, 57). Between 1980 and 1992, a similar large-scale short-term training program was in place in Germany. These were the 'programs according to §41a Employment Promotion Act' (ST8092).

This paper estimates the effects of short-term training programs in West Germany both for the time period 1980 to 1992 and 2000 to 2003 regarding both future employment and future participation in longer training programs. This is the first paper to analyze these programs for the earlier time period and to estimate long-run effects on outcomes. Our results show that short-term training shows mostly persistently positive and often significant employment effects. The effects are particularly strong when participation starts during months 7 to 12 of the unemployment spell. We tend to find smaller effects for short-term training starting during the second year of the unemployment spell. When short-term training focuses on testing and monitoring search effort, there are slightly smaller effects compared to when the focus is on training only. The lock-in periods lasted longer in the 1980s and 1990s compared to the early 2000s. Short-term training results in higher future participation in longer training programs and this effect was much stronger for the earlier time period. The employment effects of the ST8092 programs did not change significantly by year between 1980 and 1992, i.e. there is no evidence for business cycle effects in contrast to the results for longer training programs in Lechner and Wunsch (2006).

Our findings most likely reflect a change in active labor market policy between 1992 and 2000. In the 2000s, there is a strong focus on activating the unemployed. In contrast, in the 1980s and 1990s it was accepted policy to 'give the unemployed some time' and to encourage them to participate in longer training programs when this seemed advisable and the unemployed were hard to place. Our results suggest that the policy reorientation towards activation did not result in worse employment outcomes. If anything, as far as comparable, ST00 programs with a focus on training show better employment effects. As a caveat, we have to acknowledge, however, that the estimated treatment effects for the two time periods are obtained for different selective treatment samples, i.e. the effects cannot be compared without accounting for these differences.

The fact that we find some long-lasting effects of short-term training may be surprising given their short duration. These programs by themselves do not provide a sizeable human capital investment. Future research should investigate the hypothesis that the positive program effects can be traced back to the higher participation rates in longer training programs. However, as one piece of evidence against this hypothesis, we find that in the two cases with no significant participation effects the employment effects of ST00 are particularly high. However, a thorough investigation of the hypothesis would require an evaluation approach for multiple sequential treatments as e.g. the one developed by Lechner (2004), but it remains an open question whether the stringent identifying assumptions required are satisfied in applications like ours. As a final caveat, an overall assessment of the microeconomic effects of short-term training is not possible, because the necessary information for a comprehensive cost-benefit-analysis is lacking in our data.

References

Abbring, J. and G. J. van den Berg (2003). "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica* 71, 1491-1517.

Bender, S., A. Haas, and C. Klose (2000). "IAB employment subsample 1975-1995." *Schmollers Jahrbuch (Journal of Applied Social Science Studies)* 120, 649-662.

Bender, S., M. Biewen, B. Fitzenberger, M. Lechner, R. Miquel, A. Osikominu, M. Waller, C. Wunsch (2005). "Die Beschäftigungswirkung der FbW-Maßnahmen 2000 - 2002 auf individueller Ebene: Eine Evaluation auf Basis der prozessproduzierten Daten des IAB - Zwischenbericht Oktober 2005." Goethe University Frankfurt and SIAW St. Gallen.

Bender, S., A. Bergemann, B. Fitzenberger, M. Lechner, R. Miquel, S. Speckesser, C. Wunsch (2005). "Über die Wirksamkeit von Fortbildungs- und Umschulungsmaßnahmen." *Beiträge zur Arbeitsmarkt- und Berufsforschung* 289, IAB, Nürnberg.

Bergemann, A., B. Fitzenberger, and S. Speckesser (2008). "Evaluating the Dynamic Employment Effects of Training Programs in East Germany Using Conditional Difference-in-Differences." forthcoming *Journal of Applied Econometrics.*

Biewen, M., B. Fitzenberger, A. Osikominu, and M. Waller (2007). "Which Program for Whom? Evidence on the Comparative Effectiveness of Public Sponsored Training Programs in Germany." IZA Discussion Paper, No. 2885, Bonn.

Büttner, T. (2007). "Ankündigungseffekt oder Maßnahmewirkung? Eine Evaluation von Trainingsmaßnahmen zur Überprüfung der Verfügbarkeit." IAB Discussion Paper, 25/2007, Nürnberg.

Bundesagentur für Arbeit (2000-2004). "Arbeitsmarkt." Nürnberg (various issues).

Bundesagentur für Arbeit (2007). "Der Arbeits- und Ausbildungsmarkt in Deutschland - Dezember und das Jahr 2007." Monthly Report, Nürnberg.

Bundesanstalt für Arbeit (1980-1993). "Förderung der beruflichen Bildung/Weiterbildung." Nürnberg (various issues).

Bundesanstalt für Arbeit (1985, 1994). "Amtliche Nachrichten der Bundesanstalt für Arbeit." Nürnberg (various issues).

Caliendo, M., R. Hujer, S.L. Thomsen, and C. Zeiss (2004). "Einfluss von motivationalen Merkmalen aus Befragungsdaten auf die Schätzung des

Beschäftigungseffektes von ABM-Maßnahmen." Unpublished Manuscript, Goethe University Frankfurt.

Dobischat, R. and H. Wassmann (1981). "Die Teilnahme der Arbeitslosen an Maßnahmen zur Verbesserung der Vermittlungsaussichten und ihre Nutzung für die berufliche Weiterbildung." *Berufsbildung in Wissenschaft und Praxis* 6/81, 17-19.

Fay, R. (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, Paris.

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

Fitzenberger, B., A. Osikominu and R. Völter (2008). "Get Training or Wait? Long-Run Employment Effects of Training Programs for the Unemployed in West Germany." forthcoming *Annales d'Economie et de Statistique.*

Fredriksson, P. and P. Johansson (2003). "Program Evaluation and Random Program Starts." Institute for Labour Market Policy Evaluation (IFAU), Uppsala, Working Paper, 2003:1.

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

Hujer, R., S.L. Thomsen and C. Zeiss (2006). "The Effects of Short-Term Training Measures on the Individual Unemployment Duration in West Germany." Discussion Paper 06-065, ZEW Mannheim.

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

Jacobebbinghaus, P. and S. Seth (2007). "The German Integrated Employment Biographies Sample IEBS." *Journal of Applied Social Science Studies (Schmollers Jahrbuch)* 127, 335-342.

Kluve, J. (2006). "The Effectiveness of European Active Labor Market Policy." Discussion Paper, RWI Essen.

Kurtz, B. (2003) "Trainingsmaßnahmen - Was verbirgt sich dahinter?" IAB Werkstattbericht No. 8, Nürnberg.

Lechner, M. (2004). "Sequential Matching Estimation of Dynamic Causal Models." Discussion Paper 2004-06, University of St. Gallen.

Lechner, M., R. Miquel, and C. Wunsch (2004). "Long-Run Effects of Public Sector Sponsored Training in West Germany." IZA Discussion Paper No. 1443.

Lechner, M. and C. Wunsch (2006). "Are Training Programs more effective when Unemployment is high?" IZA Discussion Paper No. 2355.

Lechner, M. and C. Wunsch (2007). "What Did All the Money Do? On the General Ineffectiveness of Recent West German Labour Market Programmes." IZA Discussion Paper No. 2800.

Martin, J.P. and Grubb, D. (2001). "What works and for whom: A review of OECD countries' experiences with active labour market policies." *Swedish Economic Policy Review* 8, 9-56.

OECD (2007). "Activating the Unemployed: What Countries do?" Employment Outlook, Chapter 5, Paris.

Osikominu, A. (2008). "Is Short Training Short-Lived and Long Training Long-Lasting? A Multi-State Duration Analysis of the Dynamic Effects of Training Schemes For the Unemployed." Discussion Paper, *mimeo*, Albert-Ludwigs-University-Freiburg.

Plaßmann, G. (2002). "Der Einfluss der Arbeitslosenversicherung auf die Arbeitslosigkeit in Deutschland." *Beiträge zur Arbeitsmarkt- und Berufsforschung, BeitrAB* 255.

Rosenbaum, P.R. and D.B. Rubin (1983). "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70, 41-55.

Roy, A.D. (1951). "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers* 3, 135-146.

Rubin, D.B. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66, 688-701.

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

Schneider, H., K. Brenke, D. Hess, L. Kaiser, J. Steinwede und A. Uhlendorff (2006). "Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission - Modul 1b: Förderung beruflicher Weiterbildung und Transferleistungen." IZA Research Report, No. 7, Bonn.

Sianesi, B. (2003). "Differential Effects of Swedish Active Labour Market Programs for Unemployed Adults in the 1990s." Discussion Paper, Institute for Fiscal Studies, London.

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

Smith, J.A. and P. Todd (2005). "Rejoinder." *Journal of Econometrics* 125, 365-375.

Stephan, G., S. Rässler, and T. Schewe (2006). "Das TrEffeR-Projekt der Bundesagentur für Arbeit - die Wirkung von Maßnahmen aktiver Arbeitsmarktpolitik." Zeitschrift für ArbeitsmarktForschung, 39(3/4), 447-465.

Appendix

Table 3.1: Entries into Active Labor Market Programs in West Germany from 1979-1992 (in Thousand)

	1979	1980	1981	1982	1983
Training programs	209	247	280	266	306
– further training	149	162	190	189	220
– short-term training	0.7	14	25	23	24
– retraining	31	38	47	42	42
– job training	29	33	17	11	20
Job creation schemes	20	15	14	8	23
	1984	1985	1986	1987	1988
Training programs	353	409	530	596	566
– further training	260	298	374	420	420
– short–term training	30	38	52	63	29
– retraining	43	45	59	65	66
– job training	19	28	45	49	51
Job creation schemes	26	34	41	42	38
	1989	1990	1991	1992	
Training programs	490	574	594	575	
– further training	361	383	421	418	
– short-term training	27	59	53	47	
– retraining	61	63	70	81	
– job training	41	68	49	29	
Job creation schemes	28	27	28	18	

Source: Bundesanstalt für Arbeit (1980-1993), Bundesanstalt für Arbeit (1985, 1994).

Table 3.2: Entries into Active Labor Market Programs in Germany and West Germany from 1999–2004 (in Thousand)

	1999	2000	2001	2002	2003	2004
	Germany					
Qualification programs	1,108	1,221	1,069	1,537	1,502	1,548
– further and retraining	491	552	450	456	255	185
– short-term training	432	477	565	877	1,064	1,188
Employment subsidies	538	459	465	544	808	950
Placement and advisory services	532	601	742	934	2,920	5,134
Job creation schemes	353	314	246	220	194	170
Specific measures for youths	244	263	265	294	389	408
Other	312	391	516	457	212	309
Total	3,087	3,249	3,304	3,985	6,025	8,520
	West Germany					
Qualification programs	714	770	643	972	985	1,038
– further and retraining	307	338	261	273	161	124
– short-term training	265	286	339	545	690	789
Employment subsidies	245	225	206	245	365	481
Placement and advisory services	286	279	296	375	1,281	2,797
Job creation schemes	96	89	73	63	39	42
Specific measures for youths	181	193	191	210	262	270
Other	231	296	370	345	17	175
Total	1,753	1,852	1,778	2,210	2,949	4,803

Source: Bundesagentur für Arbeit (2001, 2002, 2003, 2004, 2005).

Table 3.3: Participation in short-term Training as a First Training Program for the Inflow Samples into Unemployment

Stratum	Months 1 to 6	Months 7 to 12	Months 13 to 24
	ST8092		
Male Participants	165	201	183
Male Control Group	59921[a]	25674[a]	15631
Female Participants	145	145	167
Female Control Group	35782[a]	22970	17020
	ST00		
Male Participants QST00	559	221	211
Male Participants MST00	531	177	214
Male Control Group	20979	8337	5122
Female Participants QST00	537	214	130
Female Participants MST00	325	126	115
Female Control Group	13848	7070	4975

[a]For these three groups, we randomly selected half of the available non-participants due to computer constraints.

Table 3.4: Cumulated Treatment Effects

QST00, Men			
	6 months	12 months	24 months[a]
Stratum 1	-0.145 (0.090)	-0.015 (0.187)	0.262 (0.378)
Stratum 2	0.274 (0.139)**	1.005 (0.308)***	2.429 (0.657)***
Stratum 3	0.319 (0.108)***	0.799 (0.241)***	1.518 (0.409)***

QST00, Women			
	6 months	12 months	24 months[a]
Stratum 1	-0.080 (0.083)	0.369 (0.191)*	1.197 (0.422)***
Stratum 2	-0.009 (0.139)	0.669 (0.313)**	2.419 (0.699)***
Stratum 3	-0.108 (0.089)	0.013 (0.254)	0.224 (0.482)

MST00, Men			
	6 months	12 months	24 months[a]
Stratum 1	-0.142 (0.089)	-0.167 (0.191)	-0.175 (0.403)
Stratum 2	-0.029 (0.115)	0.142 (0.277)	0.378 (0.560)
Stratum 3	0.166 (0.096)*	0.557 (0.236)**	0.984 (0.393)**

MST00, Women			
	6 months	12 months	24 months[a]
Stratum 1	-0.203 (0.102)**	-0.322 (0.230)	-0.044 (0.452)
Stratum 2	0.223 (0.162)	0.907 (0.404)**	2.590 (0.896)***
Stratum 3	0.038 (0.107)	0.265 (0.277)	0.533 (0.498)

ST8092, Men				
	6 months	12 months	24 months	48 months
Stratum 1	-0.448 (0.143)***	-0.343 (0.315)	-0.050 (0.631)	1.308 (1.274)
Stratum 2	-0.285 (0.118)**	-0.004 (0.300)	0.910 (0.627)	2.336 (1.168)**
Stratum 3	-0.042 (0.116)	0.161 (0.258)	0.522 (0.552)	1.271 (1.236)

ST8092, Women				
	6 months	12 months	24 months	48 months
Stratum 1	-0.150 (0.142)	0.039 (0.342)	0.289 (0.770)	0.917 (1.501)
Stratum 2	-0.302 (0.154)*	0.019 (0.389)	0.910 (0.844)	2.813 (1.595)*
Stratum 3	0.130 (0.104)	0.613 (0.245)**	1.675 (0.543)***	3.124 (1.122)***

[a] In stratum 3 the treatment effects are summed over 19 months.

Note: Sum of the monthly treatment effects from month zero (program start). *** = statistically significant at 1%, ** = at 5%, * = at 10%, bootstrapped standard errors.

Table 3.5: Average ATT after Lock-in Period

QST00		
	Men	Women
Stratum 1	0.028 (0.016)*	0.070 (0.020)***
Stratum 2	0.111 (0.031)***	0.130 (0.034)***
Stratum 3	0.092 (0.026)***	0.025 (0.032)

MST00		
	Men	Women
Stratum 1	0.014 (0.019)	0.024 (0.019)
Stratum 2	0.027 (0.025)	0.143 (0.042)***
Stratum 3	0.063 (0.025)**	0.038 (0.032)

ST8092		
	Men	Women
Stratum 1	0.043 (0.029)	0.025 (0.034)
Stratum 2	0.062 (0.027)**	0.074 (0.036)**
Stratum 3	0.030 (0.028)	0.071 (0.026)***

Note: Average of the monthly treatment effects from month six since program start until the end of the observation period (until month 48 for ST8092 programs). *** = statistically significant at 1%, ** = at 5%, * = at 10%, bootstrapped standard errors.

Table 3.6: ATT for Participation Rates

QST00		
	Men	Women
Stratum 1	0.083 (0.018)***	0.063 (0.018)***
Stratum 2	0.016 (0.025)	0.025 (0.031)
Stratum 3	0.047 (0.024)**	0.106 (0.036)***

MST00		
	Men	Women
Stratum 1	0.164 (0.020)***	0.175 (0.026)***
Stratum 2	0.183 (0.032)***	0.200 (0.045)***
Stratum 3	0.132 (0.027)***	0.110 (0.034)***

ST8092		
	Men	Women
Stratum 1	0.092 (0.033)***	0.180 (0.038)***
Stratum 2	0.218 (0.034)***	0.254 (0.044)***
Stratum 3	0.194 (0.041)***	0.229 (0.037)***

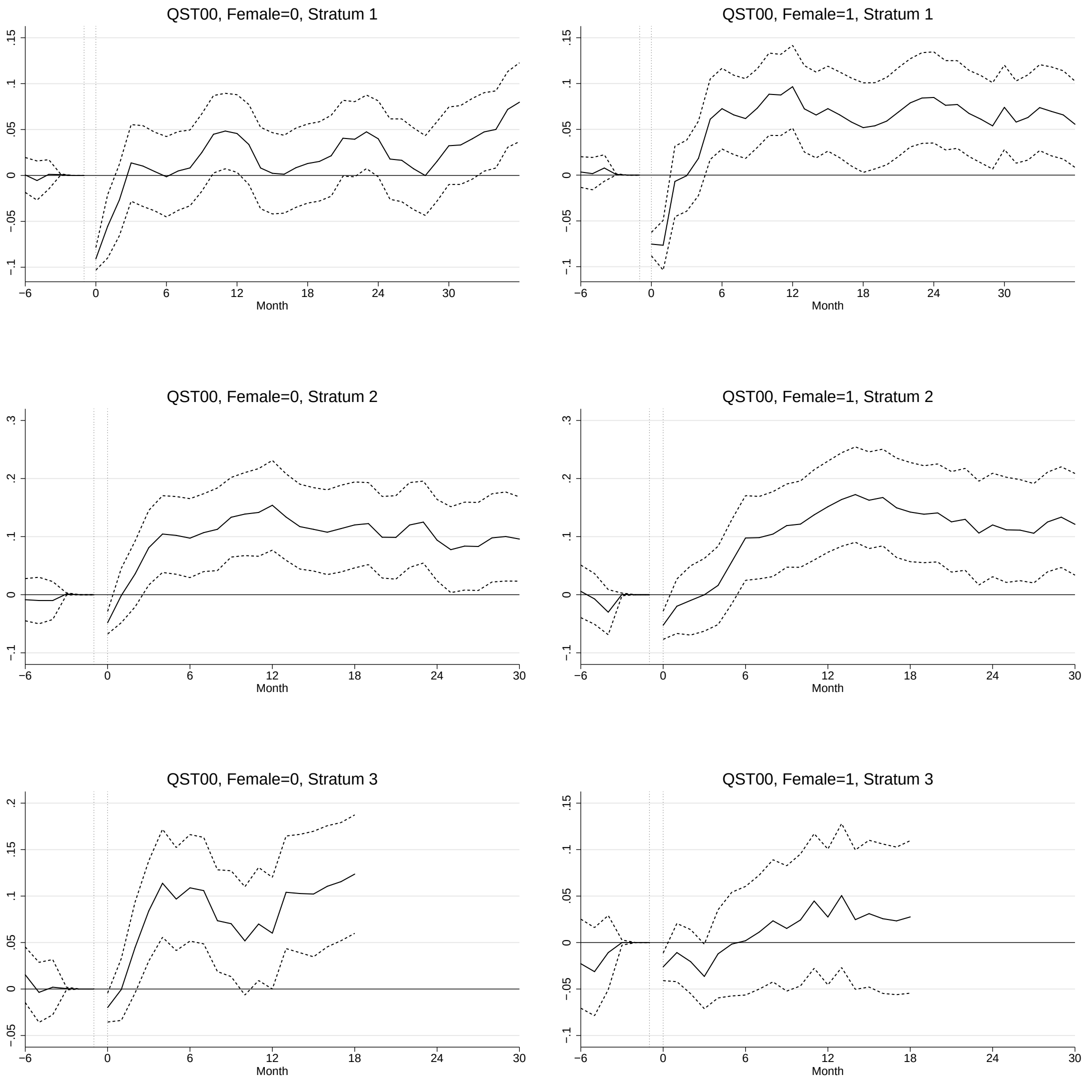
Note: Treatment effects until the end of the observation period (until month 48 for ST8092 programs) *** = statistically significant at 1%, ** = at 5%, * = at 10%, bootstrapped standard errors.

Table 3.7: Test of Heterogeneity of Employment Effects over Time

χ^2–Statistic (p-Value)		
	Men	Women
Stratum 1	8.06 (0.701)	2.91 (0.992)
Stratum 2	15.94 (0.143)	12.79 (0.235)
Stratum 3	10.32 (0.502)	7.45 (0.762)

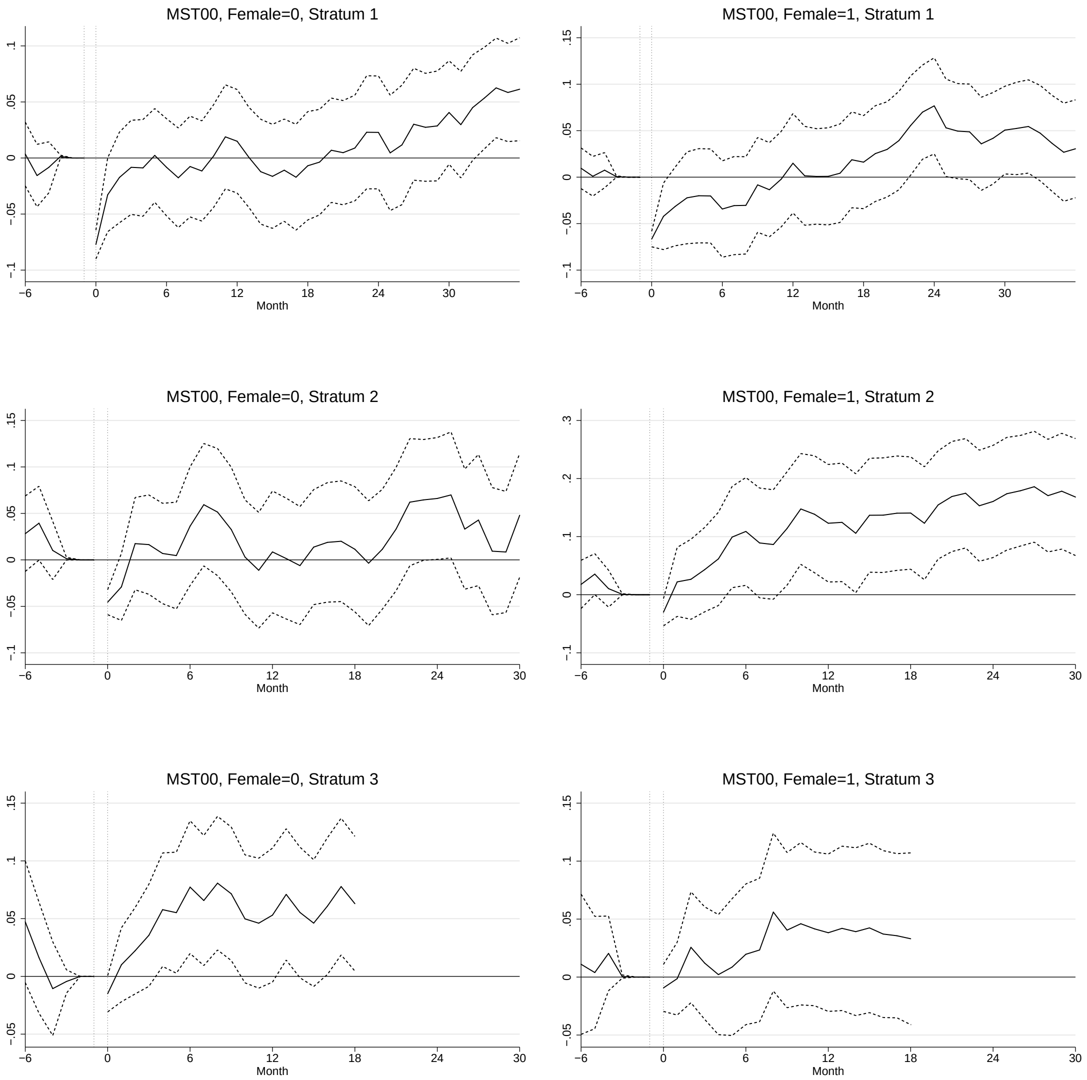
Note: Test on joint significance of all year dummies in a regression of the individual treatment effects averaged over the months after program start (see table 3.5) on an intercept, year dummies, and elapsed unemployment duration. Empirical standard errors are calculated from bootstrap resamples.

Figure 3.1: Average Treatment Effect on the Treated (ATT) QST00



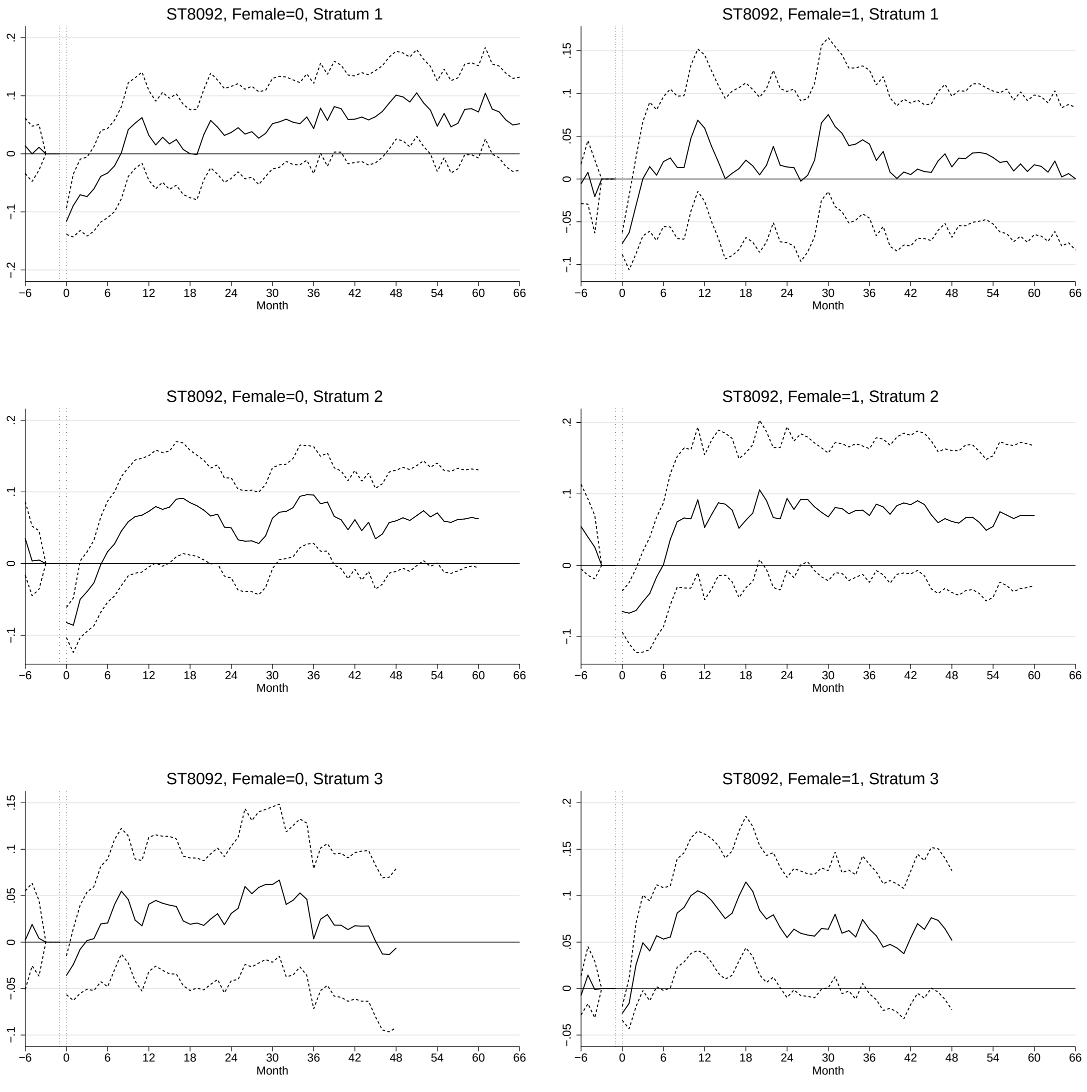
Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa.

Figure 3.2: Average Treatment Effect on the Treated (ATT) MST00



Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa.

Figure 3.3: Average Treatment Effect on the Treated (ATT) ST8092



Difference in employment rates is measured on the ordinate, pre-unemployment (< 0) and post-treatment (≥ 0) months on the abscissa.

Additional Appendix to: Déjà Vu? Short-Term Training in Germany 1982-1992 and 2000-2003

Note: The following material is made available as additional information for the paper "Déjà Vu? Short-Term Training in Germany 1982-1992 and 2000-2003".

Table 3.8: Variable Definitions for the 2000-2003 Sample

Label	Definition
Personal Attributes	
female	1 if female, 0 otherwise
agegroup	age in 6 groups
foreigner	1 if citizenship is not German, 0 otherwise
qualification	1 no degree, 2 vocational training degree, 3 university or technical college degree
schooling	1 no schooling degree, 2 Hauptschulabschluss or Mittlere Reife /Fachoberschule (degrees reached after completion of the 9th or 10th grade), 3 Fachhochschulreife or Abitur/Hochschulreife (degrees reached after completion of the 12th or 13th grade)
health	1 no health problems mentioned, 2 health problems, but considered without impact on placement, 3 health problems considered to have an impact on placement
pasthealth	same categories as health, but referring to the past two years before the beginning of the unemployment spell
disabled	1 if disabled, 0 otherwise
married	1 missing, 2 married, 3 not married
child	1 if at least one child, 0 otherwise
youngchild	1 if at least one child younger than 10 years, 0 otherwise
Last Employment	
occupation	occupation of last employment in 7 categories
industry	industry of last employment in 6 categories
endlastjob	2 termination of last occupation by employer, 3 by employee, 4 limited in time, 5 other and missing

<continued on next page>

Label	Definition
waged	daily wage in the last job(s) before the beginning of the unemployment spell
ddssec	ddsec is 1 if earnings are within the social security thresholds
lnwage	log(waged) interacted with ddssec
Employment and Program History	
problemgroup	1 if participation in a program with a social work component within the last three years, 0 otherwise
pasttreatcancel	1 if abandonment of a program in the past according to the benefit data, 0 otherwise
penalty	1 if the unemployed had a period of disqualification from benefits within the last three years, 0 otherwise
motivationlack	1 if within the last three years there is information, that the person did not appear regularly at the labor office, on lack of cooperation, availability or similar
countemp, countub, countua, countsub, countoos, countcon	number of days within the last three years before the beginning of unemployment spent in regular employment, receiving unemployment benefits, unemployment assistance, subsistance payment, out of sample, in contact with the labor office, respectively
dcount...	1 if the respective count variable is larger than 0, 0 otherwise
demp6, demp12, demp24, demp6_12, demp12_24	1 if in regular employment 6, 12, 24, 6 and 12 and 12 and 24 months, respectively, before the beginning of the unemployment spell
claimg	remaining claim on unemployment benefit in four categories
Regional Information	
area	German Bundesländer aggregated into 6 categories. 1 SH, NI, HB, HH; 2 NW, 3 HE, RP, SL; 4 BY, BW; 5 MV, BB, BE; 6 SN, ST, TH
region	classification of the districts of residence according to local labor market conditions in 5 groups
Calendar Time of Entry into Unemployment	
quarter	quarter of the end of the last employment (from 1 to 6)

Note: If not mentioned otherwise, variables are defined relative to the beginning of the time window of elapsed unemployment duration. Variables in categories are used as dummies, i.e. agegroup1 is 1 if agegroup takes the value 1 and 0 otherwise.

Estimated Propensity Scores for the 2000-2003 Sample

Table 3.9: Participation Probit for QST00, Males

	Stratum 1	Stratum 2	Stratum 3
agegroup1		-0.166 (0.098)*	-0.160 (0.120)
agegroup12	0.001 (0.041)		
agegroup2		-0.123 (0.088)	-0.090 (0.102)
agegroup4		-0.002 (0.090)	-0.082 (0.107)
agegroup5		-0.080 (0.105)	0.214 (0.102)**
agegroup6		0.140 (0.106)	0.145 (0.113)
area2		-0.082 (0.081)	-0.151 (0.085)*
area3		-0.008 (0.097)	-0.125 (0.113)
area4		-0.129 (0.112)	-0.319 (0.130)**
child	0.116 (0.048)**	0.087 (0.067)	0.179 (0.071)**
claimg0		-0.031 (0.105)	0.043 (0.095)
claimg1	0.130 (0.103)		
claimg2	0.116 (0.088)	-0.124 (0.085)	-0.072 (0.101)
claimg3	0.096 (0.098)		
claimg34		-0.383 (0.116)***	-0.210 (0.119)*
claimg4	0.169 (0.108)		
countcon	0 (0.000)	0 (0.000)	0 (0.000)*
countemp	0 (0.000)**	0.001 (0.000)***	0 (0.000)
dcountcon	0.068 (0.053)	0.021 (0.079)	0.029 (0.086)
dcountoos		0.011 (0.078)	-0.148 (0.086)*
dcountsub	0.214 (0.071)***		
ddssec	0.607 (0.242)**		
demp12_24	0.195 (0.074)***	-0.045 (0.115)	0.055 (0.144)
demp24	-0.150 (0.076)**	0.010 (0.121)	0.009 (0.144)
demp6	0.026 (0.063)	-0.033 (0.097)	0.292 (0.114)**
endlastjob2	0.045 (0.046)		
endlastjob4	0.236 (0.064)***		
industry3		-0.162 (0.086)*	-0.267 (0.095)***
industry5		0.095 (0.078)	-0.192 (0.099)*
industry6		-0.027 (0.097)	-0.151 (0.104)
lnwaged	-0.135 (0.053)**		
married2	0.165 (0.044)***	0.167 (0.065)**	0.281 (0.071)***

	Stratum 1	Stratum 2	Stratum 3
motivationlack	0.105 (0.055)*		
occupation1	0.102 (0.073)		
occupation3	-0.013 (0.065)		
occupation5	0.243 (0.080)***		
occupation6	0.139 (0.086)		
occupation7	-0.002 (0.105)		
problemgroup	0.224 (0.092)**		
quarter1	-0.199 (0.054)***	-0.196 (0.105)*	-0.154 (0.104)
quarter2		0.029 (0.104)	-0.211 (0.117)*
quarter3		0.091 (0.095)	-0.182 (0.110)*
quarter4	-0.120 (0.051)**	-0.062 (0.094)	-0.122 (0.100)
quarter5	-0.086 (0.049)*		
quarter6		0.121 (0.093)	-0.115 (0.101)
region2		-0.237 (0.097)**	-0.092 (0.096)
region4		0.090 (0.104)	0.157 (0.115)
region5		0.038 (0.095)	0.225 (0.115)*
schooling3	0.143 (0.060)**		
youngchild	-0.048 (0.062)		
_cons	-2.633 (0.169)***	-2.354 (0.231)***	-1.961 (0.265)***
N	21538	8558	5333

Table 3.10: Results for Smith and Todd (2005) Balancing Test, QST00 Males

Treatment QST00, Female=0, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	28	29	30	30
Stratum 2	27	30	31	31
Stratum 3	28	30	31	31
Treatment QST00, Female=0, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	28	28	30	30
Stratum 2	26	27	31	31
Stratum 3	31	31	31	31

Table 3.11: Participation Probit for MST00, Males

	Stratum 1	Stratum 2	Stratum 3
agegroup1		0.063 (0.095)	
agegroup2		0.231 (0.080)***	
agegroup56	-0.033 (0.050)	0.006 (0.098)	
child		0.189 (0.072)***	0.115 (0.068)*
claimg0	-0.027 (0.079)	-0.059 (0.215)	0.142 (0.097)
claimg0_dcountoos		0.370 (0.206)*	
claimg1	0.053 (0.084)	-0.040 (0.195)	-0.085 (0.108)
claimg1_dcountoos		0.286 (0.204)	
claimg2		0.221 (0.107)**	
claimg34	0.008 (0.047)		-0.124 (0.117)
countub			-0 (0.000)
dcountcon	0.091 (0.054)*		0.129 (0.084)
dcountoos	0.064 (0.042)	-0.169 (0.091)*	
dcountsub			0.237 (0.112)**
dcountub	-0.135 (0.046)***		
demp12		0.005 (0.151)	0.029 (0.138)
demp24		-0.029 (0.080)	0.054 (0.073)
demp6_12	0.124 (0.050)**	0.106 (0.153)	-0.127 (0.133)
endlastjob2	0.102 (0.046)**		
endlastjob3	-0.172 (0.117)		
endlastjob4	0.135 (0.071)*		
foreigner		-0.291 (0.096)***	
health2		0.284 (0.127)**	
health3		-0.110 (0.135)	
industry2	0.520 (0.171)***		
industry3	0.314 (0.171)*		-0.093 (0.099)
industry4	0.432 (0.170)**		-0.067 (0.085)
industry5	0.500 (0.173)***		-0.150 (0.103)
industry6	0.350 (0.180)*		-0.321 (0.121)***
married2	0.129 (0.042)***	0.220 (0.069)***	0.234 (0.071)***
motivationlack	0.109 (0.055)**		-0.027 (0.078)
pasthealth1			0.346 (0.123)***
pasthealth2		-0.063 (0.172)	

	Stratum 1	Stratum 2	Stratum 3
pasthealth3		0.312 (0.149)**	
penalty			0.219 (0.126)*
qualification1	0.107 (0.041)***	0.182 (0.068)***	0.041 (0.066)
quarter1	0 (0.070)		
quarter3	0.008 (0.075)		
quarter4	-0.056 (0.071)		0.098 (0.088)
quarter5	-0.034 (0.070)		0.186 (0.084)**
quarter6	0.161 (0.071)**		0.141 (0.092)
region2	0.536 (0.079)***		
region3	0.278 (0.074)***	-0.037 (0.086)	-0.140 (0.082)*
region4		-0.194 (0.126)	-0.258 (0.121)**
region5	0.122 (0.080)	-0.169 (0.102)*	-0.283 (0.101)***
schooling3	0.090 (0.057)		
_cons	-2.961 (0.207)***	-2.406 (0.170)***	-2.176 (0.206)***
N	21510	8514	5336

Table 3.12: Results for Smith and Todd (2005) Balancing Test, MST00 Males

Treatment MST00, Female=0, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	26	28	28	28
Stratum 2	23	23	23	23
Stratum 3	23	23	25	25
Treatment MST00, Female=0, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	26	27	28	28
Stratum 2	20	23	23	23
Stratum 3	22	23	25	25

Table 3.13: Participation Probit for QST00, Females

	Stratum 1	Stratum 2	Stratum 3
agegroup12	-0.113 (0.048)**		-0.228 (0.099)**
agegroup4		0.193 (0.080)**	
agegroup56	-0.021 (0.053)	0.155 (0.081)*	0.158 (0.094)*
child	0.116 (0.045)**	0.223 (0.068)***	0.197 (0.088)**
claimg0	-0.042 (0.127)	0.187 (0.122)	0.242 (0.118)**
claimg2	0.166 (0.095)*	0.254 (0.109)**	0.143 (0.126)
claimg34	0.173 (0.100)*	0.126 (0.149)	0.153 (0.141)
claimg34_married2		0.242 (0.170)	
countcon	0 (0.000)	0.001 (0.000)*	0 (0.000)
countemp	0 (0.000)	0 (0.000)	-0 (0.001)
countoos		0.001 (0.000)	0 (0.000)
dcountcon	0.135 (0.057)**		
dcountoos	-0.084 (0.054)		
dcountua		0.170 (0.108)	
dcountub		0.087 (0.088)	
ddssec	-4.380 (1.263)***		
demp12	0.049 (0.133)	0.075 (0.188)	-0.178 (0.277)
demp12_24	-0.037 (0.076)	0.072 (0.120)	0.239 (0.159)
demp6	0.027 (0.092)	0.027 (0.126)	0.235 (0.159)
demp6_12	-0.015 (0.136)	0.034 (0.198)	0.145 (0.275)
endlastjob2	0.227 (0.056)***	0.375 (0.084)***	
endlastjob3	0.291 (0.075)***	0.139 (0.123)	
endlastjob4	0.081 (0.072)	0.122 (0.113)	
foreigner		-0.301 (0.115)***	
health2	-0.271 (0.143)*		
health3	0.079 (0.139)		0.234 (0.117)**
industry3	0.282 (0.118)**		
industry4	0.062 (0.058)	0.172 (0.085)**	
industry5	0.068 (0.067)	0.044 (0.102)	
industry6	0.015 (0.061)	0.030 (0.090)	
lnwaged	2.241 (0.660)***		
lnwagedsq	-0.281 (0.086)***		
married2	0.222 (0.044)***	0.200 (0.074)***	0.455 (0.083)***

	Stratum 1	Stratum 2	Stratum 3
motivationlack		-0.082 (0.085)	
pasthealth2	0.326 (0.138)**		
pasthealth3	-0.041 (0.147)		
pasttreatcancel		0.155 (0.339)	
penalty		0.332 (0.159)**	
problemgroup	0.240 (0.117)**		
region2	0.075 (0.058)		
region3		0.026 (0.096)	
region4		0.086 (0.123)	
region5		0.114 (0.100)	
_cons	-2.408 (0.224)***	-3.386 (0.470)***	-2.733 (0.521)***
N	14385	7284	5105

Table 3.14: Results for Smith and Todd (2005) Balancing Test, QST00 Females

Treatment QST00, Female=1, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	30	31	31	31
Stratum 2	28	29	30	30
Stratum 3	15	15	15	15
Treatment QST00, Female=1, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	30	31	31	31
Stratum 2	24	29	30	30
Stratum 3	14	14	15	15

Table 3.15: Participation Probit for MST00, Females

	Stratum 1	Stratum 2	Stratum 3
agegroup1		-0.175 (0.117)	
agegroup2		-0.207 (0.106)*	
agegroup4			0.261 (0.115)**
agegroup5			0.280 (0.116)**
agegroup56	-0.262 (0.098)***	-0.169 (0.123)	
agegroup6			0.372 (0.117)***
area3		-0.214 (0.111)*	
child	0.213 (0.055)***	0.104 (0.089)	
claimg0	-0.045 (0.117)	0.324 (0.147)**	0.178 (0.117)
claimg1	-0.070 (0.109)	0.105 (0.126)	-0.053 (0.132)
claimg3	0.098 (0.101)	-0.143 (0.133)	
claimg34			-0.066 (0.129)
claimg3_dcountoos	-0.114 (0.116)		
claimg4	0.506 (0.140)***	0.023 (0.164)	
claimg4_dcountoos	-0.380 (0.182)**		
countoos		0 (0.000)	-0 (0.000)
countub		-0.001 (0.001)	-0.001 (0.000)
dcountcon	0.066 (0.062)	0.073 (0.094)	
dcountoos	0.154 (0.092)*		
dcountsub			0.264 (0.163)
demp12	0.059 (0.090)	-0.174 (0.242)	-0.288 (0.127)**
demp12_24	0.045 (0.131)		
demp24	0.037 (0.130)	0.021 (0.156)	-0.120 (0.141)
demp6	0.095 (0.087)		
demp6_12		0.295 (0.245)	
endlastjob2	0.147 (0.065)**	0.161 (0.089)*	
endlastjob3	0.229 (0.089)**		
endlastjob4	0.138 (0.080)*	0.189 (0.116)	
lncountemp	-0.087 (0.092)	-1.548 (1.464)	-0.626 (1.361)
lncountempsq		0.135 (0.131)	0.062 (0.119)
married2	0.201 (0.054)***	0.271 (0.080)***	0.193 (0.087)**
onlyparttime	-0.120 (0.066)*		
parttime		-0.005 (0.079)	-0.073 (0.084)

	Stratum 1	Stratum 2	Stratum 3
qualification1	0.148 (0.053)***		0.173 (0.083)**
quarter1	0.295 (0.087)***	0.282 (0.140)**	
quarter2	0.235 (0.092)**	0.263 (0.149)*	0.141 (0.167)
quarter3	0.275 (0.088)***	0.375 (0.136)***	0.169 (0.162)
quarter4			0.317 (0.151)**
quarter5	0.350 (0.083)***	0.225 (0.137)*	0.232 (0.151)
quarter6	0.231 (0.091)**	0.413 (0.136)***	0.337 (0.153)**
region2	0.457 (0.100)***	0.312 (0.132)**	0.460 (0.169)***
region3	0.309 (0.091)***	0.426 (0.103)***	0.355 (0.156)**
region4		0.129 (0.154)	
region5	0.150 (0.097)		0.099 (0.172)
schooling3	0.084 (0.064)		
youngchild		-0.232 (0.121)*	0.353 (0.105)***
_cons	-2.563 (0.540)***	1.394 (4.026)	-1.093 (3.846)
N	14173	7196	5090

Table 3.16: Results for Smith and Todd (2005) Balancing Test, MST00 Females

Treatment MST00, Female=1, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	25	28	30	30
Stratum 2	29	29	30	30
Stratum 3	22	24	25	25
Treatment MST00, Female=1, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	27	29	30	30
Stratum 2	28	30	30	30
Stratum 3	20	23	25	25

Figure 3.4: Graphical Check of Common Support for QST00

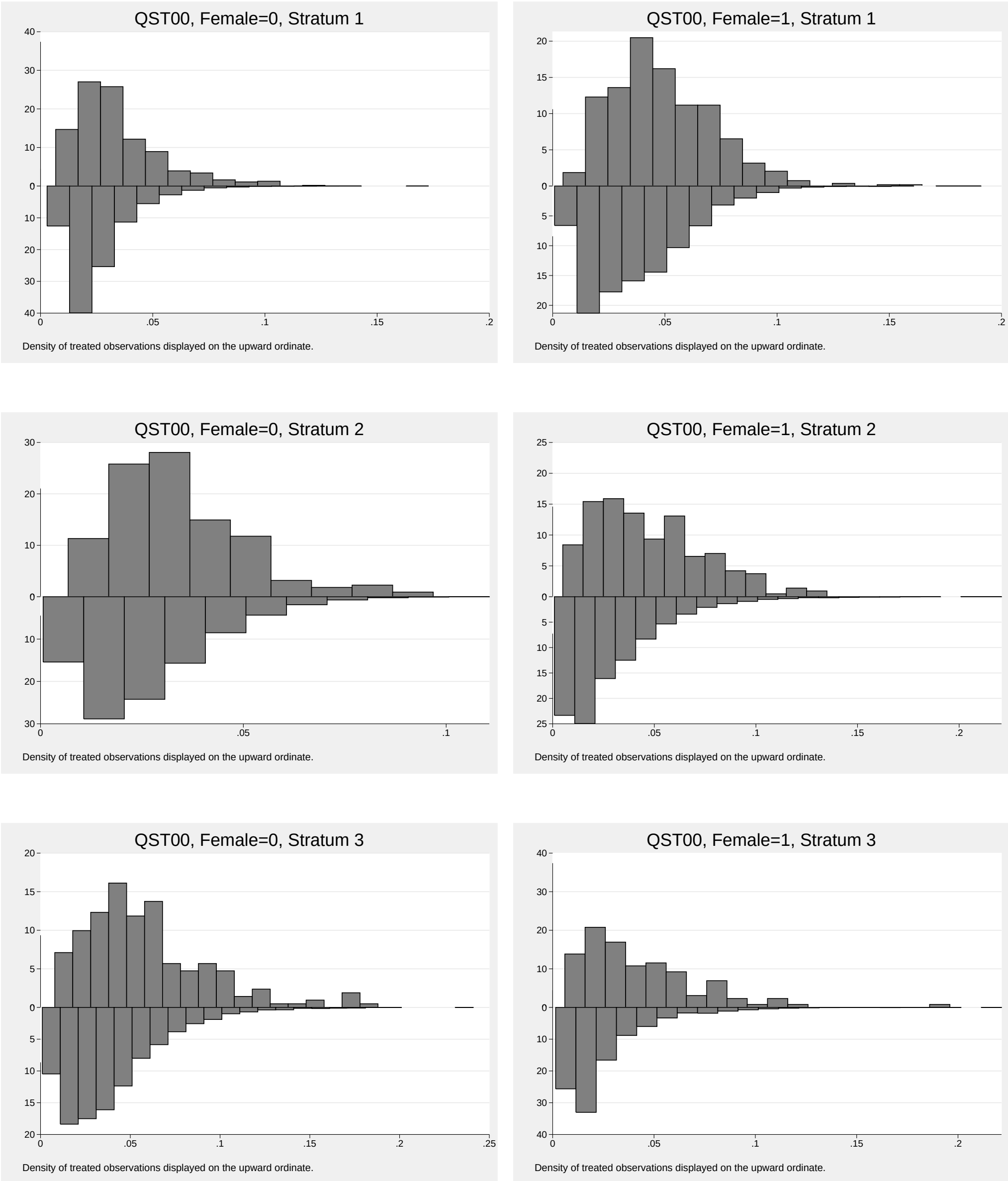


Figure 3.5: Graphical Check of Common Support for MST00

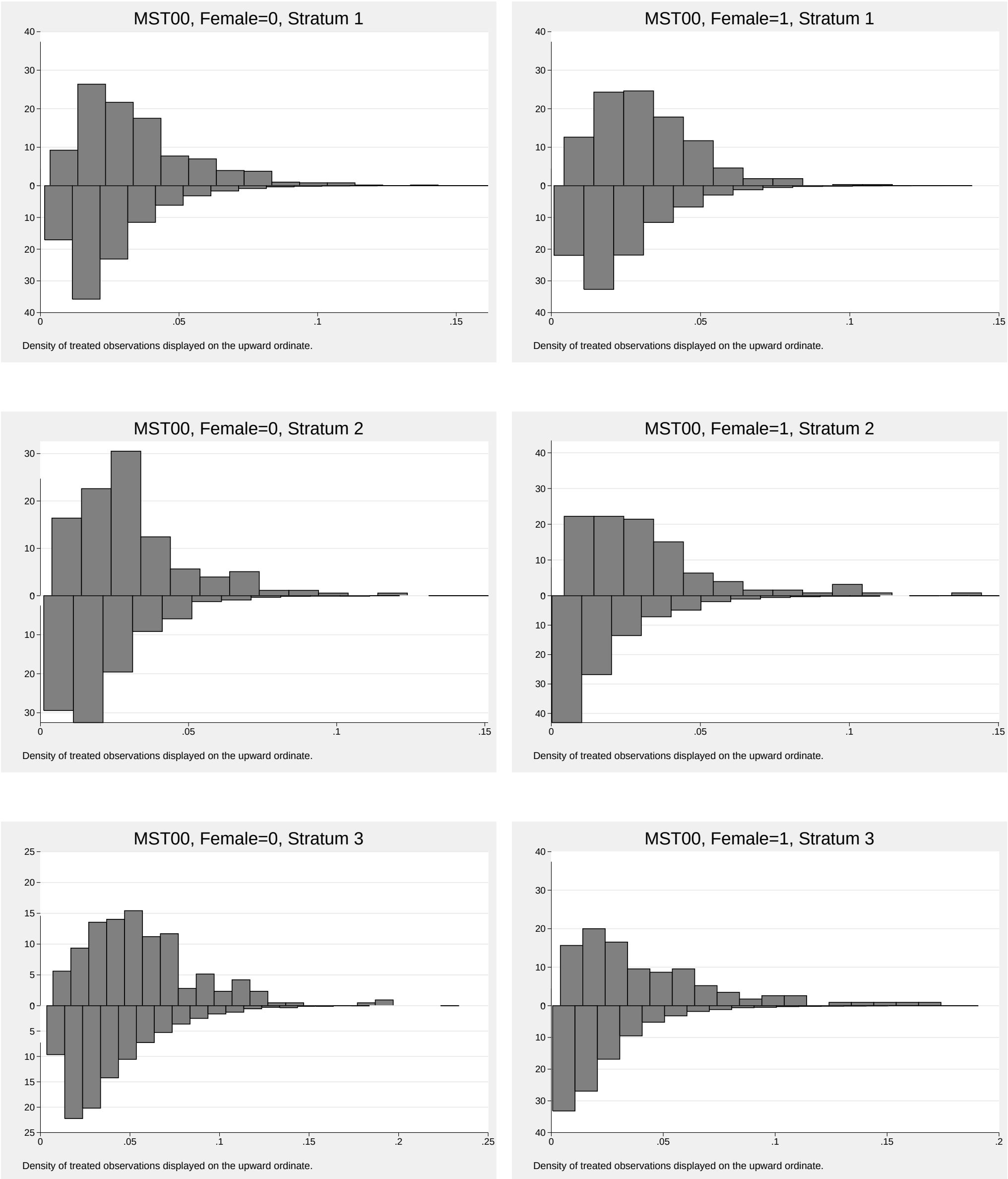


Table 3.17: Variable Definitions for the 1980-1992 Sample

Label	Definition
Personal Attributes	
aXXYY	Age at start of unemployment $\geq$XX and $\leq$ YY
age	Age at start of unemployment
lnage	ln(age) at start of unemployment
female	Female
foreign	No German citizenship
kids	Has dependent children
married	Married
BIL1	No vocational training degree
BIL2	Vocational training degree
BIL3	Abitur/No vocational training degree
BIL4	University/College degree
Last Employment	
BER1	Apprentice
BER2	Blue Collar Worker
BER3	White Collar Worker
BER4	Worker at home with low hours or BER missing
BER5	Part-time working
pentg	Daily earnings $\geq$ 15 Euro per day in 1995 Euro
entgcens	Earnings censored at social security taxation threshold
entg	Daily earnings if pentg=1 and entgcens=0, otherwise zero
logentg	log of entg if pentg=1 and entgcens=0, otherwise zero
claim0	Remaining claim on unemployment benefit at beginning of Stratum 1
claim181	Remaining claim on unemployment benefit at beginning of Stratum 2
claim361	Remaining claim on unemployment benefit at beginning of Stratum 3
lnclaimX (X = 0, 181, 361)	ln(claimX)
claimXg0	claimX=0
claimXg1	claimX>0 and claim0 $\leq$ 170

<continued on next page>

Label	Definition
claimXg2	claimX>170 and claim0 $\leq$ 350
claimXg3	claimX>350
Last Employer	
WZW1	Agriculture
WZW2	Basic materials
WZW3	Metal, vehicles, electronics
WZW4	Light industry
WZW5	Construction
WZW6	Production oriented services, trade, banking
WZW7	Consumer oriented services, organization and social services
frmsize1	Firm Size (employment) missing or $\leq$ 10
frmsize2	Firm Size (employment) > 10 and $\leq$ 200
frmsize3	Firm Size (employment) > 200 and $\leq$ 500
frmsize4	Firm Size (employment) > 500
Employment and Program History	
preexM	Employed M (M=6, 12, 24) month before unemployment starts
preex6cum	Number of months employed in the last 6 months before unemployment starts
preex12cum	Number of months employed in the last 12 months before unemployment starts
preex24cum	Number of months employed in the last 24 months before unemployment starts
preex60cum	Number of months employed in the last 60 months before unemployment starts
pretxY	Participation in any ALMP program reported in our data in year(s) Y (Y=1, 2) before unemployment starts
Regional Information	
LAND6	Schleswig-Holstein/Hamburg
LAND7	Niedersachsen/Bremen
LAND8	Nordrhein-Westfalen
LAND9	Hessen
LAND10	Rheinland-Pfalz/Saarland
LAND11	Baden-Württemberg
<continued on next page>	

Label	Definition
LAND12	Bayern
Calendar Time of Entry into Unemployment	
tnull	First unemployment month (January 1960=0)
y19YY	Unemployment begins in year 19YY
Interaction of Variables	
south	Baden-Württemberg/Bayern
middle	Hessen/Rheinland-Pfalz/Saarland
north	Schleswig-Holstein/Hamburg/Niedersachsen/Bremen
BILXBERY	Combination of education- and job-status-variables
yXXYY	Unemployment begins between year 19XX and 19YY

All variables except those referring to benefit claims are defined at the time of entry into unemployment and constant during the unemployment spell.

Estimated Propensity Scores for the 1980-1992 Sample

Table 3.18: Participation Probit for ST8092, Males

	Stratum 1	Stratum 2	Stratum 3
BER2		-0.201 (0.186)	0.123 (0.072)*
BER3	0.229 (0.072)***		
BIL1BER2	0.114 (0.072)	0.365 (0.242)	
BIL1BER3		0.636 (0.288)**	
BIL1a3034		0.341 (0.144)**	
BIL1a3544		-0.511 (0.252)**	
BIL2		0.429 (0.222)*	
BIL2BER3		-0.068 (0.192)	
BIL2a3544		-0.416 (0.223)*	
BIL4	-0.188 (0.123)		
LAND10	0.286 (0.086)***		
LAND8	0.169 (0.061)***		
LAND9	0.262 (0.084)***		
WZW1	0.235 (0.146)		
WZW2	0.231 (0.098)**		
WZW3	0.301 (0.086)***	0.177 (0.073)**	
WZW6	0.184 (0.080)**	0.125 (0.060)**	

	Stratum 1	Stratum 2	Stratum 3
WZW7	0.121 (0.095)		
a3044			0.161 (0.058)***
a3544		0.520 (0.217)**	
a4553	-0.183 (0.076)**		
claim0	-0.002 (0.001)*		
claim0g0	-0.466 (0.353)		
claim0g1	-0.479 (0.268)*		
claim0g2	-0.207 (0.130)		
claim181		-0.001 (0.000)***	
entgcens	-0.127 (0.179)		-0.254 (0.234)
foreign	-0.145 (0.085)*	-0.130 (0.083)	-0.284 (0.090)***
frmsize1	-0.310 (0.088)***		-0.025 (0.065)
frmsize2	-0.174 (0.078)**		
frmsize3	-0.049 (0.103)		0.122 (0.092)
kids	-0.134 (0.077)*		
lnage		-0.112 (0.126)	0.123 (0.133)
lnclaim361			-0.028 (0.014)**
logentg	-0.030 (0.023)	0.036 (0.030)	-0.043 (0.040)
married	-0.003 (0.069)		-0.030 (0.062)
middle		-0.012 (0.073)	0.147 (0.075)*
north		-0.151 (0.071)**	-0.184 (0.079)**
preex12cum	0.019 (0.019)		-0.043 (0.022)*
preex24cum	-0.019 (0.011)*	0.001 (0.007)	0.018 (0.008)**
preex60cum	0.001 (0.003)	0.003 (0.002)	
preex6cum			0.036 (0.049)
south		-0.147 (0.071)**	-0.182 (0.079)**
tnull		-0.001 (0.001)	
y1982		-0.222 (0.094)**	
y1983		-0.265 (0.099)***	
y1987		-0.254 (0.114)**	
y1990			0.150 (0.104)
y1991			0.385 (0.196)**
y8182	0.175 (0.069)**		
y8687	0.361 (0.060)***		
_cons	-2.159 (0.405)***	-2.130 (0.507)***	-2.676 (0.533)***

	Stratum 1	Stratum 2	Stratum 3
N	60083	25711	15814

Table 3.19: Results for Smith and Todd (2005) Balancing Test, ST8092 Males

Treatment ST8092, Female=0, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	27	29	29	29
Stratum 2	22	23	24	24
Stratum 3	18	18	18	18
Treatment ST8092, Female=0, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	22	26	29	29
Stratum 2	21	24	24	24
Stratum 3	15	18	18	18

Table 3.20: Participation Probit for ST8092, Females

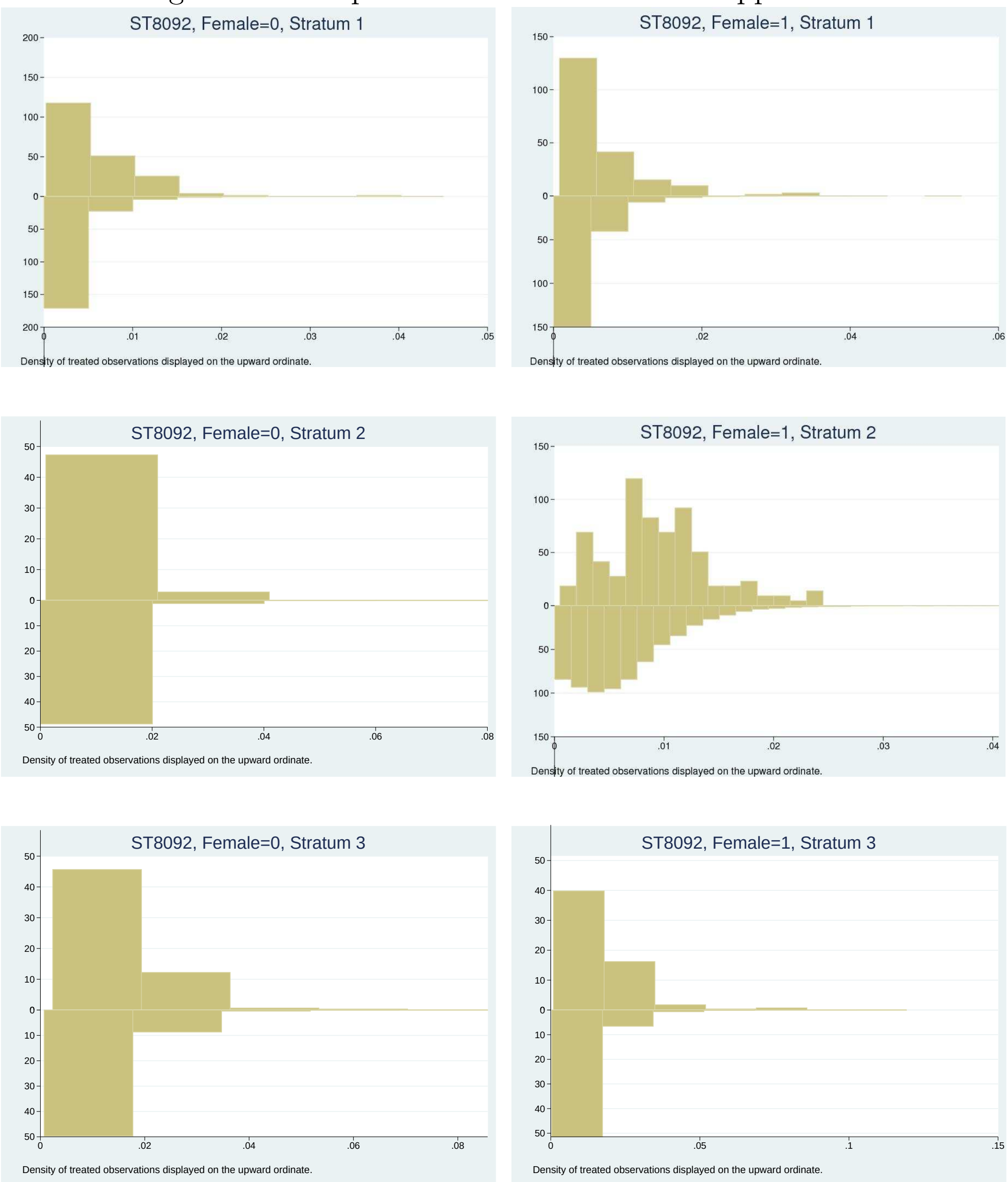
	Stratum 1	Stratum 2	Stratum 3
BER2			0.024 (0.088)
BER3			0.163 (0.083)*
BIL1	0.477 (0.172)***		
BIL2	0.443 (0.180)**		
BIL2a3034	-0.148 (0.150)		
BIL2a3544	0.171 (0.086)**		
LAND11		-0.181 (0.110)*	
LAND8	0.187 (0.060)***		
WZW3	0.094 (0.085)	-0.141 (0.114)	
WZW6		0.029 (0.077)	
WZW7		-0.170 (0.087)**	
a3034	0.322 (0.123)***		
a3544		0.123 (0.069)*	0.124 (0.067)*
claim0g1	0.578 (0.356)		
claim0g2	0.861 (0.449)*		
claim0g3	0.825 (0.482)*		
claim181g0		0.933 (0.308)***	
claim181g1		0.772 (0.183)***	
claim181g2		0.573 (0.146)***	
claim_361			-0.219 (0.088)**
entgcens			0.370 (0.298)
foreign	-0.277 (0.121)**		
frmsize1	-0.119 (0.064)*		-0.180 (0.094)*
frmsize2			-0.208 (0.090)**
frmsize4			-0.046 (0.100)
kids	-0.036 (0.081)	0.005 (0.084)	0.025 (0.087)
lnage	0.036 (0.140)	-0.155 (0.150)	
lnclaim0	-0.133 (0.080)*		
lnclaim181		0.050 (0.044)	
logentg	0.087 (0.036)**	0.088 (0.042)**	0.045 (0.047)
married	-0.128 (0.058)**	-0.125 (0.061)**	-0.150 (0.059)**
middle		0.133 (0.080)*	
north		0.022 (0.076)	-0.108 (0.075)

	Stratum 1	Stratum 2	Stratum 3
preex12		-0.111 (0.084)	0.276 (0.135)**
preex12cum			-0.061 (0.033)*
preex24		0.174 (0.078)**	
preex24cum			0.005 (0.011)
preex6	-0.145 (0.082)*		0.146 (0.189)
preex60cum			0.003 (0.003)
pretx2			0.317 (0.147)**
south			-0.265 (0.075)***
tnull	0 (0.001)	0.002 (0.001)**	0 (0.001)
y1984		-0.256 (0.130)**	
y1986	0.294 (0.084)***		
y1990	0.130 (0.117)		
y1991	0.487 (0.178)***		
y8486			0.255 (0.069)***
y8889		-0.245 (0.099)**	0.275 (0.102)***
_cons	-3.572 (0.576)***	-3.530 (0.609)***	-2.336 (0.389)***
N	35927	23115	17148

Table 3.21: Results for Smith and Todd (2005) Balancing Test, ST8092 Females

Treatment ST8092, Female=1, Cubic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	22	22	22	22
Stratum 2	20	20	20	20
Stratum 3	21	22	22	22
Treatment ST8092, Female=1, Quartic of Pscore				
	P-values>.1	P-values>.05	P-values>.01	Regressors
Stratum 1	17	19	21	22
Stratum 2	18	18	20	20
Stratum 3	20	22	22	22

Figure 3.6: Graphical Check of Common Support for ST8092



Chapter 4

Many Dropouts? Never Mind! - Employment Prospects of Dropouts from Training Programs

4.1 Introduction

Training programs represent an important part of active labor market policies in many countries and researchers have shown strong interest in analyzing the labor market effects of these programs.[1] But little is known about the number, the characteristics, and the labor market prospects of those who drop out of these programs. This is surprising as it turns out that dropouts represent a considerable group of participants: in Germany one out of five participants drops out of the program.[2] Dropouts will have a head start on the labor market, because they may already be employed while the other participants are still attending the program. But how about the medium-term and long-term effects of dropout: does it harm to drop out in the long run?

From a policy perspective knowledge about the occurrence and the labor market prospects of dropouts may be important, as institutional settings like benefits during program participation and sanctions may influence the number of those who drop out. Furthermore, studying the labor market prospects of dropouts may provide further insights in understanding the composition of average treatment effects estimated in the literature on training programs. There exists a literature for the US on dropouts of labor market programs in experimental situations (see for example Heckman, Smith and Taber (1998)). Furthermore, the threat effect of being assigned to a labor market program but not participating has been studied (see for example Rosholm and Svarer (2008)).

For the first time this paper sheds light on dropouts from training programs in western countries in a non-experimental setting.[3] Studying the effect of dropout requires to overcome two main obstacles. First, data allowing to identify which participants drop out of the program is needed. I propose a strategy to identify dropouts of further training programs using German administrative data which can be applied after

[1] In Germany from 2000 to 2002 about 1.5 million entries are registered (Bundesagentur für Arbeit 2001, 2002, 2003). The employment effects of these programs have been estimated for example by Biewen et al. (2007), Hujer et al. (2006), Kluve et al. (2007), Lechner and Wunsch (2006a, 2006b, 2007), Osikominu (2008), Rinne et al. (2007), Schneider and Uhlendorff (2006), Stephan and Pahnke (2008), Wunsch and Lechner (2008).

[2] Own calculation based on the definition of dropout and the sample of participants presented in section 4.2.

[3] The only paper on dropouts of labor market programs in a non-experimental setting is Lee and Lee (2003). Using Korean data, the authors make an attempt to deal with dropouts in program evaluation by pairwisely comparing those who complete the program, drop out or do not participate using matching.

having corrected measurement error in the registered end of participation. Second, to estimate the effect of dropping out versus completing the program it seems necessary to take into account observable and unobservable differences between dropouts and non-dropouts as well as state dependence and duration dependence. I estimate the medium-term to long-term effect of dropout using a bivariate dynamic random effects probit model. The model is identified by the timing-of-events and through functional form assumptions. It consists of a dropout equation and an employment equation. Both equations include an unobserved individual effect and these two random effects are allowed to be correlated. The two equations are estimated simultaneously using Markov Chain Monte Carlo (MCMC) methods, a technique from Bayesian statistics. I program a Gibbs sampling algorithm to simulate draws from the posterior distribution of the parameters. This approach provides information on all parameters of the model, including information on the unobserved individual specific effects. To get an estimate for the size of the dropout effect, I calculate average partial effects on the treated which account for the selection based on unobservables. This is possible because of the availability of the predictors of the unobserved individual specific effects from the MCMC estimation.

Usually, evaluation studies on the employment effects of training programs consider the start of a program as the treatment (possibly with the restriction that it has been attended for some weeks) and do not deal with the actual length of participation or the question if the program has been completed.[4] Exceptions are Kluve et al. (2007) who estimate the employment effect of variations in the length of German training programs, and Flores-Lagunes et. al (2007) who estimate earnings effects of the length of US training programs. Both papers use a matching strategy adapted to continuous treatment decisions. Fitzenberger et al. (2009) include the duration of participation in their model allowing for the end of participation to be endogenous. By contrast, the focus of this paper is on the difference between dropping out and completing the measure. Consider a sample of individuals who all experience a transition from employment to unemployment and who all start a training program. While they are in the program they decide in each period to continue or to drop out. Those who decide to drop out differ - from this period onwards - from the others by having experienced a dropout while the others will eventually have completed a program. So in the notion of the evaluation literature dropout would be the treatment. There are various reasons for dropout: an important one is that the individual is

[4]Biewen et al. (2007) for example consider program participation in medium-term further training programs if it has lasted for at least four weeks and consider shorter spells as no treatment.

lucky to receive a job offer and decides to drop out and start employment. Other examples for why some people drop out and others do not are different ex-ante information on the programs and dropout if expectations are not met, personal dislike of the teacher or classmates, temporarily higher opportunity costs (for example due to a work opportunity on the black market), changes in preferences, lack of endurance in relation to training or differences in individual discounting of the future.

There may be a specific effect on employment of dropping out versus completing a program which may be different from the effect of attending programs of different lengths. In addition to attending the program for less time, dropping out might involve missing parts of the curriculum, not obtaining a certificate, and a signal to potential employers. If, for example, the curriculum of a course covers all essential tasks of a profession one after the other, it might be less valuable to attend half of this course than attending a complete course which is of shorter planned length and more condensed. Obtaining a certificate might in particular be valuable for courses leading to officially recognized professional degrees. Also, a future potential employer might judge a dropout as a negative signal of endurance. On the other hand it is possible that attending a program for some time is enough to get the benefit out of it. This would for example be the case if program effects are due to an activation of the unemployed or an improved orientation of the individuals in which kind of job they might succeed. Furthermore, participants might even use the possibility to drop out for staying in the program only until the right moment (with regard to their skills or the economic situation) arrives to start searching for a job or until they receive a job offer which is a very good match. Thus, participants who follow this strategy might benefit from it as opposed to waiting until the planned end of the program and then starting to look for jobs.[5] To sum up, dropout may involve a negative, positive or zero effect on employment prospects.

The remainder of this paper is structured as follows: section two introduces the data, defines the evaluation sample and discusses how dropouts can be identified. Section three includes a descriptive analysis of the occurrence of dropouts and their employment prospects. Section four discusses the econometric model used to estimate the medium- and long-term effect of dropout, describes the estimation strategy and presents the results. Section five concludes.

[5]See Becker (2005) for a theoretical model which formalizes such a strategy.

4.2 Identification of Dropouts of Further Training Programs in the IEBS

4.2.1 The Integrated Employment Biographies Sample

The Integrated Employment Biographies Sample (IEBS) consists of a 2.2% random sample of individuals drawn from the universe of data records collected in four different administrative processes: the IAB Employment History (*Beschäftigten-Historik*), the IAB Benefit Recipient History (*Leistungsempfänger-Historik*), the Data on Job Search Originating from the Applicants Pool Database (*Bewerberangebot*), and the Participants-in-measures Data (*Massnahme-Teilnehmer-Gesamtdatenbank*).[6] The data contain detailed daily information on employment subject to social security contributions, receipt of transfer payments during unemployment, job search, and participation in different programs of active labor market policy (ALMP). To be specific, this study uses a draw of the administrative data which is called *IEB, Version 4.02*.[7]

The first of the four administrative data sources included in the IEBS, the IAB Employment History, consists of social insurance register data for employees subject to contributions to the public social security system. It covers the time period from 1990 to 2004. The main feature of these data is detailed daily information on the employment status of each recorded individual. For each employment spell, in addition to start and end dates, data from the Employment History contains information on personal as well as job characteristics such as wage, industry or occupation. In this study this information is used to account for the labor market history of individuals as well as to measure employment outcomes. The IAB Benefit Recipient History, the second data source, includes daily spells of unemployment benefit, unemployment assistance and subsistence allowance payments the individuals received

[6]For detailed information on the IEBS see Zimmermann et al. (2007). Information in English can be found on the website of the Research Data Center (FDZ) of the Federal Employment Office (BA) (http://fdz.iab.de/en). The website also describes the conditions under which researchers may use the IEBS and the process to get the permission.

[7]The specific version used here is described in IEB Benutzerhandbuch Version V4.02, 16.01.2006 and attendant documents, not published. This version includes some variables which are not in the standard version. The names of the additional variables I considered for this study are the following (some of them turned out to be irrelevant for the estimations): *Familienstand, FbW Abmeldedatum, Geburtsjahr juengstes Kind, Geplante Massnahmendauer, Gesundheitliche Einschraenkungen, Kapazitaet Teilnehmer FbW, Massnahmeerfolg, Massnahme - Lernort, Massnahmetraeger, Massnahmeziel - Prüfungsart, Rehabilitationsmassnahme, Zugangsgrund.*

between January 1990 and June 2005. In addition to the sort of the payment and the start and end dates of periods of transfer receipt the spells contain further information like sanctions, periods of disqualification from benefit receipt and personal characteristics. These data are mainly used to get additional information on the labor market history of these individuals.

The third data source included in the IEBS is the so-called Data on Job Search Originating from the Applicants Pool Database, which contains rich information on individuals searching for jobs covering the period from January 1997 to June 2005. The spells include detailed information concerning job search, regional information, and personal characteristics. This information is used to control for individual characteristics of the participants in the estimations. The Participants-in-measures Data, the fourth data source, contains diverse information on participation in public sector sponsored labor market programs, for example training programs, job-creation measures or integration subsidies. It covers the period from January 2000 to July 2005. Similar to the other sources, also these data come in the form of spells indicating the start and end dates at the daily level. Information for example on the type of the program and the planned end date are added. This data source is necessary to identify participation and to gain information on the program attended.

4.2.2 Sample and Further Training Programs

The focus of this study is on participation in public sector sponsored further training programs starting in between July 2000 and December 2001. Further training programs are defined in this paper as those measures that train professional skills and have a typical duration of several months up to two years. This includes all programs called *FbW - Foerderung der beruflichen Weiterbildung* in the IEBS and under the legislation except those called orientation measure, because with regard to length and content they have more in common with short-term training, a different part of German active labor market policies, than with further training programs. Because further training programs differ in how they are organized and with regard to the certificate the participant may obtain, I distinguish three groups of further training programs: general further training, practical training, and retraining. General further training teaches specific professional skills, mostly in class room. A typical example would be IT-based accounting. Participants may obtain a certificate by the school or by a professional organization. The programs subsumed under practical training take place in a training firm or include an internship in a firm, and their duration is typically a bit shorter. Retraining is, with a typical length of two years, the longest program and participants are trained in a profession which differs from the one they originally learned. In the end they may obtain a new professional degree within the German apprenticeship system.

To study the core group of participants in further training, the sample is based on programs started within the first year of an unemployment period as the first intensive active labor market program. Only individuals who experienced an inflow from continuous employment into unemployment within the year before program start are considered. Entering unemployment is defined as quitting regular (not marginal), non-subsidized employment of at least three months and subsequently being in contact with the labor agency (not necessarily immediately), either through benefit receipt, program participation, or a job search spell.[8] In order to exclude individuals eligible for specific labor market programs for young people and individuals eligible for early retirement schemes, only persons aged between 25 and 53 years at the beginning of their unemployment spell are considered. Men and women living in East and in West Germany are included.

4.2.3 Identification of Dropouts in the Data

Dropping out of a program is defined as having started a program but not completing the program, but instead quitting it before the planned end is reached. The IEBS includes a variable for the start of the program, the end of participation and the initially planned end of the program. If the data indicate that a program has been started the question is if the program has been attended (almost) as long as initially planned (planned end date) or considerably shorter. The planned length of the program is defined here as the date of the planned end minus the date of the start of the program. It is necessary to set cut-off points for the distinction of dropout and completion as well as the distinction of realized attendance and non-attendance. In this paper, program attendance is categorized as dropout as opposed to completion if the program has been attended less than 80% of the planned length.[9]

[8] Note that this implies that the same individual could appear in the sample more than once, if he or she had more than one valid unemployment spell and attended both times a further training program. This does not happen in the sample used for this study.

[9] Several further training program spells are linked to one participation if the gaps in between are less than 15 days, thus a change from one further training program into another is not counted

If attendance in the data is less than four days (and in the rare cases in which the variable *success of the program (Massnahmeerfolg)* indicates *not attended*) this is not counted as program participation for two reasons: first, dropout is understood here as having attended at least a few days and then dropping out and not as having rejected to attend a program right from the start. Second, extremely short program spells in the data may indicate in some cases that the program has not been attended at all but the registration was withdrawn too late and this was not corrected in the data. So one might count some cases as dropouts that never attended if too short spells are counted as participation. Therefore program spells which are shorter than four days do not lead to the inclusion of the individual into the sample of analysis (which is based on participants as described in section 4.2.2). As mentioned before, to distinguish between dropout and complete attendance, participation of 80% of the planned length is chosen. Choosing a higher limit, one would risk misclassifying participants as dropouts if the whole course ends a bit earlier than planned at the beginning. This may happen especially for two-year programs, particularly if they end with an external exam the date of which is not fixed when the program starts. The data reflects this - at around 90% percent of planned duration the number of finishing attendances rises. Apart from identification issues, one could argue that attending a very high percentage of the planned duration is more like full attendance than like a dropout.

For the identification of dropouts the reliability of the end date of participation as well as the planned end date are of utmost importance. But there is some measurement error in the end dates of participation in further training programs in the IEBS, see Waller (2008). This means it happens that a person quits a program but the end of participation in the data is nevertheless equal to the planned end date. To correctly identify dropouts it is necessary to correct these wrong end dates, otherwise far too few participants would be identified as dropouts. In this study the correction procedure proposed in Waller (2008) is used. It relies mainly on the information on subsistence allowance (a transfer payment made to the participants of further training programs for the time of their participation) of the IAB Benefit Recipient History, which is considered very reliable. In addition, the correction procedure in some cases uses certain contradictions with employment spells of the IAB Employment History as well as some further pieces of information from the data.

as a dropout. A gap of three months is allowed, if there is information in the data that the person was ill in between two program spells, but this turned out to be empirically irrelevant.

The planned end date of further training programs seems to be reliable in indicating until when program participation was first planned. For 7.7% of the relevant programs, the planned end date is earlier than the end date of participation. This is not necessarily measurement error - it is possible that a participant attends longer than originally planned. If the difference only amounts to a few days this is very likely to be correct, because the end of the courses can change a bit after program start. For 3.2% of the programs this difference is more than 7 days. In these cases, it may be that the participant attended for a considerably longer time period than planned - in particular if the program is not a group course but an individual program - but there might also be a problem. Thus, for the total of 3.4% of the programs for which there is an indication that the reported planned end dates should not be used as the only source for the classification, the two variables *success of the program* and *duration of the course in months* are additionally used to decide if the participation spell is classified as a dropout or not. These variables have a lot of missings and are error-prone, but - used with caution and only in addition to the planned end date - they can help to decide for the major part of the 116 programs requiring further information for classification. In the end, only for 30 (0.88%) of the programs in focus it seems impossible to classify them and these cannot be used for the analysis.

The identification strategy used in this paper has been subjected to various sensitivity checks. There is no indication of systematic problems.[10] As an alternative way to identify dropouts one might think of using the variable *success of the program* which may not only indicate non-attendance but also successful completion or dropout. Taken the information literally, one could use this variable to classify participants into dropouts and non-dropouts. But there are at least three problems in doing so: Firstly, the variable is missing or not available for 14% of relevant program spells. Secondly, it is not clear how dropout is defined in the variable and under which circumstances a dropout is registered. Thirdly, the variable suffers from severe measurement error: 49% of those classified as dropouts in this paper are coded as having completed with success. For these reasons, I prefer the strategy outlined above.

[10] In particular it may be ruled out that participants classified as dropouts in fact attended a program which was shifted to an earlier time. This hypothesis has been checked first by using the information on allowance payments to check if participants have in fact started their program the day of the indicated start date: deviations are not more frequent for those classified as dropouts than for non-dropouts. A second check was to compare those who participated shortly after becoming unemployed (for whom it is impossible that the program was shifted to an earlier start date and the start date reported in the data was not changed) with later participants, in particular with regard to the planned length of their programs and the timing of drop-out.

4.3 Descriptive Analysis

4.3.1 Occurrence of Dropout

When applying the above definition of dropout to the data, it turns out that 21% of the programs end with a dropout. The share of those who drop out differs with respect to the program type. Table 4.1 shows that the share of dropouts is lowest for general professional training and a bit higher for the very long retraining programs. The program type practical training suffers from the highest dropout rate: about 30% of the participants drop out of these programs, even though practical training programs are relatively short. The diagrams in figure 4.4 in Appendix A show when dropouts quit the program. With regard to general professional training the number of dropouts grows with the elapsed duration measured as a share of the planned duration. Considering retraining many participants drop out during the first 40% of the planned program duration and relatively few afterwards. Regarding practical training the dropout rate is especially high in the third, fourth and sixth decile of the planned duration. For the participants there are no direct financial benefits or costs of attending a program. While attending a program, participants in further training usually receive the same amount of a benefit called subsistence allowance as the amount of unemployment benefit they would receive if they did not attend. The labor agency covers the direct costs for the course and in some cases transportation costs or child care costs. If participants drop out without a good reason, they might be punished by not receiving benefits for up to six weeks, but according to the data and to what case workers told me, sanctions are usually not imposed because a participant dropped out of a further training program. To check the motivation of participants, cheaper programs, like short-term training, are used.

Table 4.1: Share of Dropouts and Program Categories

Type of program	# of participants	Share dropout	Median planned length
General Prof. Train.	1761	18.57%	8.5 months
Retraining	761	22.47%	24 months
Practical Training	544	29.78%	6 months

After dropout, the individuals either start employment immediately or they stay in non-employment for some time. The first is called *job-aligned dropout* in the following and the latter *non-job-aligned dropout*. Job-aligned dropout occurs either because participants receive a job offer and drop out due to this job offer, or because participants drop out for other reasons but start a job (which may have been available to them before but which they did not consider). On the contrary, in the case of non-job-aligned dropout participants drop out and do not start employment (subject to social security). They prefer non-employment without attending a training program to attending a training program. The law encourages participants who receive a job offer to drop out. The general rule of the German ALMP is to give priority to placement over active labor market measures. An exception is possible if the measure is necessary for a durable placement (SGB III, § 4, § 5). But it is not clear under which circumstances it is preferable to encourage participants to continue. To see the relative importance of both types of dropout, I use the information if the participant starts employment within one month after dropout to decide whether it is a job-aligned dropout or not. According to this proxy, 45% of the dropouts experience a job-aligned dropout. Alternatively, one could use the variable *success of the program* which has potentially correct information for 38% of the dropouts. Out of these, 48% are recorded to drop out due to a job offer (this information does not seem to be missing at random). Both measures indicate that a bit less than half of the dropouts experience a job-aligned dropout.

To find out which characteristics of the program and of the participants are related to dropout, a cross-sectional probit model is estimated. For the specification search variables picturing the following characteristics are considered: personal characteristics (like gender, age, nationality, occupational qualification, degree of schooling, current health problems, past health problems, disabilities, past incapacities, children), information on the last employment (occupation in last job, last job part-time, last job as a blue-collar worker, reason for the end of the last employment, last wage), regional information (labor market situation in the region, West or East Germany), information on the individual labor market history (elapsed length of unemployment period, quarter of beginning of unemployment, information on lack of motivation related to labor agency activities in the past, information on participation in programs with social assistance in the past, sanctions in the past (also interacted with number of days with transfer payments), number of days in different labor market status (unemployment benefit, unemployment assistance, program participation, out of labor market, employment) in the last three years before the start of unemployment) and information on the program (planned length of the program, capacity of the program, information on institution offering the program, the sort of the certificate the program leads to). All the above-mentioned variables have been considered, but the vast majority of them turned out not to be relevant.

Table 4.4 in Appendix A shows the average partial effects and standard errors of a specification which includes the variables that seem to have some relevance. No schooling degree or a low schooling degree is related to a higher probability of dropout for participants of general professional training and retraining. The effects are large - for general professional training the average partial effect of having no schooling degree is 16.3%, for retraining it is 20.2%. For participants of practical training, who on average have lower education than participants of the other programs, there are no significant effects of schooling. Having experienced a sanction in the past is related to a higher probability of dropout (but this effect is not significant) as well as signs for a lack of motivation with respect to labor agency activities (significant for general professional training and retraining). With regard to practical training women and people living in East Germany are less likely to drop out, which for the latter group is also true for retraining. There is slight evidence that younger people as well as those who live alone are more likely to drop out. For retraining the effect of living alone is large (13%) and highly significant. Having a child under the age of ten is related to a strongly increased dropout probability for men taking retraining. A longer planned duration of the program increases the probability of dropout (not significant for retraining) as well as having participated in a training program in the past (not significant for general professional training). An increased probability of dropout is also observed for those who have experienced unemployment in the last three years before the current unemployment period (significant only for general professional training).

4.3.2 Employment Rates and Employment Stability

In this section the employment prospects of dropouts as compared to participants who do not drop out of the program are studied descriptively.[11] The analysis is based on a panel data set in months which follows the participants from the month they start the program (t=1) until 39 months later (t=40). There is some censoring due to the end of the observation period, but every individual may be followed at least for 37 months.[12] The analysis of differences in employment chances between

[11] When considering to analyze wage differences between dropouts and non-dropouts, I found that this is not a meaningful exercise, because differences in the annual wage are very largely driven by the number of days in employment. This is due to the fact that the majority of person-months in the sample of those who received further training within the last years indicate non-employment.

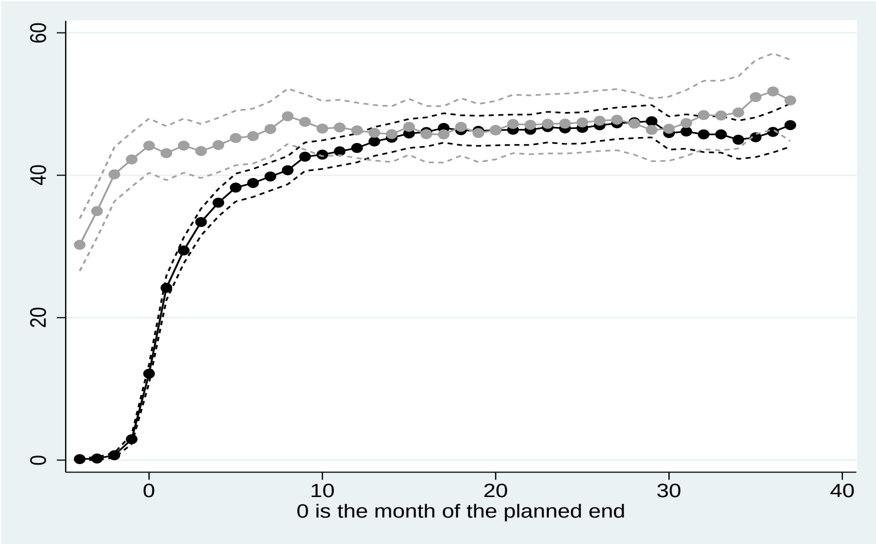
[12] A person is counted as employed in the respective month if he or she is employed for at least half of the month. The month an employment period begins is in addition counted as a month in

dropouts and non-dropouts should start at the time participants leave the program, because before that point in time all individuals experience a transition from employment to unemployment, start a further training program and they are all in non-employment while attending the program. This will be implemented when estimating the econometric model in the next section. But nevertheless the time axis of the figures (and later the time dummies in the econometric model) is aligned to the planned end of the program. In principle there are three options for the alignment: The first option is to align the time axis to the start of the programs. Dropouts leave the program earlier than non-dropouts, and as many of them take up a job, dropouts have a head start as compared to non-dropouts. While the program is still running, the employment rate of non-dropouts is zero but the employment rate of dropouts is positive, so the employment effect of dropout will be positive until the end of the program. Dropouts reduce the lock-in effect of a training program and the positive employment effects resulting from this should not be neglected when comparing employment rates of dropouts and non-dropouts. Aligning the comparison of employment rates to the start of the programs has the advantage to make the head start of dropouts visible, but if dropouts attend shorter or longer programs, the effect of this may not be distinguished from the effect of the head start. Second, if one aligned the time axis to the realized end of participation, there would be a jump in time due to dropout, because dropouts "shorten" the program. Thus, aligning the analysis to the realized end, the head start of dropouts would not be visible. The third option is an alignment to the planned end of the programs. Since this makes the head start of dropouts visible and avoids a mixture of the effect of dropout and the effect of the planned length of the program, this alignment is applied in the following.

Figure 4.1 compares the average employment status of dropouts and the average employment status of non-dropouts in each month aligned to the planned end of the programs. Consider for example month 10 after the planned end. The figure shows that 43% of those who completed the program are employed 10 months after they reached the planned end of their program and 47% of the dropouts are employed after they have reached the month of the planned end of their program (of course they are not in the program anymore at that time). Figure 4.1 shows the head start of dropouts: for example four months before the planned end of their programs, 30%

employment if the employment period is in sum more than half of a month's length (that is even if this is split into two calendar months) and the second is then only counted if the employment period is in sum more than one month.

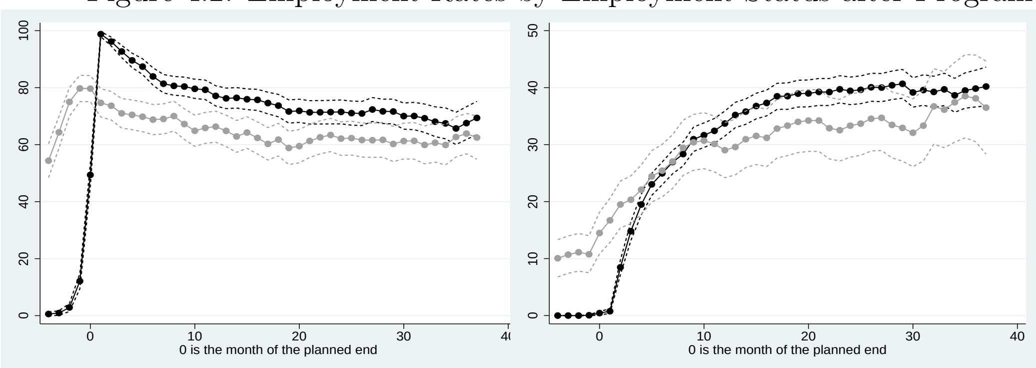
Figure 4.1: Employment Rates of Dropouts and Non-dropouts



Note: Dropouts in grey, non-dropouts in black. The dashed lines are 95% confidence intervals.

of them are employed while (per definition) none of the non-dropouts is employed. The employment rate of non-dropouts begins to rise slightly two months before the planned end (remember that non-dropout is defined as attending at least 80% of the planned duration) and rises sharply in the month of the planned end and the following months. 14 months after the planned end the head start of dropouts has vanished and employment rates of dropouts and non-dropouts are equal. In the end of the observation period dropouts do a little bit better, but it is not clear if this is significant.

Figure 4.2: Employment Rates by Employment Status after Program



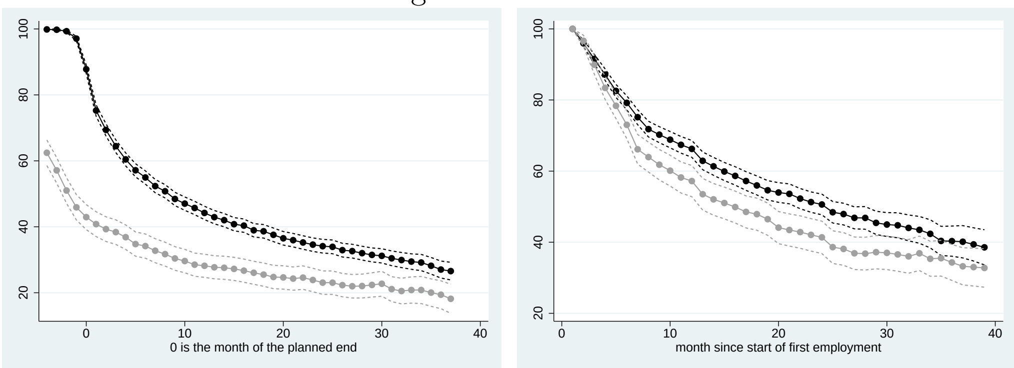
Employment rates for those employed in the month after the program

Employment rates for those not employed in the month after the program

Note: Dropouts in grey, non-dropouts in black. The dashed lines are 95% confidence intervals.

Figure 4.2 shows the same rates as figure 4.1 but separately for participants who are employed in the first month after the realized end of participation (figure on the left) and those who are not employed in the first month after the realized end of participation (figure on the right).[13] The figure on the left hand side suggests that in the long run job-aligned dropouts do a bit worse than those participants who start a job right after completing the program. According to the figure on the right hand side the employment rate of non-job-aligned dropouts is in the long run a little lower compared to the rate of those who complete but do not directly start employment after leaving the program. Note that figure 4.2 is not inconsistent with figure 4.1, because the share of those who are employed in the month after the programs is higher for dropouts than for non-dropouts.

Figure 4.3: Survival Rates



Survival rate in unemployment   Survival rate in first employment

Note: Dropouts in grey, non-dropouts in black. The dashed lines are 95% confidence intervals.

The diagram on the left hand side of figure 4.3 shows the rate of those who have not left the unemployment period in focus in the respective quarter. Non-dropouts survive longer in unemployment. The difference decreases over time but does not vanish completely. This difference to figure 4.1 indicates that dropouts leave unemployment faster but their employment is less stable. The figure on the right hand side shows the rate of those individuals who are still in their first employment period in the respective month (based on those individuals who start employment during the observation period). Month one is the first month the individual is employed in, irrespective of when he or she left the program. Two years after starting employment 41% of the dropouts and 51% of the non-dropouts are still employed. Thus, employment of non-dropouts is a bit more stable.

In sum, the descriptive analysis gives the impression that dropouts enter employment earlier but non-dropouts catch up after some time, and that employment of non-dropouts is slightly more stable. But from the descriptive analysis we can of course

[13]The realized end of participation differs between individuals relative to the planned end, therefore there is no month with an employment rate of 100%.

not infer if there exists a negative or positive effect of dropout, because the treatment dropout is not randomized, there will be selection based on observable variables and unobservables. Even the direction of selectivity is not clear ex ante. On the one hand it may be that those participants who drop out are on average those with better employment chances or a higher motivation for employment (for example having a high utility of earning a salary in the short run). Possible reasons for an increased dropout rate among them may be that they are more likely to receive a job offer, more likely to drop out due to a job offer or more likely to conclude that they do not need the program and prefer to intensify job search instead of attending the program. On the other hand it may be that those participants with characteristics which deteriorate employment prospects (like low schooling, low general ability or a low motivation for work) tend to drop out, because they find it also hard to complete the program. The nature of selectivity may also be more complex and, thus, positive and negative characteristics with respect to employment chances may partly cancel out even on the individual level: the results of the probit estimation in section 4.3.2 suggest for example a negative correlation between schooling and dropout and a positive correlation between having to support a family (proxied by men who have a child under the age of ten) and dropout for participants of retraining programs.

4.4 Joint Estimation of Dropout and Employment: Does Dropout Harm in the Long Run?

4.4.1 The Model

The descriptive analysis in the previous section suggests that dropouts enter employment earlier but non-dropouts catch up after some time. But purely descriptive analysis does not provide insights if this is an effect of dropping out. Therefore, in this section I use a bivariate dynamic random effects probit model to jointly estimate the dropout probability and the employment probability taking into account unobserved heterogeneity. To see how this may account for various differences of dropouts and non-dropouts which are potentially included in a simple comparison of employment rates like in section 4.3.2, start by considering this purely descriptive difference of average employment rates. Now, first think of estimating a simple pooled probit of an employment dummy on observable variables for the periods in which participants have finished program participation. The variable of interest would be a dummy if the individual has dropped out of the program in the past. Compared to purely descriptive analysis this will account for differences between dropouts and non-dropouts due to observables like schooling, age, last occupation, different labor market histories or the planned length of the program. It may also take into account time, season, and labor market conditions in the region. The estimation will also deal with state dependence. While in the descriptive analysis the initial luck of a job-aligned dropout will still influence the employment status a few months later if state dependence exists, the dynamic model will account for this. When estimating the effect of dropout in the past, the luck the dropout had in the past to receive a job offer will not influence the estimates in later periods. Second consider estimating a dynamic random effects probit model of an employment dummy. This will in addition include a time-constant unobserved heterogeneity term and thus account for time-constant differences of individuals' propensity to be employed. Thus, estimating a separate dynamic random effects probit model of an employment dummy will already take out much of the differences between dropouts and non-dropouts. But dropout may still be endogenous in the way that it is correlated with the error of the employment equation due to some dependence of unobserved dropout propensity and unobserved employment propensity. This problem is accounted for by introducing a dropout equation and estimating it simultaneously with the employment equation. In the following, firstly the model is presented and then the model assumptions are discussed.

The dropout equation is a random effects probit model of a dropout dummy:

$$Drop^*_{it} = \beta_D x_{it,D} + \alpha_{i,D} + \epsilon_{it,D} \tag{4.1}$$

where $Drop = \mathbf{1}[Drop^* > 0]$.

The equation is estimated for the first time in the month participants start the program and estimation goes on until either participants drop out or they have reached 80% of the planned duration so that they cannot drop out anymore according to the definition of dropout. The dropout equation may also be interpreted as a hazard model with right censoring when the program is finished. The vector of independent variables β_D includes the following information: remaining time until the planned end, a dummy if the person is still in the beginning of the program, and observable information on schooling, age, gender, family, if the last job was a blue collar job, past sanctions or signs of lack of motivation with regard to activities of the labor agency, health problems, East or West, earlier contact with the labor agency and past program participation. Other information like for example wages or occupation of the last job, year, season or detailed regional information turned out to be irrelevant.

Now consider the employment equation, which is a random effects probit model of an employment dummy:

$$E^*_{it} = \delta_E DropInPast_{it,E} + \beta_E x_{it,E} + \alpha_{i,E} + \epsilon_{it,E} \tag{4.2}$$

where $E = \mathbf{1}[E^* > 0]$.

$\alpha_{(i,E)}$ and $\alpha_{(i,D)}$ follow a **joint** normal distribution. $\epsilon_{(it,E)}$ and $\epsilon_{(it,D)}$ are independently standard normal distributed. Thus, the model includes two individual effects which are allowed to be correlated and represent the link between the two equations.

The employment equation is estimated once the individuals are again available for employment (which is in the month after they have left the program) until month 40 counted from program start onwards. On average the employment status is estimated for 32.2 periods for general professional training, for 23.9 periods for retraining and for 34.2 periods for practical training. The equation accounts for state dependence and duration dependence by including all employment lags (they are set to zero also for periods in which the individual has not reached a period in which the lag may take a one), elapsed unemployment or employment duration in months since the end of participation (also squared), planned length of the program, and information on the employment history before program participation. By using separate variables for the elapsed duration in unemployment, if applicable, and the elapsed duration in employment, if applicable, the model allows for different non-linear patterns of duration dependence in employment or unemployment, respectively. Observables like for example schooling, professional qualification, gender, health problems, region, wage in last employment, year and season are added. Dummies capture the alignment to the planned end of the program; they indicate if the current period lies before the planned end or in which month after the planned end, respectively, whereby later months are summarized. The effect of interest is the medium and long run effect of dropout. This is captured by a dummy variable called *dropout in past* indicating that the individual has dropped out in the past and has now reached at least month four after the planned end of the program.

The model also includes a dummy for dropout in the last period and for dropout which has occurred in the past given that the person has not reached month four after the planned end. The model does not provide causal estimates for the short-term effects of dropout. If participants drop out because they were lucky to receive a job offer, $\epsilon_{it,D}$ of the last period in which the dropout equation is estimated and $\epsilon_{it,E}$ of the first period the employment equation is estimated are correlated. The short-term effect of dropout is, in a way, the other side of the coin of the lock-in effect usually found when evaluating employment effects of training programs. It has to be either positive or zero. From the descriptive analysis in section 4.3.2 we know that many dropouts are employed soon after dropout while non-dropouts are by definition not employed while they are locked in the program. Thus, there is certainly a positive short-term effect of dropout, but it is not possible to infer to which extent this is a causal effect of dropout, i.e. what would be the size of the short-term effect of dropout if dropout was a randomized treatment. The medium-term and long-term effect of dropout may be estimated, because the model allows for state dependence and duration dependence and, thus, the initial luck which may have influenced the decision to drop out may be accounted for in later periods.

The model estimated in this paper shares important features with the timing-of-events approach proposed by Abbring and van den Berg (2003) in the context of continuous duration models. The model of Abbring and van den Berg (2003) and the model in the present paper both consist of two equations: one relating to the treatment and one to the outcome of interest. Both models allow for duration dependence and for unobserved heterogeneity terms in both equations which are allowed to be dependent. Abbring and van den Berg (2003) show that in their model and under the assumptions they make (mainly conditional randomness of treatment starts, a no-anticipation assumption and functional form assumptions) the treatment effect may be separated from the selection effect. The unobserved heterogeneity term of the outcome equation is identified from the competing risk part of the model and the treatment effect is then identified from differences in hazard rates (Abbring and van den Berg (2003)). The present model also uses the timing-of-events and relies in addition on functional form assumptions. Apart from a different model specification and from using discrete data, the present model differs in two main aspects from the model by Abbring and van den Berg (2003). First, the employment equation only kicks in if the individual has left the program: this is because there is no third state (like employment in the model of Abbring and van den Berg (2003)) involved at the beginning. Starting employment before having reached the planned end of the program necessarily involves the treatment dropout. Second, the period close to the one the treatment occurs must not be used for identification, because soon after dropout individuals may be employed because they were lucky to receive a job offer due to which they choose the treatment dropout. Thus, the outcome soon after the treatment is linked to the treatment due to other factors in addition to a possible causal effect of treatment. In the discrete model this endogeneity problem may be solved for later periods by accounting for state dependence and duration dependence and by relying on the functional form of the model.

To estimate a causal effect of past dropout, some exogenous variation in the decision to drop out as well as the functional form assumptions of the model (including the assumption that the random effects are uncorrelated with observed variables) are needed. The timing of the model is the following: in the first period all individuals start a program. In each of the following periods participants decide to continue to attend or to drop out. In the sense of the literature on treatment effects, dropout would be the treatment. From the month after the individual has left the program onwards, the employment status is estimated taking into account the time-constant unobserved propensity to employment, the estimation of which takes into account the time-constant unobserved propensity to drop out through the correlation of the two random effects, both being estimated simultaneously. For identification it is necessary that there is exogenous variation that influences the dropout decision. By exogenous I mean factors that do not directly influence long-term employment prospects conditional on observed and time-constant unobserved characteristics. Exogenous factors may for example be randomness in expectations between participants due to different information and dropout if expectations are not met, personal dislike of the teacher or classmates, temporarily higher opportunity costs (for example due to a work opportunity on the black market), changes in preferences, lack of endurance in relation to training or differences in individual discounting of the future. Factors leading to dropout which are not captured by observed or unobserved variables of the model and which have a direct **long-term** effect on employment (not only through state dependence) would violate the model assumptions. This means that dropout due to the luck of receiving a job offer does not violate the model assumptions, because lagged employment is endogenized in the dynamic model. If, however, the dropout occurred only because by pure luck a job with a long-term contract is offered this would violate the model assumptions because the sort of contract may not be controlled for. Similarly dropout to non-employment due to factors that do influence the long-time employment prospects and which neither can be controlled for nor are time-constant (one example would be a pregnancy) would bias the estimated effect of past dropout. These two examples show that biases may go into both directions, and if biases exist they might to some extent cancel out. As there is no instrument for dropout available, I think estimating the bivariate dynamic model is all that can be done to identify the effect of dropout. Anticipation of dropout is not a problem in this model because dropout to take up employment involves per definition dropout and employment, and dropout to leisure involves per definition dropout and non-employment - there is no third state involved. Dropout reflecting a strategy of participants to choose the optimal time to leave the program (considering to drop out in case the right job is offered or the economic situation in the region is favorable) would be an effect of dropout and does not violate the model assumptions.

4.4.2 MCMC Estimation

To estimate the bivariate model presented in the previous section complex estimation techniques are needed. In principle the estimation could be done by maximum likelihood, but as the individual specific effects $\alpha_{it,D}$ and $\alpha_{it,E}$ are not observed one would have to integrate them out and simulate the multivariate normal integrals. I made an attempt to estimate the model using GLLAMM (a Stata routine for multilevel models), but this turned out to be far too time-consuming. Even the estimation of a much simplified one factor model ran too long to be practically applicable. But with Markov Chain Monte Carlo (MCMC) simulation methods, a technique from Bayesian statistics, an attractive alternative to maximum likelihood is available.[14] The idea of MCMC methods is to obtain a large sample from the posterior distribution of the parameters. From a classical perspective, the mean of the posterior distribution converges to the maximum of the likelihood function and the variance of the posterior distribution converges to the asymptotic variance of an ML estimation. Thus, the standard deviation of the draws may be interpreted as

[14]Chib (2001) reviews important concepts of MCMC simulation methods.

standard errors from the classical perspective (Train, 2003). To obtain the sample from the posterior distribution I use a Gibbs sampler, which works by forming blocks of the model parameters and then drawing in turn from the conditional distributions of the blocks of parameters. The resulting sequence is a Markov Chain and after convergence the draws are samples from the desired posterior distribution. The key idea for the estimation of probit models is to estimate the latent variables as one step of the simulation (Albert and Chib, 1993). A similar strategy is used for the random effects (Zeger and Karim, 1991). Odejar (2002) proposes a Gibbs sampler for a model sharing important features with the one estimated in this paper. Recent examples for economic applications of very much related models are Buchinsky et al. (2005) and Fitzenberger et al. (2009). Details of the algorithm are given in Appendix B. I programmed the Gibbs sampler in Stata. For the calculation-intensive steps of the algorithm I used Mata, the matrix programming language of Stata. Conjugate but very diffuse priors are used. The results reported below are based on running the algorithm for 20,000 iterations. Convergence is monitored by comparing the means at different stages of the chains. The first 5,000 iterations are discarded (burn-in phase). Thus, the results are based on 15,000 draws. Covariates have been selected considering the size of the effects, significance (based on posterior means and standard deviations) and economic importance. Several interactions have been tested but turned out to be statistically irrelevant.

It is an important advantage of the MCMC estimation that it provides information on all parameters of the model including information on the unobserved individual specific effects $\alpha_{i,D}$ and $\alpha_{i,E}$. This information is needed to calculate average partial effects on the treated. To calculate these effects it is a natural solution to get an estimate of the average treatment effect on the treated which takes into account the selection on unobservables. To get these effects I developed the following strategy: for every tenth iteration of the MCMC estimation I calculate the partial effect for all person-periods in which the variable *dropout in past* takes a one. This strategy uses the β_E and δ_E vector of the respective iteration and the predictor of the $\alpha_{i,E}$ of the respective iteration together with the $x_{it,E}$ of the person-period. Averaging this effect over the person-periods gives a draw of the posterior distribution of the average partial effect of dropout in the past. The resulting 1,500 draws may then be used to describe the posterior distribution of the average partial effect of dropout for those who have dropped out in the past. This distribution may be described by giving the mean and the standard deviation, so information on statistical significance is readily available and there is no need to calculate standard errors (for instance using the delta method) as for classical estimators.

4.4.3 Results

Table 4.5 in Appendix A shows the results of the MCMC estimation. The posterior distributions of the parameters are summarized by means and standard deviations. First consider the variance parameters. An important part of the variance of both equations is on the individual level: in between 34% and 44% for the employment equation and in between 37% and 39% for the dropout equation. The correlation between the two random effects is relatively strong (36.9%) and significant for general professional training. A positive correlation suggests that those who have a higher propensity to be employed have also a higher propensity to drop out. For retraining, the correlation is also positive (27.9%), but insignificant. For practical training the estimation suggests a negative (-16.8%) and insignificant correlation. A negative correlation indicates that those unobserved characteristics that make a dropout more likely also decrease the employment probability. In theory both positive and negative correlations seem plausible. For the latter one could for instance think of general motivation of career improvement captured in the individual effect. It is plausible that someone who has a high motivation to study and work hard may be less likely to drop out of an offered training program and in general more likely to be employed. For a negative correlation one could also think of someone who has problems to comply with social norms and rules, such a person would have an increased risk to drop out of the program and also be less likely to succeed in finding employment or keeping a job in the long run. With respect to a positive correlation one could think of a high utility of earning money in the short run captured in the individual effect. Someone who is keen on earning a salary in the short run, for instance because of high discounting of the future or because he is the only earner in a family, is on the one hand likely to drop out of the program if he has chances to find some job. On the other hand, he will put a lot of effort in finding a job and not becoming unemployed again. A positive correlation between the random effects is also likely if participants with high ability find that the level of the programs is too low for them and thus tend to drop out.

The parameter of interest is the effect of the variable *dropout in past*. This dummy takes a one if the person experienced a dropout in the past and the current time period lies at least four months after the planned end of the program. The means of the posterior distribution (see table 4.5 in Appendix A and also the bottom line in table 4.2) suggest a negative effect of *dropout in past* for general professional training and for retraining. For practical training the effect is positive, but all three effects are insignificant on the 5% level. To estimate the size of the effect, I calculate average partial effects on the treated using the strategy described at the end of the previous section. As described above this strategy takes into account the selection based on unobservables by including the predictors of the α_E. The first line in table 4.2 depicts the results. They suggest that in the medium and long run dropout decreases employment chances of those who actually have dropped out only by 1.8 percentage points for general professional training and 3.1 percentage points for retraining. The effect of dropout is +2.2 percentage points for practical training. These effects are small compared to the effects of program participation (see for example Fitzenberger et al. (2009)). All three effects are insignificant, even on a 10% level. Thus the hypothesis that dropping out has no long run effect on dropouts can not be rejected.[15]

Table 4.2: Estimation Results

	General Prof.		Retraining		Practical	
	Average partial long-term effect of dropout:					
	Mean	SD	Mean	SD	Mean	SD
dropout in past	-0.018	0.011	-0.031	0.026	0.022	0.014
	Parameters from MCMC estimation:					
	Mean	SD	Mean	SD	Mean	SD
dropout in past	-0.250	0.134	-0.342	0.275	0.374	0.220

The small and insignificant effects of dropout in the past might hide effect heterogeneity in several dimensions. It might be that dropout has a positive long-term effect for those who drop out with a job perspective and a negative effect for non-job-aligned dropouts, and these effects might have canceled out in the estimation. Also, dropout in the past may have a different effect on finding employment as on staying in employment. The descriptive analysis suggests a lower job stability for dropouts. Also, the interaction of these two dimensions might be relevant.[16] Table 4.6 in Ap-

[15]Estimations of a simple pooled probit and a separate ML estimation of the employment equation also suggest only insignificant effects of the dummy *dropout in past*.

[16]Additional estimations (not shown in the paper) have shown that effect heterogeneity over time is not relevant. The effect of dropout in the past for example half a year after the planned end is not systematically different for the effect of dropout one year after the planned end of the program.

pendix A shows the results when separating the effect of *dropout in past* between job-aligned and non-job-aligned dropout. Job-aligned dropouts are those who are employed in the first month after dropout, which is also the first period estimated in the employment equation. Dropping out job-aligned versus non-job-aligned may be interpreted as two different treatments. The distinction is endogenous in the model. Furthermore, the dropout effect is separated with respect to those who are employed in the last period and those who are not employed in the last period (effect on finding a job or keeping a job). Table 4.3 gives the average partial effect on the treated (calculated as above). Note that the estimated partial effect of job-aligned dropout is the effect of a dropout in the past and employment in the first month after leaving the program versus completing and being employed in the first month, and the analogous for non-job-aligned dropout.

Table 4.3: Estimation Results (Flexible Specification)

	General Prof.		Retraining		Practical Tr.	
	Mean	SD	Mean	SD	Mean	SD
Average partial effect of dropout in the past:						
job-aligned and e[t-1]=0	-.030	.021	.021	.043	.013	.014
job-aligned and e[t-1]=1	-.019	.008	-.002	.015	.032	.027
non-job-aligned and e[t-1]=0	-.015	.013	-.029	.026	.012	.006
non-job-aligned and e[t-1]=1	-.009	.013	-.022	.024	.091	.053

With regard to general professional training all effects are again very small and negative. The negative effect of a job-aligned dropout on job stability is significant but very small (-1.9%). The effect of a job-aligned dropout on finding a new job is a bit larger (-3%) but insignificant. Thus, there is some slight evidence that dropping out because of a job offer might be a little bit harmful in the long run for dropouts from general professional training. For retraining the effects are a bit less negative than in the less flexible estimation and they are again all insignificant. Concerning practical training the effects of a non-job-aligned dropout on finding employment is slightly significant but very small (1.2%). There is one effect which is relatively large (though almost insignificant, p-value: 0.09%): the effect of a non-job-aligned dropout on employment stability. This says that those who are employed but have not been employed in the first period after leaving the program do better in keeping employment as opposed to the counterfactual situation in which they would have completed the program. But only 5% of the participants in practical training experience such a combination, so the estimated size of the effect is based only on a few people. To sum up, the results suggest that the effects for dropout are zero or very small. Studies evaluating the employment effects of participation in further training programs (see for example Fitzenberger et al. (2009)) conclude that further training programs have positive long-term effects. According to Fitzenberger et al. (2009) the size of these effects amounts to an increase in employment of 10 to 20 percentage points depending on the group of participants. If in this context the effect of dropout for those who actually drop out is very small, this might for example indicate that for those participants who drop out it is enough to attend part of the program (for example because the programs work through activation of the participants) or that those choose to drop out who have a low benefit from training programs.

4.5 Conclusion

This study has shown that dropout is a relevant phenomenon in further training programs for the unemployed and that it occurs job-aligned as well as without a job perspective. One out of five participants drops out of the program. The first objective of this paper was to identify dropouts in the IEBS and to gain knowledge about the occurrence of dropouts - how often and when do people drop out and which characteristics are related to an increased probability of dropout. It is possible to distinguish participants that attend at least 80% of the program from those who drop out if taking into account some particularities and sensitivities of the data. Practical training is the program type with the highest dropout rate. Less than half of the dropouts take up employment within one month. Results of a probit model estimating the probability to drop out indicate that for instance low schooling, being young and living alone is related to an increased dropout probability. Participants for whom signs of lack of motivation with regard to activities of the labor agency can be identified from the data also face an increased probability to drop out. To study the employment prospects of dropouts I first use purely descriptive analysis. Comparing employment rates of dropouts and non-dropouts shows that the head start of dropouts decreases over time and 14 months after participants have reached the date of the planned end of the program the employment rates of dropouts and non-dropouts intersect. Survival rates indicate that the first employment of dropouts is a bit less durable than the first employment of those who completed the measure.

Dropout and employment status are jointly estimated using a bivariate dynamic random effects probit model. The individual effects of the two equations are allowed to be correlated. The model is estimated using Markov Chain Monte Carlo (MCMC) methods. I programmed a Gibbs sampling algorithm to simulate draws from the posterior distribution of the parameters. This approach provides information on all parameters of the model, including the unobserved individual specific effects. To get an estimate for the size of the dropout effect, I calculated average partial effects on the treated which account for the selection based on unobservables.

Results suggest that long-run effects of dropout are very small and insignificant. A more flexible estimation with respect to job-aligned and non-job-aligned dropout and with respect to transition to employment and transition to non-employment, respectively, shows only small effects. Thus, on average the decision to drop out neither harms nor enhances future employment prospects. On the one hand, this result is in line with the hypothesis that those participants who drop out stay in the program only as long as the program offers a positive benefit to them. Dropouts would for example leave the program if they have acquired the skills they need to find a job or if they gained enough orientation to apply for jobs. Non-dropouts would exploit the whole program duration to derive the benefits of the program. On the other hand, zero or small effects of dropout may also be due to heterogeneity of programs and participants. If those who do not benefit from the program - due to their skills or due to the quality of the specific course they attend - drop out, the effect of dropout may be zero even if on average there is a positive treatment effect of training programs. Furthermore, it is possible that to some extent positive and negative effects offset each other: It might be that some participants benefit from dropping out (for example due to good timing with respect to reentering the labor market) whereas other participants are harmed by a dropout. Nevertheless, the results do not seem to support the hypothesis of strong and widespread negative effects of dropout due to a negative signal, incomplete attendance of the course or not obtaining a certificate.

References

Abbring, J. and G.J. van den Berg (2003). "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica* 71, 1491-1517.

Albert, J. and S. Chib (1993). "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association*, 88, 669-679.

Becker, S. (2005). "Introducing Time-to-Educate in a Job Search Model." IZA Discussion Paper No. 1801.

Biewen, M., B. Fitzenberger, A. Osikominu and M. Waller (2007). "Which program for whom? Evidence on the comparative effectiveness of public sponsored training programs in Germany." IZA Discussion Paper No. 2885.

Buchinsky, M., D. Fougère, F. Kramarz and R. Tchernis (2005). "Interfirm Mobility, Wages, and the Returns to Seniority and Experience in the U.S." IZA Discussion Paper No. 1521.

Bundesagentur für Arbeit (2000, 2001, 2002). *Arbeitsmarkt 2001, 2002, 2003*, Nürnberg (various issues).

Chib, S. (2001). "Markov Chain Monte Carlo Methods: Computation and Inference." In J. Heckman and E. Leamer (eds.): *Handbook of Econometrics*, Volume 5, Amsterdam: Elsevier Science, 3569-3649.

Fitzenberger, B., A. Osikominu and M. Waller (2009). "The Heterogeneous Effects of Training Incidence and Duration on Labor Market Transitions." mimeo, Albert-Ludwigs-University Freiburg.

Flores-Lagunes, A., A. Gonzalez and T. Neumann (2007). "Estimating the Effects of Length of Exposure to a Training Program: The Case of Job Corps." IZA Discussion Paper No. 2846.

Heckman, J., J. Smith and C. Taber (1998). "Accounting for Dropouts in Evaluations of Social Programs." *The Review of Economics and Statistics*, 80, 1, 1-14.

Hujer, R., S. Thomsen and C. Zeiss (2006). "The Effects of Vocational Training Programmes on the Duration of Unemployment in Eastern Germany." *Allgemeines Statistisches Archiv* 90, 299-321.

Kluve, J., H. Schneider, A. Uhlendorff and Z. Zhao (2007). "Evaluating Continuous Training Programs Using the Generalized Propensity Score." IZA Discussion Paper No. 3255.

Lechner, M. and C. Wunsch (2006a). "Active Labour Market Policy in East Germany: Waiting for the Economy to Take Off." IZA Discussion Paper No. 2363.

Lechner, M. and C. Wunsch (2006b). "Are Training Programs More Effective When Unemployment Is High?" IZA Discussion Paper No. 2355.

Lechner, M. and C. Wunsch (2007). "The Curse and Blessing of Training the Unemployed in a Changing Economy: The Case of East Germany after Unification." *German Economic Review*, 8, 468-509.

Lee, M. and S. Lee (2003). "Analysing Effects of Job-Trainings Suffering Dropouts with an Optimal Multiple Matching." mimeo, Singapore Management University.

Odejar, A. (2002). "Bayesian Analysis of Sample Selection and Endogenous Switching Regression Models with Random Coefficients Via MCMC Methods." SFB 386 Discussion Paper No. 291.

Osikominu, A. (2008). "Is Short Training Short-Lived and Long Training Long-Lasting? A Multi-State Duration Analysis of the Dynamic Effects of Training Schemes for the Unemployed." Discussion Papier, Albert-Ludwigs-University Freiburg.

Rinne, U., M. Schneider and A. Uhlendorff (2007). "Too Bad to Benefit? Effect Heterogeneity of Public Training Programs." IZA Discussion Paper No. 3240.

Rosholm, M. and M. Svarer (2008). "The Threat Effect of Active Labour Market Programmes." *Scandinavian Journal of Economics*, 110, 385-401.

Schneider, H. and A. Uhlendorff (2006). "Die Wirkung der Hartz-Reform im Bereich der beruflichen Weiterbildung." *Journal for Labor Market Research* 39, 477-490.

Stephan, G. and A. Pahnke (2008). "The Relative Effectiveness of Selected Active Labour Market Programmes and the Common Support Problem." IZA Discussion Paper No. 3767.

Train, K. (2003). "Discrete Choice Methods with Simulation." Cambridge University Press.

Waller, M. (2008). “On the Importance of Correcting Reported End Dates of Labor Market Programs.” *Schmollers Jahrbuch (Journal of Applied Social Science Studies)* 128, 213-236.

Wunsch, C. and M. Lechner (2008). “What Did All the Money Do? On the General Ineffectiveness of Recent West German Labour Market Programmes.” *Kyklos*, 61, 134-174.

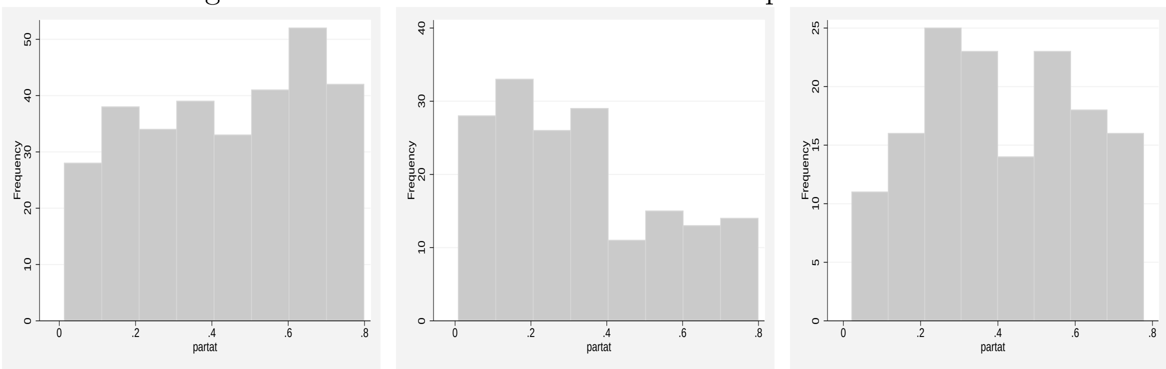
Zeger, S. and M Karim (1991). “Generalized Linear Models with Random Effects: a Gibbs Sampling Approach.” *Journal of the American Statistical Association*, 86, 79-86.

Zimmermann, R., S. Kaimer and D. Oberschachtsiek (2007). “Dokumentation des ”Scientific Use Files der Integrierten Erwerbsbiographien” (IEBS-SUF V1) Version 1.” FDZ Datenreport 1/2007, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Appendix A

Descriptive Results

Figure 4.4: Share of Planned Duration Dropouts Attend



General Professional Training Retraining Practical Training

Table 4.4: Cross-sectional Probit of Dropout Dummy by Program Type

	Gen. Prof. Train.	Retraining	Practical Train.
no schooling degree	0.163 (0.060)	0.202 (0.092)	0.080 (0.075)
lower second. schooling	0.024 (0.023)	0.078 (0.033)	-0.049 (0.046)
sanction in past	0.100 (0.083)	0.064 (0.200)	0.073 (0.119)
lack of motivation in past	0.077 (0.034)	0.117 (0.048)	0.037 (0.059)
healthproblem	0.008 (0.032)	0.012 (0.052)	0.027 (0.056)
female	0.026 (0.023)	0.059 (0.036)	-0.122 (0.048)
living in East Germany	-0.012 (0.021)	-0.068 (0.038)	-0.146 (0.045)
25 to 29 years old	0.045 (0.031)	0.063 (0.038)	0.032 (0.065)
30 to 35 years old	0.050 (0.027)	-0.036 (0.036)	0.056 (0.056)
training program in past	0.022 (0.035)	0.153 (0.077)	0.136 (0.077)
unemployed before	0.048 (0.024)	0.028 (0.037)	0.053 (0.050)
last job blue-collar	0.006 (0.022)	-0.080 (0.035)	0.039 (0.046)
industry with seasonal work	-0.029 (0.043)	-0.030 (0.060)	0.110 (0.092)
living alone	0.028 (0.023)	0.130 (0.041)	0.050 (0.045)
child under 10 * female	0.023 (0.038)	0.061 (0.069)	0.044 (0.087)
child under 10 * male	-0.008 (0.035)	0.251 (0.069)	-0.100 (0.059)
planned length in months	0.004 (0.002)	0.001 (0.002)	0.021 (0.007)
constant	-1.407 (0.146)	-1.337 (0.269)	-0.967 (0.273)

Average partial effects with standard errors in brackets.

MCMC Estimation Results

Table 4.5: Results of MCMC Estimation (Simple Specification)

	General		Retraining		Practical	
Variable	Mean	SD	Mean	SD	Mean	SD
Employment equation:						
last month dropout	0.487	0.185	0.010	0.301	1.229	0.272
dropout in recent past	-0.695	0.171	-0.917	0.281	-0.161	0.256
dropout in past	-0.249	0.151	-0.342	0.275	0.374	0.220
month of planned end or before	0.314	0.123	-0.000	0.178	0.663	0.210
month 1 after planned end	0.119	0.095	0.034	0.151	0.502	0.170
month 2 after planned end	-0.251	0.091	-0.071	0.137	0.033	0.160
month 3 after planned end	-0.147	0.083	-0.140	0.134	-0.022	0.153
month 4 to 6 after planned end	-0.184	0.058	-0.224	0.095	-0.156	0.108
month 7 to 12 after planned end	-0.144	0.041	-0.035	0.071	-0.126	0.077
month 24 to 40 after planned end	0.072	0.046	0.133	0.145	0.073	0.086
e[t-1]	2.928	0.073	2.905	0.127	3.145	0.140
e[t-2]	-0.192	0.059	-0.239	0.108	-0.172	0.114
e[t-3]	0.009	0.058	-0.097	0.102	-0.071	0.109
e[t-4]	0.014	0.059	0.109	0.107	0.004	0.108
e[t-5]	-0.000	0.063	-0.155	0.109	0.146	0.114
e[t-6]	-0.036	0.054	0.004	0.096	0.019	0.102
$\sum_{j=7}^{12} e[t-j]$	-0.001	0.010	0.003	0.020	0.001	0.019
$\sum_{j=13}^{18} e[t-j]$	-0.060	0.010	-0.061	0.023	-0.057	0.018
$\sum_{j=19}^{24} e[t-j]$	-0.035	0.012	-0.036	0.029	0.002	0.021
$\sum_{j=38}^{25} e[t-j]$	-0.028	0.010	-0.057	0.026	-0.026	0.018
elap. mon. in current state & e[t-1]=1	0.001	0.011	-0.029	0.023	-0.001	0.019
elap. mon. in current state squ. & e[t-1]=1	0.001	0.000	0.002	0.001	0.000	0.001
elap. mon. in current state & e[t-1]=0	-0.038	0.009	-0.054	0.015	-0.010	0.016
elap. mon. in current state squ. & e[t-1]=0	0.001	0.000	0.002	0.000	-0.000	0.000
planned length in days/31	-0.010	0.006	0.016	0.007	-0.027	0.018

<continued on next page>

Results of MCMC Estimation (Simple Specification) <continued>

Variable	General Mean	General SD	Retraining Mean	Retraining SD	Practical Mean	Practical SD
female	0.050	0.049	0.168	0.097	0.133	0.108
living in East Germany	-0.060	0.058	-0.414	0.137	-0.023	0.122
no vocational degree	0.005	0.066	-0.144	0.091	-0.224	0.115
no schooling degree	-0.093	0.118	0.277	0.232	-0.327	0.183
lower secondary (Hauptschule)	-0.064	0.057	-0.109	0.108	0.055	0.112
high school (Abitur)	-0.054	0.059	-0.202	0.128	0.002	0.147
25-29 years old	0.156	0.065	0.023	0.108	0.098	0.135
30-34 years old	0.066	0.059	-0.163	0.105	0.184	0.119
35-40 years old	-0.122	0.060	-0.179	0.187	-0.112	0.119
50-54 years old	-0.405	0.082	-0.071	0.360	-0.637	0.148
child under 10 * male	0.144	0.075	-0.059	0.150	0.204	0.153
child under 10 * female	-0.056	0.087	-0.311	0.184	-0.018	0.183
health problems	-0.105	0.085	-0.152	0.171	-0.076	0.151
unemployed before (last three years)	0.056	0.058	0.042	0.105	-0.314	0.117
days/31 unempl. assistance last 3 years	-0.023	0.007	-0.020	0.012	-0.028	0.011
training program before	-0.054	0.079	-0.272	0.204	-0.156	0.172
days/31 employed last 3 years	0.006	0.003	0.009	0.006	0.004	0.006
spring (second quarter)	0.023	0.030	0.019	0.058	0.072	0.056
fall (fourth quarter)	-0.110	0.030	-0.035	0.056	-0.208	0.055
winter (first quarter)	-0.061	0.031	-0.190	0.061	-0.049	0.060
year 2002	-0.079	0.035	-0.085	0.078	0.101	0.065
year 2003	0.028	0.045	-0.053	0.077	0.178	0.084
year 2004	0.061	0.060	-0.108	0.094	0.107	0.108
log last real wage	-1.094	0.682	-0.652	1.530	-0.511	1.129
log last real wage squared	0.058	0.036	0.033	0.082	0.028	0.061
last job in industry with seasonal work	-0.042	0.103	0.089	0.171	0.008	0.185
urban region high unempl. in East	-0.107	0.089	0.287	0.248		
urban region, good conditions in West	-0.066	0.096	-0.143	0.163	0.525	0.243
non-urban region, good conditions in West	0.129	0.061	0.267	0.110	0.037	0.107
constant	3.991	3.203	2.057	7.118	1.090	5.282

<continued on next page>

Results of MCMC Estimation (Simple Specification) <continued>

Variable	General Mean	General SD	Retraining Mean	Retraining SD	Practical Mean	Practical SD
Dropout equation:						
days/31 until planned end	-1.962	0.453	-0.134	0.192	-3.068	0.873
near to program begin	-0.073	0.088	-0.093	0.109	-0.209	0.132
no schooling degree	0.600	0.176	0.644	0.251	0.161	0.210
lower secondary (Hauptschule)	0.178	0.090	0.301	0.114	-0.148	0.145
sanction in last 3 years	0.507	0.257	-0.016	0.567	0.304	0.346
lack of motivation w.r.t. agency's activities	0.229	0.112	0.351	0.143	0.059	0.172
health problems	0.037	0.148	-0.023	0.215	-0.220	0.214
female	0.101	0.092	0.127	0.130	-0.377	0.156
living in East Germany	-0.100	0.084	-0.286	0.166	-0.533	0.163
25-30 years old	0.143	0.109	0.193	0.126	0.203	0.202
30-34 years old	0.140	0.098	-0.199	0.139	0.201	0.169
training program before	0.102	0.126	0.502	0.218	0.415	0.209
unemployed before (last three years)	0.177	0.105	0.152	0.147	0.183	0.168
last job as blue-collar worker	0.063	0.089	-0.279	0.126	0.173	0.144
last job in industry with seasonal work	-0.118	0.194	-0.085	0.234	0.220	0.254
living alone	0.147	0.094	0.474	0.127	0.041	0.140
child under 10 * female	0.135	0.137	0.212	0.203	0.066	0.262
child under 10 * male	0.036	0.140	0.759	0.201	-0.391	0.232
constant	-2.641	0.165	-3.018	0.225	-1.629	0.239
Individual level variances:						
individual level variance employ. equ.	0.505	0.064	0.811	0.161	0.612	0.118
individual level variance dropout equ.	0.653	0.179	0.660	0.258	0.610	0.211
individual level covariance	0.215	0.100	0.209	0.156	-0.107	0.123
share on individual level, employ. equ.	0.334	0.028	0.444	0.048	0.376	0.045
share on individual level, dropout equ.	0.388	0.061	0.385	0.083	0.369	0.074
correlation between equations	0.134	0.057	0.118	0.083	-0.064	0.072
correl. between random effects	0.369	0.139	0.279	0.182	-0.168	0.185

Table 4.6: Results of MCMC Estimation (Flexible Specification)

	General		Retraining		Practical	
Variable	Mean	SD	Mean	SD	Mean	SD
Employment equation:						
last month dropout	0.487	0.190	0.185	0.258	1.317	0.305
dropout in recent past	-0.688	0.176	-0.762	0.244	-0.039	0.290
job-aligned dropout in past & e[t-1]=0	-0.262	0.172	0.156	0.274	0.284	0.276
job-aligned dropout in past & e[t-1]=1	-0.356	0.167	-0.081	0.255	0.390	0.276
non-job-aligned dropout in past & e[t-1]=0	-0.198	0.174	-0.306	0.248	0.501	0.285
non-job-aligned dropout in past & e[t-1]=1	-0.134	0.179	-0.257	0.254	0.745	0.303
month of planned end or before	0.311	0.123	0.017	0.179	0.644	0.217
month 1 after planned end	0.112	0.093	0.061	0.158	0.465	0.170
month 2 after planned end	-0.260	0.089	-0.056	0.144	0.001	0.166
month 3 after planned end	-0.155	0.082	-0.131	0.135	-0.044	0.157
month 4 to 6 after planned end	-0.185	0.058	-0.223	0.098	-0.160	0.110
month 7 to 12 after planned end	-0.143	0.040	-0.034	0.072	-0.135	0.077
month 24 to 40 after planned end	0.073	0.045	0.135	0.145	0.070	0.086
e[t-1]	2.919	0.074	2.923	0.126	3.103	0.135
e[t-2]	-0.198	0.060	-0.237	0.108	-0.177	0.116
e[t-3]	0.008	0.059	-0.092	0.103	-0.077	0.109
e[t-4]	0.010	0.060	0.115	0.105	0.001	0.112
e[t-5]	0.001	0.062	-0.148	0.110	0.142	0.113
e[t-6]	-0.041	0.054	-0.001	0.097	0.018	0.102
$\sum_{j=7}^{12} e[t-j]$	-0.001	0.010	0.004	0.020	0.001	0.019
$\sum_{j=13}^{18} e[t-j]$	-0.061	0.010	-0.064	0.023	-0.057	0.018
$\sum_{j=19}^{24} e[t-j]$	-0.036	0.012	-0.037	0.030	0.003	0.021
$\sum_{j=38}^{25} e[t-j]$	-0.027	0.010	-0.063	0.026	-0.025	0.018
elap. mon. in current state & e[t-1]=1	0.003	0.011	-0.029	0.022	-0.000	0.019
elap. mon. in current state sq. & e[t-1]=1	0.001	0.000	0.002	0.001	0.000	0.001
elap. mon. in current state & e[t-1]=0	-0.038	0.009	-0.055	0.016	-0.010	0.016
elap. mon. in current state sq. & e[t-1]=0	0.001	0.000	0.002	0.000	-0.000	0.000

<continued on next page>

Results of MCMC Estimation (Flexible Specification)
<continued>

Variable	General		Retraining		Practical	
	Mean	SD	Mean	SD	Mean	SD
planned length in days/31	-0.010	0.006	0.015	0.006	-0.033	0.020
female	0.052	0.049	0.168	0.096	0.158	0.112
living in East Germany	-0.054	0.059	-0.377	0.141	-0.033	0.134
no vocational degree	0.009	0.066	-0.131	0.090	-0.240	0.127
no schooling degree	-0.099	0.127	0.237	0.208	-0.337	0.204
lower secondary (Hauptschule)	-0.067	0.057	-0.125	0.104	0.049	0.114
high school (Abitur)	-0.050	0.060	-0.202	0.122	0.011	0.152
25-29 years old	0.161	0.067	0.015	0.104	0.099	0.144
30-34 years old	0.072	0.060	-0.145	0.104	0.188	0.127
35-40 years old	-0.121	0.059	-0.169	0.177	-0.130	0.125
50-54 years old	-0.406	0.080	-0.012	0.339	-0.679	0.163
child under 10 * male	0.143	0.079	-0.084	0.153	0.255	0.158
child under 10 * female	-0.063	0.087	-0.312	0.179	-0.028	0.198
health problems	-0.109	0.087	-0.150	0.165	-0.085	0.159
unemployed before (last three years)	0.065	0.062	0.033	0.105	-0.334	0.136
days/31 unempl. assistance last 3 years	-0.024	0.007	-0.020	0.012	-0.030	0.012
training program before	-0.056	0.079	-0.294	0.191	-0.172	0.183
days/31 employed last 3 years	0.006	0.003	0.009	0.005	0.005	0.006
spring (second quarter)	0.023	0.029	0.021	0.058	0.072	0.055
fall (fourth quarter)	-0.109	0.030	-0.035	0.054	-0.210	0.056
winter (first quarter)	-0.059	0.032	-0.185	0.060	-0.049	0.060
year 2002	-0.080	0.035	-0.082	0.080	0.115	0.065
year 2003	0.028	0.044	-0.057	0.077	0.187	0.085
year 2004	0.058	0.058	-0.116	0.091	0.110	0.112
log last real wage	-1.101	0.651	-0.610	1.490	-0.576	1.238
log last real wage squared	0.058	0.035	0.030	0.080	0.031	0.067
last job in industry with seasonal work	-0.048	0.107	0.091	0.161	-0.022	0.208
urban region high unempl. in East	-0.116	0.086	0.250	0.240		
urban region, good conditions in West	-0.067	0.098	-0.148	0.154	0.571	0.259
non-urban region, good conditions in West	0.136	0.062	0.250	0.098	0.046	0.112

<continued on next page>

Results of MCMC Estimation (Flexible Specification) <continued>

Variable	General Mean	General SD	Retraining Mean	Retraining SD	Practical Mean	Practical SD
constant	4.020	3.064	1.850	6.916	1.456	5.800
Dropout equation:						
days/31 until planned end	-1.986	0.450	-0.118	0.186	-3.150	0.902
near to program begin	-0.074	0.095	-0.060	0.099	-0.209	0.131
no schooling degree	0.605	0.179	0.627	0.231	0.176	0.212
lower secondary (Hauptschule)	0.184	0.091	0.283	0.109	-0.139	0.153
sanction in last 3 years	0.500	0.258	-0.023	0.536	0.319	0.344
lack of motivation w.r.t. agency's activities	0.219	0.120	0.361	0.134	0.053	0.170
health problems	0.040	0.153	-0.018	0.205	-0.234	0.224
female	0.109	0.089	0.133	0.123	-0.362	0.153
living in East Germany	-0.099	0.079	-0.279	0.144	-0.533	0.164
25-30 years old	0.155	0.112	0.189	0.121	0.202	0.196
30-34 years old	0.142	0.098	-0.174	0.126	0.204	0.167
training program before	0.099	0.129	0.482	0.202	0.424	0.214
unemployed before (last three years)	0.191	0.115	0.138	0.133	0.186	0.168
last job as blue-collar worker	0.066	0.088	-0.259	0.119	0.182	0.145
last job in industry with seasonal work	-0.123	0.195	-0.089	0.211	0.209	0.248
living alone	0.143	0.094	0.453	0.121	0.044	0.135
child under 10 * female	0.121	0.148	0.195	0.207	0.057	0.265
child under 10 * male	0.025	0.141	0.720	0.191	-0.385	0.224
constant	-2.659	0.184	-2.974	0.183	-1.651	0.236
Individual level variances:						
individual level variance employ. equ.	0.530	0.066	0.741	0.143	0.718	0.147
individual level variance dropout equ.	0.662	0.212	0.553	0.161	0.619	0.200
individual level covariance	0.217	0.104	0.108	0.110	-0.148	0.151
share on individual level, employ. equ.	0.345	0.028	0.422	0.046	0.414	0.049
share on individual level, dropout equ.	0.389	0.073	0.350	0.061	0.373	0.072
correlation between equations	0.134	0.060	0.065	0.066	-0.085	0.084
correl. between random effects	0.363	0.147	0.167	0.168	-0.208	0.201

Appendix B

Algorithm for the MCMC Estimation

The following independent priors are set: the prior distributions of the coefficients $\eta_E = \beta_E$ and δ_E are given by independent normal priors with distribution $\mathcal{N}(b_{E,0}, B_{E,0})$. $\mathcal{N}(\bullet)$ denotes the normal distribution. Setting very large values for the variance $B_{E,0}$, I use extremely diffuse priors. The same is done for the coefficients of the β_D vector, the prior distributions are given by $\mathcal{N}(b_{D,0}, B_{D,0})$. The prior distribution of the random effects is $\mathcal{N}(0, \Sigma)$. The hyperparameter Σ^{-1} follows the prior distribution $\mathcal{W}^{-1}(H_0, h_0)$, where H_0 is the inverse scale matrix and h_0 denotes the degrees of freedom. $\mathcal{W}^{-1}$ denotes the inverse Wishart distribution. To use a diffuse prior I set a small h_0. For the diagonal elements of H_0 the individual level variances of a separate ML estimation of the two equations times h_0 are set and I set the off-diagonal elements to zero. The algorithm is presented in the following. Let $z_{it,E}$ and $z_{it,D}$ denote the whole set of covariates in the employment or dropout equation, respectively.

- Set starting values for the coefficient vectors η_E and β_D, the individual specific effects ($\alpha_{i,E}$, $\alpha_{i,D}$) and the variance covariance matrix of the individual specific effects Σ.
- Step 1a: Sample E^*_{it} from $\mathcal{N}(z_{it,E}\eta_E + \alpha_{i,E}, 1)$ with support $[0, \infty]$ if $E_{it} = 1$ and with support $[-\infty, 0]$ if $E_{it} = 0$ (if the employment equation is to be estimated). $\mathcal{N}(\bullet)$ denotes the normal distribution.
- Step 1b: Sample D^*_{it} from $\mathcal{N}(z_{it,D}\beta_D + \alpha_{i,D}, 1)$ with support $[0, \infty]$ if $D_{it} = 1$ and with support $[-\infty, 0]$ if $D_{it} = 0$ (if the dropout equation is to be estimated).
- Step 2: Sample $(\alpha_{i,E}, \alpha_{i,D})'$ from its bivariate normal conditional posterior distribution $\mathcal{N}(\mu, V_{\alpha_i})$, where $\mu = V_{\alpha_i} \cdot \begin{pmatrix} T_{i,E} & 0 \\ 0 & T_{i,D} \end{pmatrix} \cdot \begin{pmatrix} (\bar{E}^*_i - \bar{z_{i,E}}\eta_E) \\ (\bar{D}^*_i - \bar{z_{i,D}}\beta_D) \end{pmatrix}$ and $V_{\alpha_i} = \left(\Sigma^{-1} + \begin{pmatrix} T_{i,E} & 0 \\ 0 & T_{i,D} \end{pmatrix} \right)^{-1}$, a bar over a variable denotes its mean across time, $T_{i,E}$ the number of observations for person i for which the employment equation is to be estimated, and $T_{i,D}$ the number of observations for person i for which the dropout equation is to be estimated.

- Step 3a: Sample the η_E vector from its multivariate normal conditional posterior distribution $\mathcal{N}(M_E, V_E)$, where $M_E = V_E(B_{E,0}^{-1} b_{E,0} + \sum_{i=1}^{N} \sum_{t=1}^{T_{i,E}} z'_{it,E}(E^*_{it,E} - \alpha_{i,E}))$ and $V_E = (B_{E,0}^{-1} + \sum_{i=1}^{N} \sum_{t=1}^{T_{i,E}} z'_{it,E} z_{it,E})^{-1}$. N is the number of persons in the data using all person-periods for which the employment equation is to be estimated.

- Step 3b: Sample the β_D vector from its multivariate normal conditional posterior distribution $\mathcal{N}(M_D, V_D)$, where $M_D = V_D(B_{D,0}^{-1} b_{D,0} + \sum_{i=1}^{N} \sum_{t=1}^{T_{i,D}} x'_{D,it}(D^*_{D,it} - \alpha_{i,D}))$ and $V_D = (B_{D,0}^{-1} + \sum_{i=1}^{N} \sum_{t=1}^{T_{i,D}} z'_{it,D} z_{it,D})^{-1}$ using all person-periods for which the dropout equation is to be estimated.

- Sample Σ^{-1} from its conditional posterior distribution $\mathcal{W}^{-1}\left(\begin{pmatrix} \sum\limits_{i=1}^{N} \alpha_{i,E}^2 & \sum\limits_{i=1}^{N} \alpha_{i,E}\alpha_{i,D} \\ \sum\limits_{i=1}^{N} \alpha_{i,E}\alpha_{i,D} & \sum\limits_{i=1}^{N} \alpha_{i,D}^2 \end{pmatrix} + H_0,\ N + h_0\right)$. $\mathcal{W}^{-1}$ denotes the inverse Wishart distribution.

- Go to Step 1. Always use current values.

Chapter 5

The Heterogeneous Effects of Training Incidence and Duration on Labor Market Transitions

5.1 Introduction

There exists a huge literature which estimates the effects of government sponsored training programs on outcome variables such as earnings and employment.[1] The literature differs regarding the econometric methods used. In one strand of the literature, the application of matching methods presumes that sufficiently rich data are available to justify that there is no remaining selection on unobservables after controlling for observable variables. A different strand of the literature uses the timing of events to identify the treatment effects. Abbring and van den Berg (2003) (henceforth AVdB) show that single spell data for unemployment duration and time until start of treatment allow to identify the effect of treatment on exit rates from unemployment provided a mixed proportional hazard (MPH) model holds and treatment can not be anticipated. Furthermore, the effects of treatment on hazard rates in subsequent spells are identified. The model allows for the presence of permanent unobserved heterogeneity terms which enter the hazard rates in a multiplicative fashion and which are independent of the observed covariates.[2] These assumptions are crucial because AVdB show that in a more general setting without the MPH assumption, the effect of treatment start on exit rates from unemployment can not be identified. AVdB view the duration until treatment start and the duration until exit from unemployment as two competing risks which are linked through their dependence upon the unobserved permanent heterogeneity. Based on the MPH and the no-anticipation assumption, the timing of events allows to identify the treatment effect because in this competing risks setting the exit rates from unemployment after treatment starts can be contrasted with the model estimates for the exit rates in the case where treatment would not have been started. The latter are estimated through the competing risks part of the model (see Abbring and van den Berg (2004), page 15). Because the unobserved heterogeneity term is permanent, its spurious selection

[1]See e.g. Heckman, LaLonde and Smith (1999), Martin (2000), Kluve (2006), Card, Kluve, and Weber (2009) for surveys and Heckman, Ichimura, Smith, and Todd (1998), Sianesi (2004), Richardson and van den Berg (2008), Lechner, Miquel, and Wunsch (2009), Bergemann, Fitzenberger, Speckesser (2009), Biewen, Fitzenberger, Osikominu, and Waller (2007), Osikominu (2008) and Stephan (2009) as exemplary empirical evaluation studies for various countries.

[2]AVdB also discuss identification based on multiple spell data under weaker functional form assumptions on the hazard rates. However, the approach is only applicable for subsets of observations with multiple unemployment spells and the crucial assumption to account for unobserved heterogeneity is that the latter enters the hazard rates multiplicatively in the same way for both spells. Abbring and van den Berg (2004) discuss the link between treatment effects estimation using duration models based on the timing-of-events approach on the one hand and cross-sectional binary treatment models as well as linear panel data models with individual fixed effects on the other hand.

(or sorting) effects can be distinguished from the effects of treatment starts which show random variation in time across individuals.

In a different strand of the timing-of-events approach, Sianesi (2004) suggests to estimate the effects of treatment versus waiting by using a sequential dynamic matching approach to account for the changing selection of treated individuals who are treated at a certain elapsed duration of unemployment. Treated individuals are matched to individuals who are still unemployed after the same elapsed duration of unemployment and who have not yet started treatment. The latter includes individuals who will start treatment later during the course of their unemployment spell.[3]

As a common feature, all the cited empirical studies in footnote 1 estimate the effects of the incidence of training - typically measured by the time of the start of participation in a training program - on future outcome variables. The literature typically ignores the actual duration of treatment, which is likely to be endogenous.[4] Endogeneity works through dropout from treatment before the planned end of the treatment (see also Kluve, Schneider, Uhlendorff, and Zhao, 2007, and Waller, 2009, on the endogeneity of dropouts). One form of endogeneity of dropouts reflects the fact that individuals may drop out from a program because they exit from unemployment. At the same time, it may be the case that treatment so far has affected the chances to find a new job. This argument suggests that dropouts are a positive selection compared to individuals who complete the treatment until the planned end. A second form of endogeneity of dropouts results if participants are not able to participate regularly in the programs (e.g. because of lack of endurance), if the courses are too difficult for them, or if they simply find out that the program does not match their interests or needs. These reasons for dropouts suggest a negative selection of dropouts. For similar reasons, individuals who participate in the program beyond the planned duration (e.g. due to not reaching the goals of the programs in time) are likely to involve a negative selection of program participants compared to participants who finish the program as planned.

Furthermore, it is to be expected that training programs involve "mechanical" lock-in effects, i.e. between the start of the program and the planned end of the program

[3]The approach of Adda, Costa Dias, Meghir and Sianesi (2007) differs from the aforementioned strands of the literature. The authors build and estimate a structural dynamic model of labor supply incorporating labor market policies.

[4]The treatment AVdB consider in their example does not involve a duration. It would be conceptually straightforward to add a third duration variable accounting for the duration of treatment.

exit rates from unemployment are considerably lower for participants compared to similar non-participants. The size of the lock-in effect may change over the course of treatment and the time until the planned end of the program may have a negative effect on exits from unemployment because job finding efforts will increase the sooner the program is going to end. In a mechanical sense, the moment of the dropout marks the end of the lock-in effect. The recent literature typically accounts for the lock-in effect by estimating the causal program effect from the start of the treatment onwards, which includes the lock-in effect over the course of treatment.

One may consider to take different training duration as different doses of treatment where the treatment effect differs by the length of the treatment such that training programs with different lengths may be considered as multiple distinct treatments (Imbens, 2000). The endogenous nature of actual durations of program participation and the asymmetry induced by lock-in effects suggest that it may not be sufficient to account for different lengths of the programs or to estimate a dose response function both based on the conditional independence assumption.[5] To account for the selection of dropouts, it is important to allow for dropouts who immediately find a job after dropout and other dropouts. The former group is likely to involve a positive selection of dropouts while the latter is likely to involve a negative selection of dropouts. By specifying separate equations for employment transitions and for exits from treatment, our empirical approach accounts for both types of selection.

In the meta-analysis of active labor market policy conducted by Card, Kluve, and Weber (2009), Germany is the country with the largest number of evaluation studies for training. Most of the recent studies on long-term training programs in Germany have used propensity score matching to estimate the average effect of treatment on the treated, see among others Bergemann, Fitzenberger, and Speckesser (2009), Biewen, Fitzenberger, Osikominu, and Waller (2007), Lechner and Wunsch (2006, 2007), Schneider and Uhlendorff (2006), and Stephan (2009). These studies (except the first one in the above list) are based on German administrative data and they condition on a rich set of observed variables including detailed information on the

[5] Kluve, Schneider, Uhlendorff, and Zhao (2007) account for dropouts and discuss the possibility that early dropouts are a positive or a negative selection of program participants. They provide IV estimates of outcome regressions on actual duration where the endogenous actual duration is instrumented by the exogenous (with and without conditioning on other covariates) planned duration. Hausman test results indicate that the OLS and the IV estimates do not differ significantly. This result is not very informative for two reasons. First, both the OLS and the IV estimates are very imprecisely estimated. Second, if both positive and negative selection effects for dropouts are present then the two types of biases induced may be offsetting each other.

employment history. They do not account explicitly for unobserved heterogeneity.[6] Based on AVdB's approach, Osikominu (2008) estimates the effects on the exits from unemployment and the stability of subsequent employment accounting for selection on unobservables. Regarding long-term training, the study finds no reduction in unemployment duration but strong positive effects on the employment duration.

Our paper implements the timing-of-events approach in discrete time and estimates training effects both on the exit rate from unemployment and the rate at which individuals keep employment. We estimate the effects of long-term training based on rich administrative labor market data in Germany. We account for endogenous dropout and selection on unobservables. We specify a two-equation model for discrete transitions between employment and non-employment as well as for entry into and exit from training while being non-employed. The model is a bivariate random effects probit model which accounts for two dependent random effects and which is estimated by Markov Chain Monte Carlo (MCMC) techniques.[7] The analysis is based on an inflow sample into unemployment and we account for state dependence and duration dependence in a very flexible way in order to avoid strict functional form restrictions. The model is based on the timing-of-events assumption similar to AVdB where training can impact upon the two labor market transition rates only after the beginning of training. Also the impact of training is modeled in a very flexible way in order to avoid strict functional form restrictions. The model is identified in an analogous way as in the timing-of-events approach by AVdB based on the non-anticipation assumption and the assumption that the random effects are uncorrelated with the observed covariates.

There are the following innovative methodological aspects in our paper. First, we estimate the effect on employment both of the incidence and the duration of training accounting for endogenous dropout. Second, estimating a discrete time model for labor market transition we account for the full observed observation vector for each individual over time. This is in contrast to almost all of the papers using continuous time duration models which restrict the attention to a small, fixed maximum number of spells analyzed for a given individual. Third, though using a parametric bivariate random effects probit model for employment and training, we specify the model in a very flexible way in order to account for state dependence, duration dependence,

[6] Some of these studies (the second and the last one in the above list) apply the treatment versus waiting approach suggested by Sianesi (2004).

[7] See Chib (2001) for a survey on MCMC methods and Buchinsky, Fougère, Kramarz, and Tchernis (2005) for a recent application in modelling discrete time transition models.

interaction effects, and heterogeneity of treatment effects. Our large sample size allows us to integrate such flexibility into our model. Our model estimates are semi-parametric in nature and they allow to estimate a variety of interesting treatment effects. The model replicates well the observed time path of employment for the treated individuals. Fourth, using an inflow sample into unemployment, the probit specifications are chosen to account for differences in the employment history before entering the unemployment spell. This way, we make use of the richness of the administrative data at hand to account for the selection of treated individuals. Fifth, using Bayesian MCMC techniques allows for a numerically very robust estimation of our flexible model specification. In addition to the estimation of the posterior distribution of the parameter estimates, the MCMC techniques also provide predictions of the individual random effects in both equations. We suggest how to use these estimates in a simulation approach for the estimation of the posterior distribution of various treatment effects of interest, such as the average effect of treatment on the treated and the effect of different given planned program lengths. Furthermore, the MCMC estimates allow to assess explicitly the selectivity of the treated and the nontreated individuals based on the predicted random effects. Sixth, by accumulating the discrete time variation in the treatment and outcome variables, i.e. using the full information of the discrete data, the estimation approach does not have to assume the same model specification across spells.

To put our paper into perspective in comparison to the AVdB approach, it has to be mentioned that our estimation approach has also a number of disadvantages. First, compared to continuous time duration models, we have to aggregate the time dimension into a small number of discrete time points. In order to limit our analysis to a manageable number of observations (persons$\times$time periods), we have to restrict our data to a quarterly frequency. These quarterly data do not allow to identify quick successions of events which allow to identify the sign of the effect of treatment start under less stringent assumptions in the AVdB approach, see e.g. Abbring and van den Berg (2004, page 17). Second, we had to aggregate the timing of events within a quarter into a quarterly employment dummy and a quarterly treatment dummy. This could be problematic for short treatment durations. Thus, we only analyze longer training programs which are planned to last at least a couple of months. Third, when an individual, who is still participating in training at the beginning of a quarter, exits from non-employment to employment during the same quarter, this implies that treatment must have ended during this quarter, i.e. our estimation approach can only estimate the effects of dropout for the subsequent quarter. This is another reason why a quick succession of events can not be analyzed with our approach. Fourth, even though we use data on individuals with multiple transitions, i.e. the discrete time equivalent of multiple spell data, our estimation approach assumes independence between the covariates and the individual random effects. This is in contrast to mutiple spell continuous time duration models based on the MPH assumption which allow to difference out the individual random effect. However, such an approach could only be implemented for individuals with multiple transitions. Furthermore, to estimate the treatment effect on the probability of remaining employed one would have to model employment before the inflow into the unemployment spell which defines our sample. Thus, one would have to use a much larger sample than we use and for this sample one would most likely have to address a difficult left censoring problem. Finally, such a model requires the same specification of the hazard rate across spells, whereas our approach allows us to accumulate information on model outcome variables over time. For these reasons, we prefer our approach. Fifth, we account for state dependence in a flexible way but we do not model completely separate discrete choice models for both non-employment and employment. We check our model fit in order to investigate whether a misspecification problem exists. These five caveats point to limitations of our estimation framework. For the reasons discussed above, we think it is nevertheless well suited for the problem analyzed in this paper.

Our estimation results imply positive effects of long training on exits from non-employment soon after leaving the program for all cases considered. Ten quarters after program start, the effect of treatment on unconditional employment rates for the treated individuals lies between 12 and 21 percentage points (ppoints). These effects are more positive than the results reported in the literature, especially for East Germany. This is consistent with our model estimates implying a strong negative selection of treated individuals.

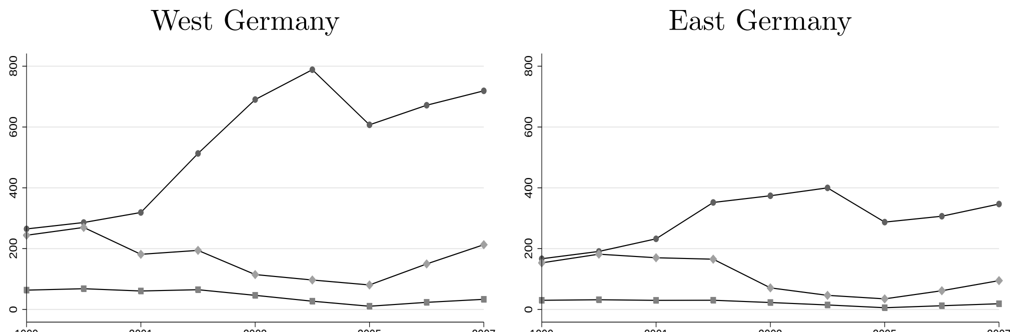
The remainder of this paper is structured as follows: section 2 describes the data and the training program. Section 3 discusses our evaluation method and MCMC estimation of the model. Section 4 presents the results and section 5 concludes. The appendix includes further details on the data, the implementation of the estimation approach, and detailed estimation results.

5.2 Institutional Background and Data

5.2.1 Training in Germany

Training schemes have traditionally dominated active labor market policy in Germany. The legislation distinguishes three main types of training, further training (*Berufliche Weiterbildung*), retraining (*Berufliche Weiterbildung mit Abschluss in einem anerkannten Ausbildungsberuf*), and short-term training (*Trainingsmaßnahmen und Maßnahmen der Eignungsfeststellung*). Figure 5.1 shows the evolution of entries into the three different training programs in West and East Germany during the period 1999 to 2007. Until 2000, enrolment into further training (henceforth also referred to as long-term training) was around 260 thousand in West Germany and 170 thousand in East Germany. A policy reorientation favoring programs supposed to activate the unemployed in the short run led to a decline in further training and retraining and a sharp increase of short-term training. In 2004, participation in further training was about 100 thousand in West Germany and about 50 thousand in East Germany. The corresponding figures for short-term training were 800 thousand and 400 thousand, respectively, up from around 200 thousand in 1999. After a low point in 2005, participation recovered somewhat in 2006 and 2007.

Figure 5.1: Entries into Training Programs in West and East Germany (in 1000)



Source: BA (2001, 2006, 2007, 2008); own calculations.

The main goal of active labor market policy in Germany is to reintegrate unemployed individuals into employment. In this study we focus on further training programs. They are used to adjust the skills of the unemployed to changing requirements of the labor market and possibly to changed individual conditions of employability (due to health problems for example). Further training courses typically last several months to one year and are usually conducted as full-time programs. Teaching takes place in class rooms or on the job in training firms. The course curriculum may also include internships. Typical examples of further training schemes are courses on IT based accounting or on customer orientation and sales training. Similar to the much longer retraining schemes, that lead to a complete new degree within the German apprenticeship system, further training programs aim at improving the human capital and productivity of the participant. Short-term training, in contrast, primarily aims at improving job search and lasts typically about four weeks.

In order to become eligible for training, job seekers have to register personally at the local employment agency. This involves a counseling interview with a caseworker. In principle, they have in addition to fulfill a minimum work requirement and be entitled to unemployment benefits. However, there are exceptions to this rule. The most important criterion is that the training scheme has to be considered necessary by the caseworker for the unemployed to find a new job. Participation in training can occur at any time during an unemployment spell.

Before 2003, training measures were assigned by the caseworker. This was often done in agreement with the job seeker, considering his or her willingness to receive training and to work in a specific field. The final decision was subject to the discretion of the caseworker. Assignment into programs was to a large extent driven by the supply of courses that were booked in advance for a year by the employment agencies from training providers. Assignments to training often occurred at very short notice in order to fill course capacities (Schneider et al., 2006).

In 2003, the assignment procedure changed to a system where the job seeker receives a training voucher from the caseworker valid between one and three months. The voucher specifies the maximal length, the content and the objective of the eligible training program. The job seeker then chooses by himself a suitable course from a pool of certified training providers. The 2003 reform meant to make the allocation process more targeted and selective. However, potential participants were uncertain about the actual starting date because it turned out that training providers tended to collect vouchers until a critical number of participants was reached or they shortly canceled scheduled courses if there were too few participants (Kühnlein and Klein, 2003, Schneider et al., 2006). Moreover, in the first months of 2003, programs that were assigned under the old system still started. 93% of the programs in our analysis sample start before the reform. An additional 2% starts in the first quarter of 2003, thus about 5% of the programs fall in the time when vouchers were used.

During training most participants receive a subsistence allowance of the same amount as the unemployment compensation they would receive otherwise. Participants not eligible for subsistence allowance may receive similar payments from the European Social Fund. In addition, travel and child-care costs may be covered by the employment agency.

Once a particular program or a training voucher has been assigned, participation is mandatory. Non-compliance may be sanctioned with a temporary suspension of unemployment compensation. The planned duration of the further training programs considered in this paper is eight months on average. However, not all participants who start a program complete it. In fact, according to the study in Waller (2009), one out of five participants who have started a program and attended it for at least one week drop out before having reached 80% of the planned duration. About half of the dropouts start employment soon after quitting a program. In many cases this behavior is encouraged by the employment agency because in general employment has priority over participation in active labor market programs. Exceptions from this rule are possible if completing the program is deemed necessary for a stable placement. Those dropping out for other reasons are typically not sanctioned. As opposed to dropouts, it also happens in some cases that participation in training is prolonged. Due to dropout and possible prolongment of participation the actual duration of training is endogenously determined.

5.2.2 Constructing a Panel Data Set

For the empirical analysis, we construct a panel data set from a rich administrative database, the Integrated Employment Biographies Sample (IEBS). The IEBS is a 2.2% random sample from a merged data file containing individual data records collected in four different administrative processes: the IAB Employment History (*Beschäftigten-Historik*), the IAB Benefit Recipient History (*Leistungsempfänger-Historik*), the Data on Job Search Originating from the Applicants Pool Database (*Bewerberangebot*), and the Participants-in-Measures Data (*Maßnahme-Teilnehmer-Gesamtdatenbank*).[8] The data contain detailed daily information on employment subject to social security contributions, receipt of transfer payments during unem-

[8]For further information on the data see Appendix A.

ployment, job search, and participation in different active labor market programs.

We consider an inflow sample into unemployment consisting of individuals who became unemployed between the first of July 1999 and the end of December 2000, after having been continuously employed for at least 125 days. Entering unemployment is defined as the transition from non-subsidized employment to non-employment plus subsequently (not necessarily immediately) some contact with the employment agency, either through benefit receipt, program participation, or a job search spell. In order to exclude individuals eligible for specific labor market programs targeted to youths and individuals eligible for early retirement schemes, we only consider persons aged between 25 and 53 years at the beginning of their unemployment spell.

We discretize the spell information in the original data by calendar quarters. We follow a person in the sample from the quarter of his/her first inflow into unemployment over the next 16 quarters or until the end of 2004, whichever occurs first. For 76% of the individuals in the sample we observe the full sequence of 17 quarters. The sequences of the remaining individuals are shorter either because we observe less than 17 quarters from their inflow until the end of 2004, or because we censor the time path for training participants in the quarter in which they enter a long-term active labor market program other than training. We ignore participation in short-term training and do not censor employment sequences in this case.

We distinguish the two outcome states non-subsidized employment (henceforth denoted as employment) and non-employment as alternative states. We aggregate the employment information measured at a daily level into quarters as follows. First, for short gaps of a length up to 45 days between sequences of longer employment or non-employment spells we extend the longer spells through the gap. Second, we map the start of non-employment and employment spells to the quarterly employment dummy in the following way. If a transition to non-employment occurs during a calendar quarter, the employment dummy is set to zero during this quarter. It continues to equal zero in the following quarter if the elapsed duration of non-employment at the end of the quarter exceeds 90 days. From the third quarter of non-employment onwards, the employment dummy is set to zero if the share of days in non-employment exceeds one half. Third, we take care of not dropping short employment spells by defining the individual to be employed in the later quarter.

Participation in further training is coded as follows. We construct a dummy variable that equals one in the quarter in which the job seeker starts a training program and attends it for at least 27 days. In order to model the duration of the training program we apply the same rules as for the employment dummy above to the qualification dummy. Because not only the start of a program but also the program status in each following quarter is used for the estimation, it is important to use reliable information on the realized program duration. We correct the reported end dates of training programs using the correction procedures proposed in Waller (2008). Participation can already occur in the first quarter we observe for an individual.

The definition of the quarterly employment and training dummy variables mimics the timing of events. When a person starts a training program in one quarter, he is also coded to be non-employed in that quarter. While being on the program, a participants remains non-employed. When a program participant exits into employment in one quarter, even though he has been in the program at the beginning of this quarter, the training dummy changes to zero in that quarter. Consequently, our empirical analysis imposes a lag in the effect of training, such that training in one quarter is only allowed to have a causal effect on employment in future quarters.

The panel data set for the analysis is complemented by adding personal, occupational and regional information. Some of the covariates are updated at the beginning of each quarter. The estimations are carried out separately for males and females and West and East Germany.

5.2.3 Descriptive Analysis

Table 5.1 gives an overview of the four samples and their basic characteristics. On average we observe 13 to 15 quarters per person, with the number of non-employment quarters ranging from eight to ten. This corresponds to 1.5 to 1.9 unemployment spells and about one employment spell on average per person. One in ten to one in five persons participate in training throughout the observation period with participation rates being higher in East Germany and among females.

Figure 5.2 illustrates the evolution of the employment and training rates from the quarter of inflow into unemployment onwards. In the calendar quarter of the inflow, all individuals are defined as non-employed. The employment rates subsequently recover, but those of females remain at a slightly lower level than those of males. While participation rates barely reach five percent in West Germany, they peak at about eight to nine percent in East Germany.

Table 5.1: Descriptive Statistics

	Male, West	Female, West	Male, East	Female, East
Individuals	16,317	12,328	8,737	4,869
Quarters per Person	14.7	15.3	13.2	13.5
Quarters Employed p. P.	6.0	5.2	5.1	4.0
Quarters Unemployed p. P.	8.70	10.1	8.2	9.5
Quarters in Training p. P.	0.27	0.31	0.45	0.56
Employment Spells p. P.	1.21	0.85	1.06	0.74
Unemployment Spells p. P.	1.89	1.53	1.78	1.48
Training Spells p. P.	0.11	0.12	0.16	0.19

Figure 5.2: Employment and Participation Rates over Time

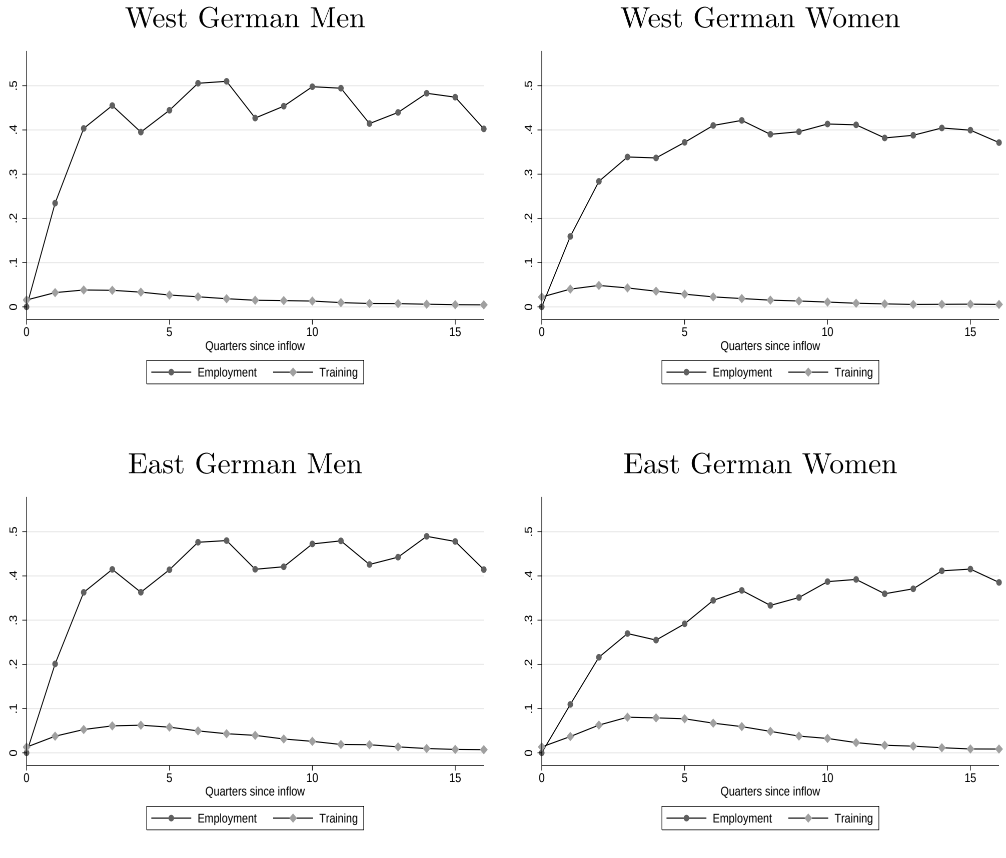
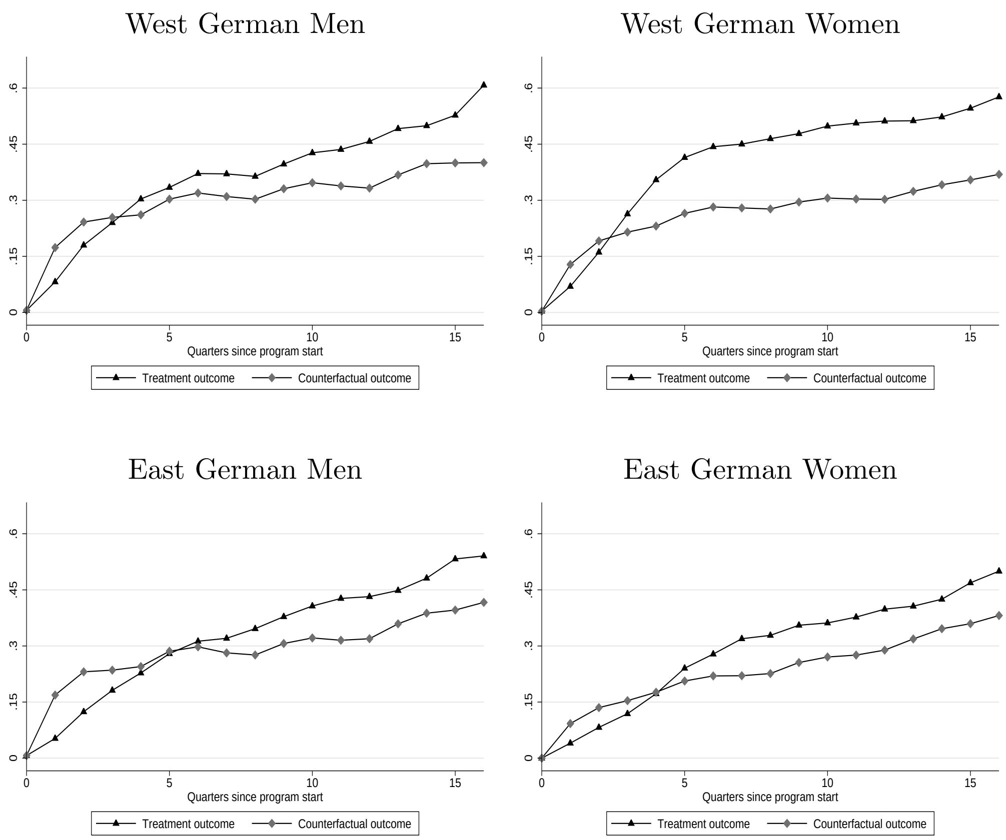


Figure 5.3 gives a first impression of the likely order of magnitude of the treatment effects. It shows the actual employment rates and estimates of the counterfactual associated with starting a training program in a given quarter versus waiting for the treated individuals where treated and matched controls are merely aligned in time. Treatment status is a time-varying variable. This means that training participants who enrol later are counted as controls for those who enrol in an earlier quarter.

The matching is performed with respect to the calendar quarter of the first inflow and the elapsed unemployment duration in the current unemployment spell. No adjustments are made for other potential sources of selection bias. West German females show the largest employment differences five to ten quarters after program start, which amount to more than 15 ppoints. The initial lock-in periods characterized by negative employment effects are substantially longer in East Germany than in West Germany.

Figure 5.3: Raw Treatment and Nontreatment Employment Rates



Note: Raw estimates of the treatment effect on the treated, where treated and controls are aligned in the time dimension only. In particular, treated and nontreated individuals are matched on the calendar quarter of their first inflow and elapsed unemployment duration in the current spell. No adjustments are made for other potential sources of selection bias.

5.3 Evaluation Framework

5.3.1 Estimation Approach

We estimate the effect of participating in training, measured by the sequence of quarterly training dummy variables Q_{it}, on the employment dummy $E_{it'}$ in subsequent quarters ($t' > t$). We model the employment and the training decision as a two equation system with possibly dependent individual specific effects. This means that we allow for selection into and out of training based on unobservables. As both dependent variables are binary we specify a random effects probit model for each. Following Chib and Hamilton (2002) and Chib and Jacobi (2007), we estimate the model using Bayesian MCMC techniques. The treatment effects are estimated using simulations for both outcomes based on the MCMC iterations. Chib and coauthors analyze the binary treatment case with continuous outcomes and allow separate outcome equations. We estimate the effect of both treatment incidence and the duration of treatment on a discrete outcome variable (employment) and we allow state and duration dependence in the outcome variable.

Estimation of discrete choice-models for labor market transitions can be viewed as a discrete time version of the timing-of-events approach by AVdB which uses a continuous time duration model with unobserved heterogeneity, where time until treatment start and unemployment duration constitute two competing risks. Note, however, that our approach also models the length of the training program. The goal of the timing-of-events approach by AVdB is to estimate the causal treatment effect on the hazard to leave unemployment. Identification of the causal effect of entering a program relies on the conditional randomness of program starts and a no-anticipation condition as well as functional form assumptions involving e.g. a mixed proportional hazard model and functional form assumptions qualitatively similar to the ones used here. Similar to AVdB, our approach relies on a selection on unobservables strategy. Our estimates allow for heterogeneity of treatment effects and we estimate both the effect of training incidence and duration. Identification of the treatment effects in our model is implied in an analogous way by the standard assumptions of the timing-of-events approach by AVdB, i.e. the non-anticipation assumption and the assumption that the random effects are uncorrelated with the observed covariates. We rule out that anticipation of participation in the future affects current or future employment. These identifying assumptions are plausible in the present context as enrolment into training largely depends on short-term indicators (cf. section 5.2.1).

Now, we describe the model more formally. Consider first the employment equation. In order to model the employment dynamics we introduce employment lags up to the order of 15 (i.e. $E_{i(t-1)}$, $E_{i(t-2)}$, $\ldots$, $E_{i(t-15)}$, where i indexes individuals and t quarters) as explaining variables of current employment status. A lagged variable only kicks in if the inflow into unemployment has not been too recent for the corresponding lag to be available, i.e. the jth lag kicks in if $t - j \geq t_0$, where t_0 denotes the quarter of the inflow into unemployment. This way we account for the entire employment history since the inflow into unemployment. Furthermore, we include a vector of observed characteristics, $\boldsymbol{x}_{it,E}$, in the employment equation. In particular, we use information on schooling and occupational qualification, age, occupation and salary in the previous employment, number of days employed in the last three years before the inflow into unemployment, health, children, labor market characteristics of the residential municipality, season and year. In addition, we control in a flexible way for the elapsed number of quarters an individual is in the panel, t, and the elapsed duration in the current employment or non-employment spell, denoted $\tau_{it,E}$.

Following the timing-of-events assumption, we assume that participation in training in a given quarter affects in a causal way the employment probability only in subsequent quarters. Thus, the employment equation includes a flexible specification involving lagged information of training status. A dummy variable $Q_{i(t-1)}$ indicates if the individual attended a training program in the previous quarter. If this dummy takes a one, lagged training information is depicted by a dummy if the individual attended training from two quarters ago onwards $[Q_{(t-1)} = 1] \times [Q_{(t-2)} = 1] \times [Q_{(t-3)} = 0]$, from three quarters ago onwards $[Q_{(t-1)} = 1] \times [Q_{(t-3)} = 1] \times [Q_{(t-4)} = 0]$, from four quarters ago onwards $[Q_{(t-1)} = 1] \times [Q_{(t-4)} = 1] \times [Q_{(t-5)} = 0]$ and so on. If the dummy on participation in training in the last quarter $Q_{i(t-1)}$ takes a zero, we add a dummy D_{it} indicating whether an individual has ever participated in training since the inflow quarter. For trainees who have already exited the program, thus individuals with $Q_{i(t-1)} = 0$ and $D_{it} = 1$, we account for their training history by adding $\tau_{it,Q}$, which in this case indicates the completed duration, as well as a polynomial of tsb (time since begin), indicating how many quarters have passed since the quarter the former trainee started training. An interaction of $\tau_{it,Q}$ and tsb is also added. To allow for effect heterogeneity, the variables reflecting training history are interacted with other explanatory variables in the employment equation and the first lag of employment status $E_{i(t-1)}$. This way we can distinguish between the effect of training on entering employment and on remaining employed.

Specifically, the employment dummy E_{it} is modelled by:

$$
\begin{aligned}
(5.1) \qquad E_{it} = \mathbb{1}[&\boldsymbol{g}\left(E_{i(t-1)}, \ldots, E_{i(t-15)}, \boldsymbol{x}_{it,E}, \tau_{it,E}, t\right)\boldsymbol{\beta}_E + \gamma_{0E} Q_{i(t-1)} + \\
&D_{it}(1 - Q_{i(t-1)})\boldsymbol{m}\left(E_{i(t-1)}, \boldsymbol{x}_{it,E}, \tau_{i(t-1),Q}, tsb\right)\boldsymbol{\delta}_E \\
&Q_{i(t-1)}\boldsymbol{h}\left(Q_{i(t-2)}, \ldots, Q_{i(t-16)}, \boldsymbol{x}_{it,E}, \tau_{i(t-1),Q}\right)\boldsymbol{\gamma}_{1E} + \alpha_{i,E} + \epsilon_{it,E} > 0]
\end{aligned}
$$

where $\mathbb{1}[.]$ is the indicator function, $\boldsymbol{g}(.)$, $\boldsymbol{h}(.)$, and $\boldsymbol{m}(.)$ denote vector-valued functions, and $\boldsymbol{\beta}_E$, $\boldsymbol{\gamma}_{1E}$, $\boldsymbol{\delta}_E$ denote conformable column-vectors of coefficients. These functions may involve various interaction effects. $\alpha_{i,E}$ is the individual specific effect and $\epsilon_{it,E}$ the idiosyncratic error term. The treatment effect of training on employment is captured through the coefficients γ_{0E}, $\boldsymbol{\gamma}_{1E}$, and $\boldsymbol{\delta}_E$. It is allowed to differ by the duration of training, the time since training began, and a number of covariates. Specification (5.1) allows the information determining the employment probability to accumulate endogenously over time. Therefore, the effects of lagged employment are allowed to change by allowing for interactions with time t and other covariates. We account for duration and state dependence in a flexible way. The exact specification of the functions $g(.)$, $h(.)$, and $m(.)$ is guided by whether some coefficients are significant.[9]

Consider next the participation equation modeling the transition into and out of training. It is estimated simultaneously with the employment equation if the individual is not employed in the respective quarter and has not yet left a training program. Since participation can only occur during non-employment the two equation system reduces to a single equation for observations for which the employment status, E_{it}, is equal to one. Then, the treatment equation is switched off. This triangular structure provides additional identifying restrictions. We do not consider reentry into training after the completion of a first training program because this only occurs very rarely in our data and we consider this to be a different treatment.

The training equation includes a vector of observed regressors, $\boldsymbol{x}_{it,Q}$. In particular,

[9]We use the posterior variances and covariances to undertake the analogy of classical significance tests.

we include variables driving the decision to enter and to stay in a program. The vector comprises a dummy indicating whether the individual was enrolled in training in the previous quarter, $Q_{i(t-1)}$, a variable for the elapsed quarters in the program $\tau_{it,Q}$, a polynomial of the time until the planned end (in case enough planned duration is left to allow for another quarter in the program) and a dummy if the planned end is missing in the data. These variables are equal to zero if the individual was not yet enrolled in training ($Q_{i(t-1)} = 0$). Furthermore, the vector of independent variables includes variables summarizing the employment history since the inflow quarter, dummy variables indicating whether the current quarter is the inflow quarter, as well as whether a repeated transition from employment to non-employment has occurred, and a polynomial of the elapsed unemployment duration in days. Finally, information on age, schooling, vocational training, last job, number of days in employment in the last three years before the inflow, health, children and entitlement to unemployment compensation, season, and year is incorporated.

Specifically, the participation dummy Q_{it} is modelled by:

$$Q_{it} = \mathbb{1}[\boldsymbol{p}\left(Q_{i(t-1)}, \boldsymbol{x}_{it,Q}, \tau_{it,Q}\right)\boldsymbol{\beta}_Q + \alpha_{i,Q} + \epsilon_{it,Q} > 0] \tag{5.2}$$

where $\boldsymbol{p}(\centerdot)$ denotes a vector-valued function, $\boldsymbol{\beta}_Q$ the conformable coefficient vector, $\alpha_{i,Q}$ the individual specific effect, and $\epsilon_{it,Q}$ the idiosyncratic error term.

The two individual specific random effects, $\alpha_{(i,E)}$ and $\alpha_{(i,Q)}$, follow a *joint* normal distribution, $(\alpha_{i,E},\, \alpha_{i,Q})' \sim \mathcal{N}(\mathbf{0}, \Sigma)$. The error terms $\epsilon_{it,E}$ and $\epsilon_{it,Q}$ are independent standard normals. Thus, the model includes two individual specific effects which are allowed to be correlated. Let $\boldsymbol{z}_{it,y}$, $y \in \{E, Q\}$, denote the entire vector of covariates including lagged endogenous variables and interaction terms in the employment and qualification equation, respectively, $\boldsymbol{\eta}'_E = (\boldsymbol{\beta}_E, \gamma_{0E}, \boldsymbol{\gamma}_{1E}, \boldsymbol{\delta}_E)'$, $\boldsymbol{\eta}_Q = \boldsymbol{\beta}_Q$. T_i is the number of quarters individual i is in the panel. Then the likelihood contribution of individual i is as follows:

$$L_i = \int \prod_{t=1}^{T_i} f(E_{it}|\boldsymbol{z}_{it,E}, \alpha_{i,E}; \boldsymbol{\eta}_E) \cdot f(Q_{it}|\boldsymbol{z}_{it,Q}, \alpha_{i,Q}; \boldsymbol{\eta}_Q)^{C_{it}} dG(\alpha_{i,E}, \alpha_{i,Q}) \tag{5.3}$$

where $f(y_{it}) = \Phi(\boldsymbol{z}_{it,y}\boldsymbol{\eta}_y + \alpha_{i,y})^{y_{it}} \cdot (1 - \Phi(\boldsymbol{z}_{it,y}\boldsymbol{\eta}_y + \alpha_{i,y}))^{(1-y_{it})}$, $y \in \{E, Q\}$, $\Phi(\centerdot)$ denotes the standard normal cumulative distribution function, and C_{it} is a dummy equal to one if the individual is non-employed and has not yet completed a training program. As the individual specific effects are not directly observed one would have to integrate them out, as suggested in equation (5.3), in order to estimate the model by maximum likelihood. In this paper, however, we follow a different approach, which we describe in the following section. This approach avoids maximization of the likelihood function and it has the important advantage that it provides Bayesian estimates of the posterior distribution of all parameters of the model including the random effects. We exploit this feature to estimate the posterior distribution of different treatment effects of interest using a simulation approach.

5.3.2 MCMC Estimation of a Random Effects Probit Model

We estimate the model introduced in the previous section using Bayesian MCMC techniques (see Chib, 2001, for an overview on MCMC techniques). The values of the parameters along the MCMC iterations allow to estimate the posterior distribution of the parameters and of other model parameters of interest. From a classical perspective, the mean of the posterior distribution converges to the maximum of the likelihood function and the variance of the posterior distribution converges to the asymptotic variance of an ML estimation. Thus, the standard deviation of the draws may be interpreted as standard errors from the classical perspective (Train, 2003). To obtain a sample from the posterior distribution we use a Gibbs sampler, which works by forming blocks of the model parameters and then drawing in turn from the conditional distributions of the blocks of parameters. The resulting sequence is a Markov Chain and after convergence the draws are samples from the desired posterior distribution. The key idea for the estimation of probit models is to estimate the latent variables as one step of the simulation (Albert and Chib, 1993). A similar strategy is used for the estimation of the random effects in a best prediction sense (Zeger and Karim, 1991).[10] Odejar (2002) proposes a Gibbs sampler for a model sharing important features with the one estimated in this paper. Buchinsky et al. (2005) and Horny et al. (2008) apply similar models to study employment and wage mobility but they do not make use of the predictions of the random effects.

Details of the algorithm are given in Appendix B. Conjugate but very diffuse priors are used. The results reported below are based on running the algorithm for 50,000 iterations. We monitor convergence by comparing the means at different stages of

[10]The means of the posterior distribution of the individual specific random effects estimate the expected values of the random effects of one individual given the data and the prior distribution of parameters.

the chains. We discarded the first 5,000 iterations (the burn-in phase). Thus the results are based on 45,000 draws. We implement the Gibbs sampler in Stata.

5.3.3 Estimation of the Treatment Effects of Interest

The raw coefficient estimates are difficult to interpret because of the complex dynamic structure of the model involving many interaction effects. Therefore, we analyze directly the posterior distribution of the treatment parameters of interest. We consider the following average effects of treatment on the treated (ATT):

Classical ATT. This is the ATT of training versus non-participation during the observation period.

Training versus Waiting. At any given quarter, those enrolling into training during this quarter are counted as treated whereas those not yet enrolling are assigned to the control group. The latter may potentially participate in a later quarter. This effect mimics the treatment parameter suggested by Sianesi (2004) and estimated in several subsequent papers for European training programs using propensity score matching (see for example Biewen et al., 2007).

Effect of a Given Planned Program Duration. The ATT is estimated for different given planned program durations, allowing for the realized program length to be endogenous. In particular, we compare the effect of attending a program with a planned length of one, three, and four quarters, respectively, to attending a program with a planned duration of two quarters.

To estimate these treatment effects, we simulate draws from the posterior distribution of these treatment effects based on the sequence of MCMC iterations. To account for selection based on unobservables, we use the draws of the individual random effects $\alpha_{i,E}$ and $\alpha_{i,Q}$ from the MCMC estimation of the model. The details of the simulation procedure are given below.

First, we describe the simulation of the *Classical ATT*. For every 30th draw of the MCMC iterations (after the burn-in phase), we go through the following steps:

Step 1. For each participant, predict the treatment outcome E^1_{it} starting with the first period after program participation ($t - t_a = 1$, where t_a denotes the last quarter of program participation).[11] In particular, go from $t - t_a = 1$ to $t - t_a = 9$ and predict the employment status for each period based on the corresponding draw from the vector of coefficients η_E, the vector of explanatory variables $z_{it,E}$, the corresponding draw of the $\alpha_{i,E}$ and a draw of the idiosyncratic error term $\epsilon_{it,E}$. The dynamic elements of $z_{it,E}$, such as lags of employment status, are updated when moving from one quarter to the next. $\epsilon_{it,E}$ is drawn from a standard normal distribution.

Step 2. For each participant, simulate the counterfactual employment outcome (i.e. the outcome if the participant had not participated in a program) E^0_{it} for each period beginning with the quarter of program start ($t - t_s = 0$, where t_s denotes the first quarter of program participation). Again go through the dynamic process and predict the employment status for each period based on the same η_E, $\alpha_{i,E}$ and $\epsilon_{it,E}$ as before. Adapt the $z_{it,E}$ to a situation with no participation and update them while going through the process.

Step 3. To get a draw of the ATT aligned to the end of the program, average the difference of the two predictions over all treated individuals (N_1), i.e. $\frac{1}{N_1}\sum_{i=1}^{N_1}(E^1_{it} - E^0_{it})$ (t aligned to end of program), for each period $t > t_a$. This gives a draw from the posterior distribution of the ATT for each quarter.

Step 4. For a draw of the ATT aligned to the start of the program, average the difference of the two predictions over all treated individuals, i.e. $\frac{1}{N_1}\sum_{i=1}^{N_1}(E^1_{it} - E^0_{it})$ (t aligned to start of program), for each period $t \geq t_s$.

The resulting 1,500 draws provide an estimate of the posterior distribution of the *Classical ATT*. We estimate the ATT by the mean of the posterior distribution and we use the standard deviation as our estimate of estimation uncertainty.

Second, the estimation of the posterior distribution of the effect of *Training versus Waiting* proceeds in an analogous way. Step 1 remains the same. The counterfactual employment outcome E^w_{it} relates to a situation in which the participant does not start a program in the observed start quarter (t_s), so the employment status $E^w_{it_s}$ is simulated and $Q^w_{it_s}$ is set to zero. The individual may start a program later. Thus, from $t = t_s$ onwards, the counterfactual employment status E^w_{it} and the counterfactual participation status Q^w_{it} are simulated in turn, adapting the elements of $\boldsymbol{z}_{it,E}$ and $\boldsymbol{z}_{it,Q}$ that include lagged employment or participation status while going

[11] While in the program, the employment dummy is equal to zero.

through the dynamic processes. Note that $Q_{it}^{w}=1$ is not allowed if E_{it}^{w} is one in a given period or if an individual has already left a program. The simulation of E_{it}^{w} uses the respective draw of $\boldsymbol{\eta}_E$ and $\alpha_{i,E}$. The simulation of Q_{it}^{w} relies on the respective draw of $\boldsymbol{\eta}_Q$ and $\alpha_{i,Q}$ from the MCMC estimation of the model. The same $\epsilon_{it,E}$ as in step 1 are used and the $\epsilon_{it,Q}$ are drawn from a standard normal distribution. In order to calculate a draw of the effect of *Training versus Waiting* aligned to the start of the program, average the difference of the two predictions for each period over all participants ($\frac{1}{N_1}\sum_{i=1}^{N_1}(E_{it}^{1}-E_{it}^{w})$, t aligned to start of program). Likewise, calculate $\frac{1}{N_1}\sum_{i=1}^{N_1}(E_{it}^{1}-E_{it}^{w})$ (t aligned to end of program) to get a draw aligned to the end of the program.

Finally, consider the *Effect of a Given Planned Program Duration*. For each 30th draw, we simulate the employment status for different planned program durations. First, we simulate the employment status and participation status for a situation in which all participants are assigned to a program with a planned length of two quarters.[12] For the quarter when a participant starts the program ($t=t_s$), Q_{it}^{p2} is set to one and $E_{it_s}^{p2}$ is set to zero, as in the original data. In the next quarter ($t-t_s=1$), E_{it}^{p2} also remains the same as in the original data. The participation status $Q_{it_S}^{p2}$ in t_s is then simulated and in the following quarters $t-t_s>1$, E_{it}^{p2} and Q_{it}^{p2} are simulated in turn for each period. The elements of $\boldsymbol{z}_{it,E}$ and $\boldsymbol{z}_{it,Q}$ that include lags of employment or participation status are adapted while going through the dynamic processes. Again, the $\boldsymbol{\eta}_E$, $\boldsymbol{\eta}_Q$, $\alpha_{i,E}$ and $\alpha_{i,Q}$ of the respective draw of the MCMC estimation are used. $\epsilon_{it,E}$ and $\epsilon_{it,Q}$ are drawn from a standard normal distribution, respectively. Similarly, the employment status and the participation status are simulated for each period for the alternative scenario in which the planned program duration is set to one quarter (E_{it}^{p1} and Q_{it}^{p1}), three quarters (E_{it}^{p3} and Q_{it}^{p3}), and four quarters (E_{it}^{p4} and Q_{it}^{p4}), respectively. The same $\epsilon_{it,E}$ and $\epsilon_{it,Q}$ as before are used. As the median planned duration of the programs in the data is slightly more than two quarters, we take E_{it}^{p2} as a benchmark and calculate the effect of a planned duration of one quarter as opposed to two quarters ($\frac{1}{N_1}\sum_{i=1}^{N_1}(E_{it}^{p1}-E_{it}^{p2})$), of three quarters versus two quarters ($\frac{1}{N_1}\sum_{i=1}^{N_1}(E_{it}^{p3}-E_{it}^{p2})$), and four quarters versus

[12] In terms of the model specification, this means that the explanatory variables in the participation equation involving the planned end date (i.e. *days/91 until planned end if enough duration left* and *days/91 until planned end if enough duration left squared*) are adapted to this scenario. As these variables are measured in days, the decision whether there is still enough planned duration left and the values these variables take depend on the day within a quarter at which the program starts. Note that it is possible in the simulation as well as in the original data that realized program participation continues beyond the planned end date. The variable *planned end missing* is set to zero.

two quarters ($\frac{1}{N_1}\sum_{i=1}^{N_1}(E_{it}^{p4} - E_{it}^{p2})$) for each period $\{t : 0 \leq t - t_s \leq 9\}$. Again, we estimate the posterior distribution based on 1,500 draws from the simulations for the MCMC iterations.

5.4 Estimation Results

We estimate the impact of incidence and duration of training on the transition probabilities between employment and non-employment using the MCMC estimation approach described in the previous section. Our empirical model accounts for selection into training based on unobservables. Estimation is carried out separately for West German males, West German females, East German males, and East German females. The detailed estimation results are given in table 5.3 in Appendix C. The first column for each sample refers to the mean of the coefficients and the second to their standard deviation over MCMC iterations after the burn-in phase. We interpret them in an analogous same way as the point estimates and standard errors of the coefficients obtained by a frequentist approach. Next, we first briefly discuss the overall fit of the model and the individual level variances of the error terms. Because of the complexity of the model (it comprises about 160 parameters), we refrain from further discussing single parameters, but then discuss results for different treatment effects of interest.

5.4.1 Model Fit and Selection on Unobservables

Evidence on the fit of the model is provided in table 5.2 for the treated individuals from the start of the program onwards along with information on the number of observations available in each quarter. Actual and predicted employment rates of the trainees match closely in all four samples. Thus, our rich model specification does a good job in replicating the employment dynamics found in the data which gives some evidence that our model is not misspecified.

The last panel of table 5.3 in Appendix C displays the individual level variances of the composite error terms of the employment and the training equation, respectively, as well as the covariance between the two. The share of the variance that is due to the random effects varies between 36% and 51% for the employment equation and between 22% and 31% for the training equation. Except for the sample of

Table 5.2: Employment Rate and Number of Participants Still Observed Aligned to Start of Program

	Male West			Female West			Male East			Female East		
$t - t_s$	$\bar{E}_t$	$\hat{E}_t$	N_t	$\bar{E}_t$	$\hat{E}_t$	N_t	$\bar{E}_t$	$\hat{E}_t$	N_t	$\bar{E}_t$	$\hat{E}_t$	N_t
0	0.000	0.000*	1740	0.000	0.000*	1431	0.000	0.000*	1300	0.000	0.000*	848
1	0.079	0.079*	1740	0.070	0.070*	1431	0.048	0.048*	1300	0.039	0.039*	848
2	0.179	0.177	1721	0.162	0.165	1411	0.120	0.123	1290	0.081	0.078	840
3	0.239	0.244	1696	0.264	0.267	1385	0.178	0.177	1282	0.118	0.121	834
4	0.302	0.301	1664	0.354	0.355	1366	0.226	0.218	1265	0.172	0.185	825
5	0.334	0.346	1623	0.415	0.412	1338	0.278	0.284	1229	0.240	0.239	816
6	0.371	0.370	1577	0.442	0.440	1316	0.311	0.320	1201	0.279	0.274	795
7	0.371	0.377	1526	0.450	0.454	1291	0.318	0.335	1159	0.320	0.300	765
8	0.364	0.387	1465	0.464	0.468	1253	0.345	0.355	1106	0.329	0.330	741
9	0.396	0.405	1393	0.478	0.481	1213	0.378	0.391	1039	0.356	0.355	710

Notes: t_s denotes the quarter of program start, and $\bar{E}_t$ the sample mean of the employment dummy in quarter t. $\hat{E}_t$ is the mean of the employment dummy as predicted using the simulation strategy (prediction of treatment outcomes).
* Observed value is taken for the first period.

West German females, the correlation between the two random effects tends to be significantly negative. The correlation coefficient is -.22 and significant at the five percent level for females in East Germany. It is -.11 to -.12 for West and East German males and significant at the ten percent level. This suggests that those individuals who have a higher unobserved propensity to enter a program and to stay in a program tend to have a lower unobserved propensity to be employed.

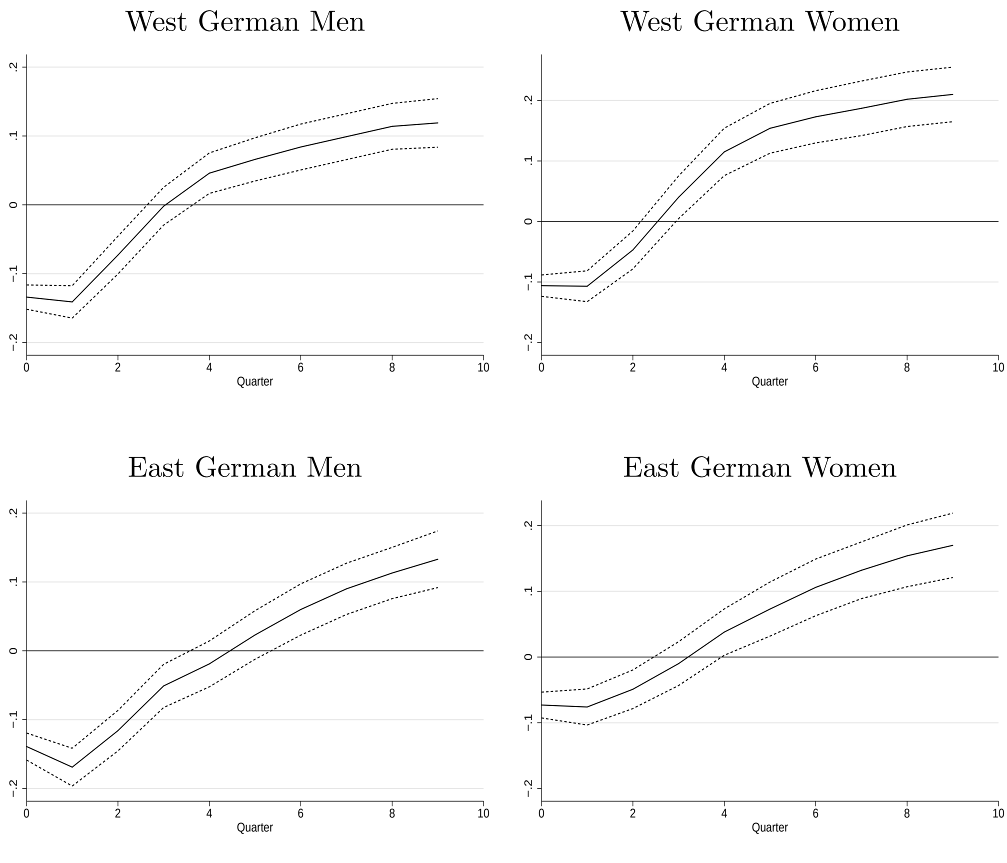
5.4.2 Classical Treatment Effect on Employment Probability

Figure 5.4 shows the average effect of training versus no training for those who participate on the quarterly employment probability.[13] More precisely, we compare the average of the actual employment outcomes of trainees with the expected counterfactual outcome obtained by setting the lags of training status in the employment equation to zero. In figure 5.4 the average difference in the quarterly employment rates is depicted on the vertical axis, while quarters since program start are measured on the horizontal axis. The dashed lines around the estimated treatment

[13]The corresponding numbers are given in table 5.4 in Appendix C.

effects are 95 percent confidence bands. Treatment effects for a particular quarter are statistically significant if zero is not contained in the confidence bands. As can be seen, a participation in training reduces the employment probability during the first three to five quarters after program start. During the first two quarters ($t - t_s \leq 1$) the employment probabilities of participants decline between seven (East German females) and 15 (East German males) ppoints compared to the situation of no participation. This lock-in effect lasts one quarter longer in the East German samples compared to the West German ones. After the first year counted from the quarter of program start, the difference in employment rates turns positive and continues to increase until the end of the observation window. Ten quarters after program start ($t - t_s = 9$), West German females have a 21 ppoints higher employment probability than in the absence of training. The effects are of similar magnitude for East German females (17 ppoints) and somewhat smaller for the male samples with around 12 ppoints.

Figure 5.4: Classical Average Treatment Effect on the Treated



5.4.3 Training versus Waiting

In the recent evaluation literature dealing with dynamic program starts, researchers often focus on the effect of treatment at a given point in time versus no treatment at that point in time, implying that the treatment may take place at some later point in time. Thus, the control group is a mixture of individuals who never participate and those who defer participation. This effect is commonly referred to as the effect of training versus waiting (Sianesi, 2004). In order to mimic this parameter, we simulate the training status of the actual trainees imposing that they postpone participation at least one quarter beyond their observed true program start. This entails the possibility that the simulated training dummies are zero during all quarters for some individuals. Table 5.5 in Appendix C depicts the actual and the simulated participation rates. From the last row of table 5.5, it can be seen that under the simulated waiting scenario only 68.7 % of the original participants ever enrol into training.

Figure 5.5: ATT of Training versus Waiting

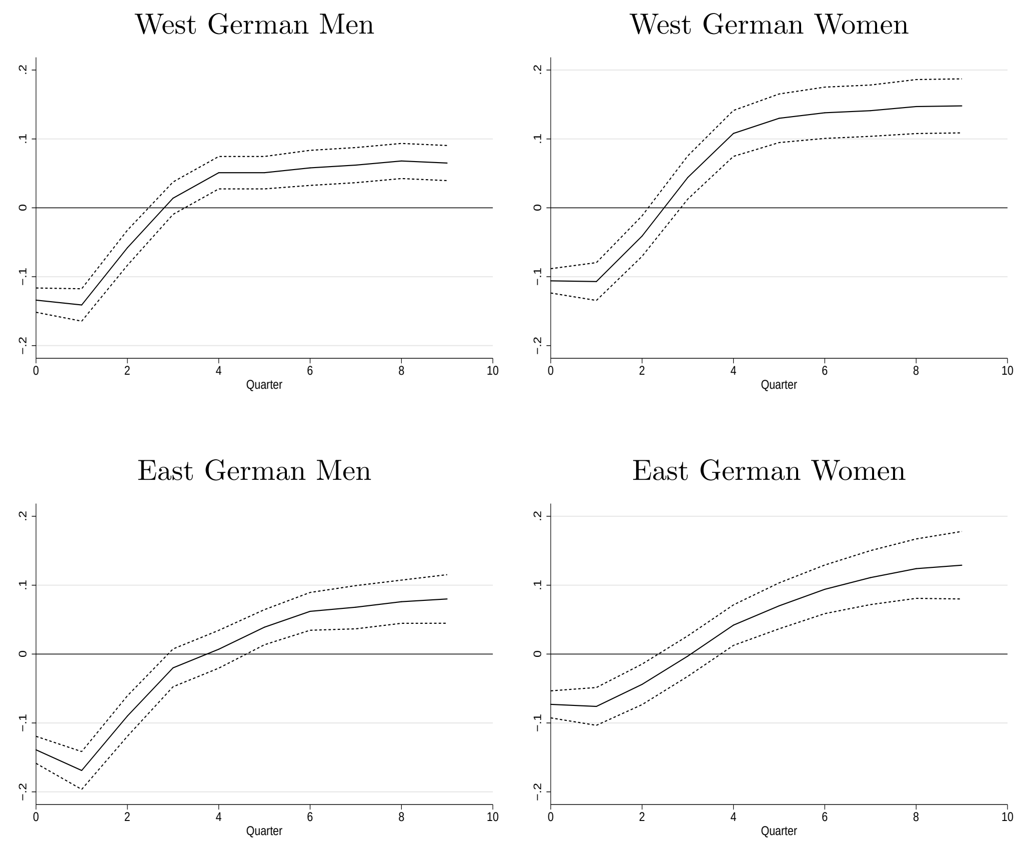


Figure 5.5 displays the evaluation results for the scenario of training versus wait-

ing.[14] There are negative lock-in effects of similar magnitude and length as for the classical effect of training in figure 5.4. In the quarter in which the program starts, participants in training have a seven to 14 ppoints lower probability to be in employment than compared to the situation of not yet starting a program. After about four to five quarters ($t - t_s = 3, 4$), the treatment effects turn positive and then increase further during the subsequent quarters. In quarter ten since program start ($t - t_s = 9$), they lie in the range of 7 to 15 ppoints. This is about a third less than compared to the case of a pure no-training control group.

Based on our model estimates, the estimated effect for treatment versus waiting underestimates the causal (classical) treatment effect in the medium and long run. This finding is due to the fact that control persons who obtain training in the near future also experience positive treatment effects in the medium and long run.

5.4.4 Variation in Planned Training Duration

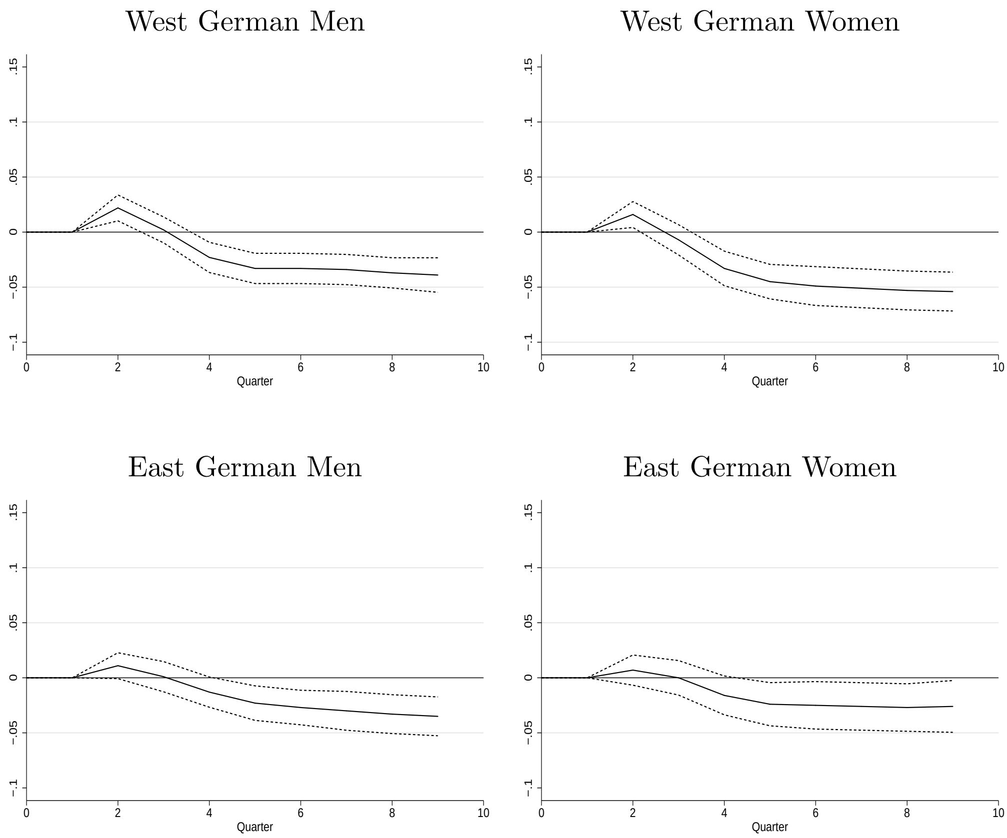
Here, we use our model estimates to analyze how treatment effects vary with the planned program duration. Indeed, the optimal length of training is a question of particular interest to policy makers. For this purpose, we simulate the training and employment histories of the actual trainees in the data that result after fixing the planned program duration to a prespecified value. In particular, we consider planned program durations of one, two, three, and four quarters. We then evaluate the effect of participating in a program scheduled over one, three, and four quarters, respectively, as opposed to two quarters, the median of planned duration in the data. Tables 5.7 to 5.10 in Appendix C show the simulated participation and employment probabilities associated with different planned program durations. Note that the simulated realized program duration can be shorter or longer than the planned one. However, the tables suggest that there is a strong positive correlation between planned and realized program durations.

Figure 5.6 displays the treatment effects associated with a planned duration of one versus two quarters.[15] The gains of a shorter participation are small and only transient. In the third quarter after program start ($t - t_s = 2$), the employment probability is between one and two ppoints higher. In the medium and long run,

[14]The corresponding numbers are given in table 5.6 in Appendix C.

[15]The corresponding numbers are given in table 5.11 in Appendix C.

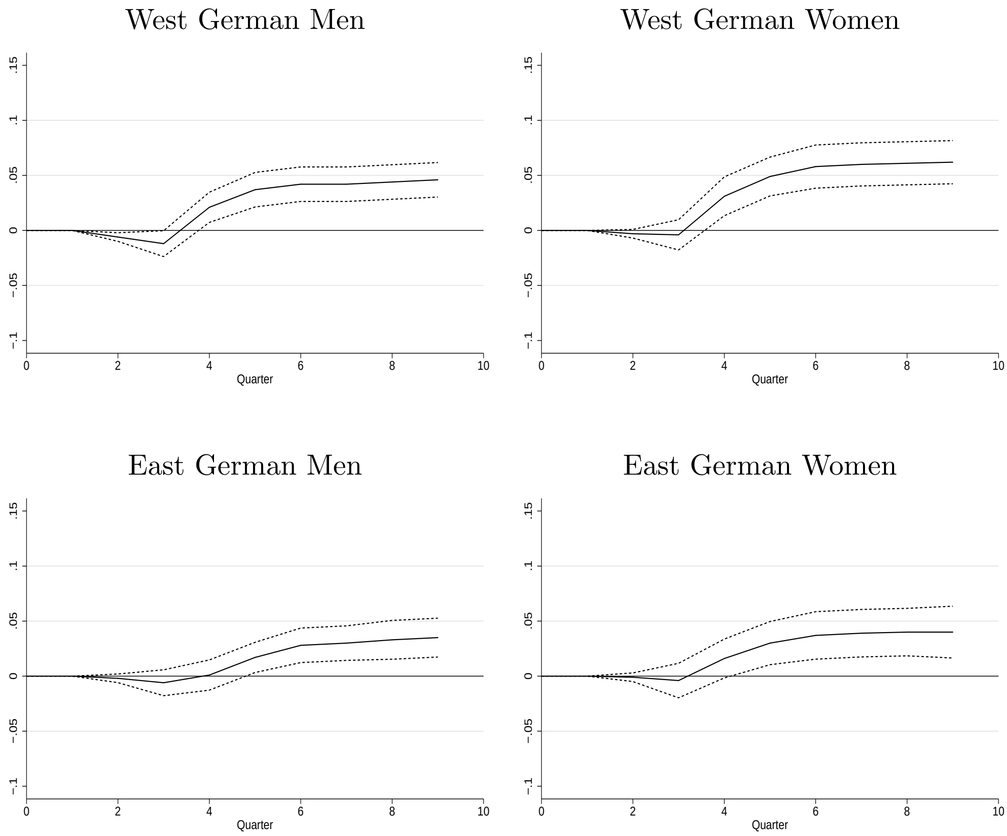
Figure 5.6: ATT of Attending a Program Scheduled for One versus Two Quarters



those attending programs with a scheduled length of two quarters fare better, exhibiting employment rates that are consistently higher by three to five ppoints. A similar picture arises when comparing programs with a scheduled length of three and four quarters, respectively, with those planned to last two quarters, cf. figures 5.7 and 5.8.[16] Trainees attending longer programs are only slightly worse off during the additional quarters they are supposed to be in the program. After the scheduled end of the longer program, they have consistently higher employment rates than compared to the benchmark case of a six-month program. Indeed, compared to a planned duration of two quarters, the employment rates associated with attending a nine-month program are four to six ppoints higher and those associated with attending a one-year program are six to eleven ppoints higher.

[16]The corresponding numbers are given in tables 5.12 and 5.13 in Appendix C.

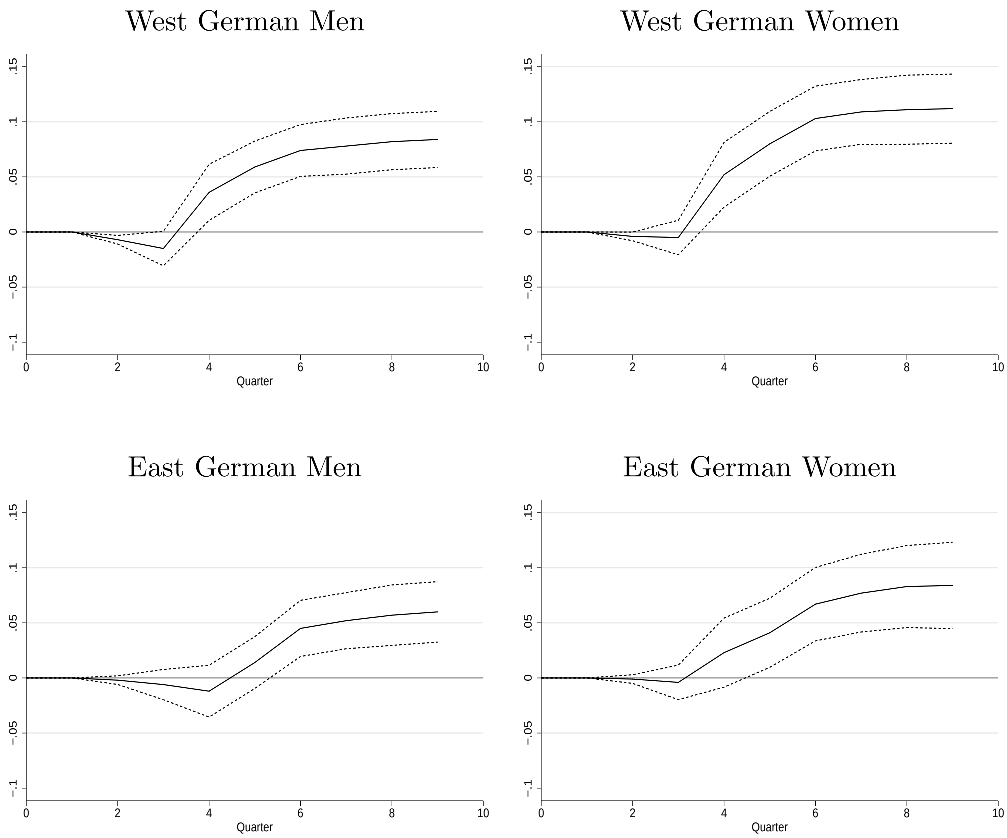
Figure 5.7: ATT of Attending a Program Scheduled for Three versus Two Quarters



5.5 Conclusion

This paper estimates the effects of long-term training on discrete time labor market transitions in Germany using a dynamic random effects probit model with an employment equation and a participation equation. The participation equation models the start of participation in long-term training as well as the end of participation accounting for endogenous dropout. We control for selection on unobservables by allowing the random effects of both equations to be dependent. The models are specified in a flexible way and we account for various forms of effect heterogeneity. Using Bayesian Markov Chain Monte Carlo (MCMC) methods, we estimate the posterior distribution of the model parameters, including the individual random effects and the treatment effects of interest. The employment equation and the training equations are estimated simultaneously. We estimate separate models for West and East Germany and for males and females. The interpretation of the means of the parameters is not straightforward because of the complexity and the dynamic na-

Figure 5.8: ATT of Attending a Program Scheduled for Four versus Two Quarters



ture of the model. The posterior distributions of the treatment effects are estimated simulating employment and treatment outcomes based on the model estimates along the MCMC iterations.

Our results imply positive reemployment effects soon after the participants have left the program and positive employment effects persist until the end of our observation period. Two years after the end of program participation, the employment effect of the program lies in between 12 and 21 ppoints. These effects are more positive than the results reported in the literature, especially for East Germany. This is consistent with our model estimates implying a negative selection of treated individuals. Our estimation approach allows us to estimate various other treatment effects of interest. The effects of treatment versus waiting turn out to be positive in the medium and long run but it is smaller than the classical treatment effect. Increasing the planned duration of training shows positive effects on the treatment effect after the end of the program.

On the methodological side, there are several extensions to the literature as discussed in the introduction. Most importantly, we model the transitions in and out of the program. This enables us to estimate the effect of different planned program durations, which has rarely been estimated in the literature. Furthermore, the model estimation shows that many coefficients on state dependence, duration dependence and various interactions as well as part of the coefficients on the employment history are significant. Thus it seems to be important to use a highly flexible model. Using MCMC methods enables a robust estimation of our model and provides information on the random effects which we need for our simulation approach. Our simulation approach allows to directly interpret the treatment effects of interest, which greatly eases the interpretation of the results. It is difficult to judge what we might have gained by using the full information of the discrete data and what we might have lost by using only a quarterly frequency, by assuming independence between the exogenous covariates and the individual random effects, and by not modeling completely separate discrete choice models for both non-employment and employment. Overall, as discussed in section 5.4.1, the model fits the data well. In sum, the model and estimation techniques used in this paper permit to estimate additional treatment effects of interest as compared to standard methods, while accounting very flexibly for selection and heterogeneity. Thus, the effort of using a non-standard evaluation approach has been worthwhile.

References

Abbring, J. and G.J. van den Berg (2003). "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica* 71, 1491-1517.

Abbring, J. and G.J. van den Berg (2004). "Analyzing the Effect of Dynamically Assigned Treatments using Duration Models, Binary Treatment Models, and Panel Data Models." *Empirical Economics* 29, 5-20.

Adda, J., M. Costas Dias, C. Meghir and B. Sianesi (2007). "Labour market programmes and labour market outcomes: a study of the Swedish active labour market interventions." IFAU Working Papier, 2007: 27.

Bergemann, A., B. Fitzenberger and S. Speckesser (2009). "Evaluating the Dynamic Employment Effects of Training Programs in East Germany Using Conditional Difference-in-Differences." *Journal of Applied Econometrics*, forthcoming.

Biewen, M., B. Fitzenberger, A. Osikominu and M. Waller (2007). "Which Program for Whom? Evidence on the Comparative Effectiveness of Public Sponsored Training Programs in Germany." IZA Discussion Paper, No. 2885, Bonn.

Buchinsky, M., D. Fougère, F. Kramarz and R. Tchernis (2005). "Interfirm Mobility, Wages, and the Returns to Seniority and Experience in the U.S." IZA Discussion Paper No. 1521.

Bundesanstalt für Arbeit, BA, (2001). *Arbeitsstatistik 2000 - Jahreszahlen.* Nürnberg.

Bundesagentur für Arbeit, BA, (2006). *Arbeitsstatistik 2005 - Jahreszahlen.* Nürnberg.

Bundesagentur für Arbeit, BA, (2007, 2008). *Arbeitsmarkt 2006, 2007.*" Nürnberg.

Card, D., J. Kluve, and A. Weber (2009). "Active Labor Market Policy Evaluations: A Meta-Analysis." CESifo Working Paper, No. 2570, Munich.

Chib, S. (2001). "Markov Chain Monte Carlo Methods: Computation and Inference." In J. Heckman and E. Leamer (eds.): *Handbook of Econometrics*, Volume 5, Amsterdam: Elsevier Science, 3569-3649.

Chib, S. and B.H. Hamilton (2002). "Semiparametric Bayes Analysis of Longitudinal Data Treatment Models", *Journal of Econometrics* 110, 67-89.

Chib, S. and L. Jacobi (2007). "Modeling and Calculating the Effect of Treatment at Baseline from Panel Outcomes." *Journal of Econometrics* 140, 781-801.

Heckman, J., H. Ichimura, J.A. Smith and P. Todd (1998). "Characterizing Selection Bias using Experimental Data." *Econometrica*, 65, 1017-1098.

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

Horny, G., R. Mendes and G.J. van den Berg (2009). "Job Durations and Firm Specific Effects: MCMC Estimation with Longitudinal Employer-Employee Data." IZA Discussion Paper No. 3992, Bonn.

Imbens, G. (2000). "The Role of the Propensity Score in Estimating Dose-Response Functions." *Biometrika.* 87, 706-710.

Kluve, J. (2006). "The Effectiveness of European Active Labor Market Policy." Discussion Paper, RWI Essen.

Kluve, J., H. Schneider, A. Uhlendorff and Z. Zhao (2007). "Evaluating Continuous Training Programs Using the Generalized Propensity Score." IZA Discussion Paper No. 3255, Bonn.

Kühnlein, G. and B. Klein (2003). "Bildungsgutscheine - mehr Eigenverantwortung, mehr Markt, mehr Effizienz? - Erfahrungen bei der Neuausrichtung der beruflichen Weiterbildung." Arbeitspapier Band 74, Hans-Böckler-Stiftung.

Lechner, M., R. Miquel and C. Wunsch (2009). "Long-Run Effects of Public Sector Sponsored Training in West Germany." *Journal of the European Economic Association*, forthcoming.

Lechner, M. and C. Wunsch (2006). "Active Labour Market Policy in East Germany: Waiting for the Economy to Take Off." *The Economics of Transition*, forthcoming.

Lechner, M. and C. Wunsch (2007). "What Did All the Money Do? On the General Ineffectiveness of Recent West German Labour Market Programmes." *Kyklos* 61, 134-174.

Martin, J.P. (2000). "What works among active labour market policies: Evidence from OECD countries' experiences." *OECD Economic Studies* No. 30, 79-113.

OECD (2005). "Labour Market Programmes and Activation Strategies: Evaluating the Impacts", *OECD Employment Outlook*, Ch. 4.

Odejar, A. (2002). "Bayesian Analysis of Sample Selection and Endogenous Switching Regression Models with Random Coefficients Via MCMC Methods." SFB 386 Discussion Paper No. 291.

Osikominu, A. (2008). "Is Short Training Short-Lived and Long Training Long-Lasting? A Multi-State Duration Analysis of the Dynamic Effects of Training Schemes for the Unemployed." Discussion Paper, Albert-Ludwigs-University Freiburg.

Plaßmann, G. (2002). "Der Einfluss der Arbeitslosenversicherung auf die Arbeitslosigkeit in Deutschland." *Beiträge zur Arbeitsmarkt- und Berufsforschung, BeitrAB* 255.

Richardson, K. and G.J. van den Berg (2008). "Duration dependence versus unobserved heterogeneity in treatment effects: Swedish labor market training and the transition rate to employment." IFAU Working Paper, 2008:7, Uppsala.

Schneider, H., K. Brenke, D. Hess, L. Kaiser, J. Steinwedel and A. Uhlendorff (2006). "Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission - Modul 1b: Förderung beruflicher Weiterbildung und Transferleistungen." IZA Research Report No. 7, Bonn.

Schneider, H. and A. Uhlendorff (2006). "Die Wirkung der Hartz-Reform im Bereich der beruflichen Weiterbildung." *Journal for Labor Market Research* 39, 477-490.

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

Stephan, G. (2009). "The effects of active labor market programs in Germany - an investigation using different definitions of non-treatment." *Jahrbücher für Nationalökonomie und Statistik*, forthcoming.

Train, K. (2003). *Discrete Choice Methods with Simulation.* Cambridge: Cambridge University Press.

Waller, M. (2008). "On the Importance of Correcting Reported End Dates of Labor Market Programs." *Schmollers Jahrbuch* 128, 213-236.

Waller, M. (2009). Many Dropouts? Never Mind! - Employment Prospects of Dropouts from Training Programs, Freiburg University, not published.

Zeger, S. and M. Karim (1991). "Generalized linear models with random effects: a Gibbs sampling approach." *Journal of the American Statistical Association*, 86, 79-86.

Zimmermann, R., S. Kaimer, D. Oberschachtsiek (2007). "Dokumentation des Scientific Use Files der Integrierten Erwerbsbiographien (IEBS-SUF V1) Version 1.0." FDZ Datenreport 1/2007, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Appendix A: Detailed Information on the Data

This study uses data from the IEBS Version 4.02. A German description of the IEBS Version 3.01 can be found in Zimmermann et al. (2007). Information in English can be found on the website of the Research Data Center of the Federal Employment Agency (http://fdz.iab.de/en.aspx). The website also describes the conditions under which researchers may obtain access to the IEBS.

The first of the four administrative data sources included in the IEBS, the IAB Employment History, consists of social insurance register data for employees subject to contributions to the public social security system. It covers the time period from 1990 to 2004. The main feature of these data is detailed daily information on the employment status of each recorded individual. For each employment spell, in addition to start and end dates, data from the Employment History contain information on personal as well as job and firm characteristics such as wage, industry or occupation.

The IAB Benefit Recipient History, the second data source, includes daily spells of unemployment benefit, unemployment assistance and subsistence allowance payments the individuals received between January 1990 and June 2005. In addition to the sort of the payment and the start and end dates of periods of transfer receipt the spells contain further information like sanctions, periods of disqualification from benefit receipt and personal characteristics. Furthermore, the information in the Employment and the Benefit Recipient History allows one to calculate the individual entitlement periods to unemployment benefits.[17]

The third data source included in the IEBS is the so-called Data on Job Search Originating from the Applicants Pool Database, which contains rich information on individuals searching for jobs. It contains all the records starting January 2000 to June 2005 and partly also those beginning before 2000 if the person in question keeps the same client number throughout. The database includes a rich variety of information on personal characteristics (in particular education, family status and health condition), information related to placement fields (e.g. qualification and experience in the target profession), and regional information.

The Participants-in-Measures Data, the fourth data source, contains diverse information on participation in public sector sponsored labor market programs, for example training programs, job-creation measures, integration subsidies, business start-up allowances covering the period January 2000 to July 2005. Comparing the entries into different programs in 1999 with the figures for later years shows that information on programs starting in 1999 seems to be already complete for most active

[17] For the calculation of the claims, the present study relies on Plaßmann (2002) that contains a summary of the different regulations.

labor market programs. Furthermore, this database allows to distinguish subsidized employment in the context of active labor market policy from regular employment. Similar to the other sources, information comes in the form of spells indicating the start and end dates at the daily level, the type of the program as well as additional information on the program such as the planned end date or if the program ends with a certificate.

Appendix B: Algorithm for the MCMC Estimation

The posterior distribution combines the likelihood (see section 5.3.1) and the priors. We set the following independent priors: the prior distributions of the coefficients $\boldsymbol{\eta}_E$ are given by independent normal priors with distribution $\mathcal{N}(b_{E,0}, B_{E,0})$. $\mathcal{N}(\boldsymbol{.})$ denotes the normal distribution. Setting very large values for the variance $B_{E,0}$, we use extremely diffuse priors. The same is done for the elements of the coefficient vector $\boldsymbol{\eta}_Q$, whose prior distributions are given by $\mathcal{N}(b_{Q,0}, B_{Q,0})$. The prior distribution of the random effects is $\mathcal{N}(\mathbf{0}, \Sigma)$. The hyperparameter Σ^{-1} follows the prior distribution $\mathcal{W}^{-1}(H_0, h_0)$, where H_0 is the inverse scale matrix and h_0 denotes the degrees of freedom. $\mathcal{W}^{-1}$ denotes the inverse Wishart distribution. In order to set a diffuse prior, we choose a small value for h_0. The diagonal elements of H_0 are set to the individual level variances of separate Maximum Likelihood estimations of the two equations multiplied by h_0, and the off-diagonal elements are set to zero.

- Step 0: Set starting values for the coefficient vectors $\boldsymbol{\eta}_E$ and $\boldsymbol{\eta}_Q$, the random effects ($\alpha_{i,E}$, $\alpha_{i,Q}$), and the variance covariance matrix of the random effects Σ.
- Step 1a: Sample E^*_{it} from $\mathcal{N}(\boldsymbol{z}_{it,E}\boldsymbol{\eta}_E + \alpha_{i,E}, 1)$ with support $[0, \infty]$ if $E_{it} = 1$ and with support $[-\infty, 0]$ if $E_{it} = 0$.
- Step 1b: Sample Q^*_{it} from $\mathcal{N}(\boldsymbol{z}_{it,Q}\boldsymbol{\eta}_Q + \alpha_{i,Q}, 1)$ with support $[0, \infty]$ if $Q_{it} = 1$ and with support $[-\infty, 0]$ if $Q_{it} = 0$ (provided the training equation is to be estimated).
- Step 2: Sample $(\alpha_{i,E}, \alpha_{i,Q})'$ from its bivariate normal conditional posterior distribution $\mathcal{N}(\boldsymbol{\mu}, V_{\alpha_i})$, where $\boldsymbol{\mu} = V_{\alpha_i} \cdot \begin{pmatrix} T_{i,E} & 0 \\ 0 & T_{i,Q} \end{pmatrix} \cdot \begin{pmatrix} (\bar{E}^*_i - \bar{\boldsymbol{z}}_{i,E}\boldsymbol{\eta}_E) \\ (\bar{Q}^*_i - \bar{\boldsymbol{z}}_{i,Q}\boldsymbol{\eta}_Q) \end{pmatrix}$ and $V_{\alpha_i} = \left(\Sigma^{-1} + \begin{pmatrix} T_{i,E} & 0 \\ 0 & T_{i,Q} \end{pmatrix}\right)^{-1}$, a bar over a variable denotes its mean across time, $T_{i,E}$ the number of observations for person i, and $T_{i,Q}$ the number of observations for person i for which the training equation is to be estimated.

- Step 3a: Sample the $\boldsymbol{\eta}_E$ vector from its multivariate normal conditional posterior distribution $\mathcal{N}(M_E, V_E)$, where $M_E = V_E(B_{E,0}^{-1} b_{E,0} + \sum_{i=1}^{N} \sum_{t=1}^{T_{i,E}} \boldsymbol{z}'_{it,E}(E_{it}^* - \alpha_{i,E}))$ and $V_E = (B_{E,0}^{-1} + \sum_{i=1}^{N} \sum_{t=1}^{T_{i,E}} \boldsymbol{z}'_{it,E}\boldsymbol{z}_{it,E})^{-1}$. N is the number of persons in the data.

- Step 3b: If the training equation is to be estimated, sample the $\boldsymbol{\eta}_Q$ vector from its multivariate normal conditional posterior distribution $\mathcal{N}(M_Q, V_Q)$, where $M_Q = V_Q(B_{Q,0}^{-1} b_{Q,0} + \sum_{i=1}^{N} \sum_{t=1}^{T_{i,Q}} \boldsymbol{z}'_{it,Q}(Q_{it}^* - \alpha_{i,Q}))$ and $V_Q = (B_{Q,0}^{-1} + \sum_{i=1}^{N} \sum_{t=1}^{T_{i,Q}} \boldsymbol{z}'_{it,Q}\boldsymbol{z}_{it,Q})^{-1}$.

- Step 4: Sample Σ^{-1} from its conditional posterior distribution
$$\mathcal{W}^{-1}\left(\begin{pmatrix} \sum_{i=1}^{N} \alpha_{i,E}^2 & \sum_{i=1}^{N} \alpha_{i,E}\alpha_{i,Q} \\ \sum_{i=1}^{N} \alpha_{i,E}\alpha_{i,Q} & \sum_{i=1}^{N} \alpha_{i,Q}^2 \end{pmatrix} + H_0,\ N + h_0\right).$$ Go to Step 1. Always use the current parameter values.

Appendix C: Detailed Estimation Results

Table 5.3: Means and Standard Deviations of Parameters from MCMC Estimation

Name	Male West Mean	Male West SD	Female West Mean	Female West SD	Male East Mean	Male East SD	Female East Mean	Female East SD
Employment Equation								
Q_{t-1}	-0.478	0.064	-0.300	0.079	-0.761	0.080	-0.351	0.112
$[Q_{t-1} = 1] \times \ldots$								
$\ldots [Q_{t-2} = 1] \times [Q_{t-3} = 0]$	0.320	0.078	0.323	0.088	0.474	0.097	0.120	0.147
$\ldots [Q_{t-3} = 1] \times [Q_{t-4} = 0]$	0.681	0.091	0.844	0.096	0.776	0.108	0.526	0.143
$\ldots [Q_{t-4} = 1] \times [Q_{t-5} = 0]$	1.222	0.111	1.418	0.120	0.834	0.124	0.996	0.148
$\ldots \sum_{j=5}^{6} [Q_{t-j} = 1 \times Q_{t-7} = 0]$	0.666	0.137	1.012	0.132	1.090	0.121	0.711	0.171
$\ldots \sum_{j=7}^{16} [Q_{t-j} = 1]$	0.596	0.131	0.373	0.084	0.919	0.158	0.216	0.293
$\ldots \tau_{t,Q} \times$ unskilled	0.013	0.032	-0.035	0.036	0.079	0.060	-0.075	0.097
$\ldots \tau_{t,Q} \times$ high school	-0.041	0.031	-0.004	0.033	-0.039	0.037	0.013	0.047
$\ldots \tau_{t,Q} \times$ health probl.	0.052	0.044	0.090	0.049	-0.077	0.090	0.089	0.108
$\ldots \tau_{t,Q} \times$ age ≥ 50	0.043	0.051	-0.127	0.058	-0.021	0.045	-0.142	0.062

<continued on next page>

Means and Standard deviations of Parameters from MCMC Estimation <continued>

Name	Male West Mean	SD	Female West Mean	SD	Male East Mean	SD	Female East Mean	SD
$[Q_{t-1}=0]\times[E_{t-1}=0]\times\ \ldots$								
$\ldots D_t$	0.125	0.134	-0.008	0.162	-0.599	0.161	-0.602	0.236
$\ldots D_t \times tsb$	-0.095	0.027	-0.026	0.031	0.083	0.035	0.109	0.050
$\ldots D_t \times tsb^2$	0.008	0.002	0.004	0.002	-0.003	0.002	-0.001	0.003
$\ldots D_t \times \tau_{t,Q}$	0.362	0.068	0.432	0.076	0.303	0.077	0.307	0.098
$\ldots D_t \times \tau^2_{t,Q}$	0.001	0.009	-0.002	0.011	-0.012	0.011	0.018	0.014
$\ldots D_t \times tsb \times \tau_{t,Q}$	-0.029	0.005	-0.030	0.006	-0.009	0.006	-0.033	0.008
$\ldots D_t \times$ unskilled	-0.205	0.111	-0.093	0.139	0.238	0.208	0.394	0.341
$\ldots D_t \times$ high school	-0.228	0.150	-0.322	0.183	0.514	0.206	0.539	0.234
$\ldots D_t \times$ health problems	-0.275	0.168	-0.074	0.229	-0.295	0.316	0.607	0.334
$\ldots D_t \times$ age > 50	0.099	0.177	0.004	0.191	-0.283	0.192	-0.616	0.262
$\ldots D_t \times \tau_{t,Q} \times$ unskilled	0.035	0.045	0.011	0.054	-0.078	0.080	-0.293	0.131
$\ldots D_t \times \tau_{t,Q} \times$ high school	0.047	0.049	0.062	0.063	-0.199	0.058	-0.114	0.066
$\ldots D_t \times \tau_{t,Q} \times$ health problems	0.045	0.065	-0.040	0.095	-0.080	0.100	-0.039	0.112
$\ldots D_t \times \tau_{t,Q} \times$ age ≥ 50	-0.038	0.071	-0.080	0.075	0.030	0.058	0.105	0.072
$[Q_{t-1}=0]\times[E_{t-1}=1]\times\ \ldots$								
$\ldots D_t$	0.107	0.167	0.349	0.213	-0.018	0.215	0.650	0.348
$\ldots D_t \times tsb$	0.025	0.033	-0.044	0.039	0.051	0.043	0.041	0.065
$\ldots D_t \times tsb^2$	-0.3^{-4}	0.002	0.002	0.002	-0.002	0.003	-0.003	0.004
$\ldots D_t \times \tau_{t,Q}$	0.203	0.082	0.220	0.090	0.123	0.097	-0.170	0.169
$\ldots D_t \times \tau^2_{t,Q}$	-0.011	0.011	-0.004	0.011	0.001	0.013	0.040	0.026
$\ldots D_t \times tsb \times \tau_{t,Q}$	-0.006	0.006	-0.001	0.007	-0.003	0.007	0.001	0.010
$\ldots D_t \times$ unskilled	-0.169	0.133	0.172	0.162	0.502	0.299	0.723	0.536
$\ldots D_t \times$ high school	-0.140	0.173	0.146	0.202	0.133	0.221	-0.152	0.277
$\ldots D_t \times$ health problems	-0.442	0.200	-0.118	0.260	1.523	0.650	-0.478	0.495
$\ldots D_t \times$ age > 50	-0.692	0.226	-0.166	0.215	-0.255	0.234	-0.316	0.342
$\ldots D_t \times \tau_{t,Q} \times$ unskilled	0.006	0.055	-0.081	0.062	-0.146	0.114	-0.288	0.205
$\ldots D_t \times \tau_{t,Q} \times$ high school	-0.028	0.055	-0.077	0.065	-0.086	0.065	0.006	0.084
$\ldots D_t \times \tau_{t,Q} \times$ health prob.	0.206	0.075	0.096	0.093	-0.987	0.322	0.521	0.165
$\ldots D_t \times \tau_{t,Q} \times$ age ≥ 50	0.206	0.094	0.022	0.084	0.058	0.083	0.094	0.108

<continued on next page>

Means and Standard deviations of Parameters from MCMC Estimation <continued>

Name	Male West		Female West		Male East		Female East	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
E_{t-1}	1.913	0.129	2.027	0.152	1.721	0.204	2.342	0.280
E_{t-2}	-0.220	0.032	-0.316	0.044	-0.267	0.046	-0.359	0.075
E_{t-3}	-0.278	0.034	-0.293	0.047	-0.272	0.049	-0.279	0.081
E_{t-4}	0.425	0.014	0.277	0.021	0.268	0.021	0.290	0.035
$\sum_{j=5}^{8} E_{t-j}$	-0.178	0.007	-0.212	0.009	-0.141	0.010	-0.250	0.015
$\sum_{j=9}^{12} E_{t-j}$	-0.123	0.008	-0.128	0.010	-0.070	0.011	-0.138	0.018
$\sum_{j=13}^{16} E_{t-j}$	-0.220	0.012	-0.229	0.017	-0.197	0.018	-0.284	0.031
$t > 1$	0.327	0.022	0.396	0.028	0.348	0.031	0.430	0.048
$t > 2$	0.041	0.021	0.078	0.026	0.046	0.030	0.091	0.046
$t > 3$	-0.154	0.021	-0.052	0.027	-0.114	0.030	-0.146	0.047
$t > 4$	0.286	0.021	0.233	0.027	0.294	0.030	0.297	0.048
$t > 5$	0.035	0.020	0.076	0.025	0.004	0.028	0.120	0.044
$t > 9$	0.207	0.019	0.192	0.025	0.187	0.027	0.231	0.042
$t > 13$	0.244	0.021	0.200	0.026	0.164	0.030	0.246	0.045
$E_{t-1} \times t$	0.079	0.045	0.014	0.064	-0.068	0.065	-0.070	0.109
$E_{t-2} \times t$	-0.003	0.048	-0.001	0.068	0.076	0.069	-0.114	0.113
$E_{t-3} \times t$	0.482	0.049	0.412	0.070	0.407	0.073	0.462	0.118
$\tau_{t,E} \times [E_{t-1} = 0]$	0.101	0.019	-0.121	0.020	0.028	0.029	-0.116	0.034
$\tau_{t,E} \times [E_{t-1} = 1]$	0.002	0.023	-0.039	0.027	0.088	0.036	-0.105	0.051
$\tau^2_{t,E} \times [E_{t-1} = 0]$	0.003	0.001	0.006	0.001	0.004	0.001	0.003	0.001
$\tau^2_{t,E} \times [E_{t-1} = 1]$	0.005	0.001	0.001	0.001	0.000	0.001	0.001	0.002
last job: assisting workers	-0.096	0.023	0.017	0.044	-0.149	0.032	-0.096	0.079
last job: jobs in service	-0.097	0.035	0.057	0.038	-0.081	0.057	-0.077	0.065
last job: office or business job	-0.077	0.038	0.082	0.042	-0.199	0.068	-0.039	0.073
last job: technician or related	-0.032	0.039	0.054	0.047	-0.068	0.058	-0.050	0.084
last job: academic or managers	-0.050	0.044	0.092	0.049	-0.109	0.066	-0.043	0.090
share last wages censored	0.860	0.136	0.303	0.199	0.647	0.262	1.105	0.378
log last average real wage	0.202	0.042	0.148	0.054	0.168	0.140	0.242	0.118
log last average real wage squared	0.017	0.006	-0.007	0.008	0.022	0.020	-0.009	0.017

<continued on next page>

Means and Standard deviations of Parameters from MCMC Estimation <continued>

	Male West		Female West		Male East		Female East	
Name	Mean	SD	Mean	SD	Mean	SD	Mean	SD
last job: whitecollar job	-0.077	0.032	-0.037	0.035	-0.128	0.050	-0.017	0.060
last job: seasonal worker	0.198	0.031	0.215	0.038	0.202	0.040	0.305	0.062
last job: parttime worker	-0.069	0.039	-0.107	0.031	-0.044	0.067	-0.009	0.055
days/91 employed last 3 years	0.125	0.014	0.128	0.019	0.061	0.021	0.085	0.034
days/91 employed last 3 years squ.	-0.006	0.001	-0.008	0.001	-0.001	0.001	-0.005	0.002
age	-0.650	0.136	0.846	0.183	-0.720	0.187	-0.571	0.327
no vocational degree	-0.104	0.021	0.072	0.028	-0.014	0.042	-0.220	0.071
no schooling degree	-0.047	0.022	-0.193	0.038	-0.200	0.046	-0.074	0.094
high school (Abitur)	-0.275	0.034	-0.039	0.036	-0.147	0.054	-0.154	0.072
health problems	-1.467	0.039	-1.280	0.051	-1.328	0.069	-1.825	0.133
at least one child	0.084	0.014	0.259	0.019	-0.050	0.020	-0.044	0.032
region with bad conditions	-0.042	0.069	0.094	0.100	-0.379	0.053	-1.021	0.087
urban region with high unempl.	-0.145	0.023	-0.060	0.032	-0.494	0.057	-1.010	0.095
unemployment rate in community	-0.001	0.002	-0.5^{-4}	0.003	-0.010	0.003	0.008	0.004
winter (Jan.-Mar.)	0.033	0.015	0.017	0.018	-0.089	0.022	-0.032	0.033
spring (Apr.-Jun.)	0.502	0.013	0.215	0.016	0.483	0.018	0.290	0.028
summer (Jul.-Sept.)	0.362	0.012	0.110	0.015	0.420	0.017	0.190	0.026
year 1999 or 2000	0.477	0.057	0.208	0.072	0.365	0.081	-0.108	0.130
year 2001	0.331	0.044	0.165	0.056	0.257	0.063	-0.075	0.101
year 2002	0.211	0.033	0.120	0.042	0.120	0.048	-0.069	0.075
year 2003	0.201	0.022	0.086	0.028	0.139	0.032	-0.029	0.049
age × E_{t-1}	-0.591	0.158	-0.994	0.212	0.023	0.230	-1.512	0.383
low skilled × E_{t-1}	-0.099	0.026	-0.354	0.037	-0.105	0.057	-0.037	0.095
high school (Abitur) × E_{t-1}	0.577	0.043	0.168	0.046	0.540	0.070	0.439	0.090
health problems × E_{t-1}	-0.076	0.054	-0.190	0.073	0.033	0.099	-0.001	0.199
share last wages censored × E_{t-1}	-0.201	0.158	-0.828	0.219	-0.925	0.225	-1.709	0.386
log last average real wage × E_{t-1}	-0.147	0.028	-0.091	0.030	-0.166	0.045	-0.158	0.061
age × $\tau_{t,E}$	-0.282	0.022	-0.055	0.024	-0.213	0.033	-0.021	0.042
low skilled × $\tau_{t,E}$	-0.003	0.004	-0.014	0.004	-0.015	0.008	-0.008	0.011
high skilled × $\tau_{t,E}$	0.032	0.005	0.012	0.005	0.025	0.008	0.016	0.010
health problems × $\tau_{t,E}$	0.043	0.007	0.015	0.008	0.038	0.011	0.066	0.016

<continued on next page>

Means and Standard deviations of Parameters from MCMC Estimation <continued>

Name	Male West Mean	SD	Female West Mean	SD	Male East Mean	SD	Female East Mean	SD
share last wages censored $\times$ $\tau_{t,E}$	-0.122	0.020	-0.051	0.025	-0.057	0.027	-0.065	0.048
log last average real wage $\times$ $\tau_{t,E}$	-0.029	0.003	-0.012	0.003	-0.021	0.006	-0.004	0.006
age $\times$ $\tau_{t,E}$ $\times E_{t-1}$	0.274	0.034	0.245	0.040	0.128	0.050	0.262	0.077
low skilled $\times$ $\tau_{t,E}$ $\times E_{t-1}$	0.027	0.006	0.043	0.007	0.054	0.013	0.009	0.020
high school (Abitur) $\times$ $\tau_{t,E}$ $\times E_{t-1}$	-0.029	0.008	-0.024	0.008	-0.021	0.013	-0.066	0.016
health problems $\times$ $\tau_{t,E}$ $\times E_{t-1}$	-0.012	0.011	0.025	0.014	-0.013	0.019	-0.064	0.042
share last wages cens. $\times$ $\tau_{t,E}$ $\times E_{t-1}$	0.126	0.032	0.140	0.044	0.118	0.047	0.355	0.091
log last av. real wage $\times$ $\tau_{t,E}$ $\times E_{t-1}$	0.028	0.006	0.020	0.006	0.020	0.009	0.032	0.012
constant	-2.667	0.141	-2.665	0.180	-1.900	0.304	-1.393	0.340
Qualification Equation								
$\tau_{t,Q}$	0.038	0.048	0.071	0.060	0.060	0.056	0.230	0.080
Q_{t-1}	0.663	0.138	0.688	0.157	0.429	0.162	-0.534	0.247
planned end missing	1.458	0.108	0.892	0.151	1.482	0.185	0.942	0.329
days/91 to pl. end if enough dur. left	1.852	0.061	2.009	0.070	2.206	0.075	2.523	0.100
days/91 to pl. end if . . . squ.	-0.175	0.007	-0.168	0.008	-0.206	0.010	-0.214	0.011
inflow quarter	0.035	0.044	0.012	0.047	-0.211	0.058	-0.178	0.079
days/91 elapsed unempl. duration	0.061	0.014	-0.036	0.016	0.092	0.019	0.046	0.021
days/91 elapsed unempl. duration sq.	-0.008	0.001	0.001	0.001	-0.010	0.001	-0.006	0.002
days inflow to end quarter if $t = 0$	-0.009	0.001	-0.009	0.001	-0.009	0.001	-0.009	0.002
repeated inflow	-0.031	0.043	0.022	0.062	-0.005	0.056	-0.057	0.088
winter (Jan.-Mar.)	0.206	0.032	0.301	0.037	0.254	0.041	0.339	0.051
spring (Apr.-Jun.)	0.134	0.032	0.176	0.036	0.194	0.040	0.375	0.048
summer (Jul.-Sept.)	0.117	0.031	0.147	0.035	0.124	0.041	0.207	0.050
year 1999 or 2000	0.559	0.108	0.590	0.124	0.706	0.156	0.415	0.187
year 2001	0.379	0.099	0.382	0.111	0.639	0.144	0.452	0.173
year 2002	0.338	0.093	0.232	0.099	0.601	0.137	0.401	0.158
year 2003	0.111	0.094	0.033	0.095	0.293	0.137	0.029	0.152
younger than 30	0.037	0.039	-0.169	0.050	0.038	0.057	-0.041	0.074
30-34 years old	-0.022	0.035	-0.154	0.042	0.001	0.049	-0.043	0.061
40-44 years old	0.021	0.037	0.083	0.044	-0.018	0.049	0.074	0.059
45-49 years old	-0.067	0.042	0.057	0.047	-0.051	0.050	-0.015	0.064

<continued on next page>

Means and Standard deviations of Parameters from MCMC Estimation <continued>

Name	Male West Mean	Male West SD	Female West Mean	Female West SD	Male East Mean	Male East SD	Female East Mean	Female East SD
50 years or more	-0.263	0.047	-0.218	0.055	-0.165	0.058	-0.079	0.068
no schooling degree	0.028	0.040	-0.263	0.074	-0.194	0.081	-0.234	0.126
high school (Abitur)	0.224	0.036	0.045	0.037	0.160	0.054	0.036	0.052
no vocational degree	-0.007	0.028	-0.039	0.036	-0.065	0.054	-0.136	0.070
last job: office or business jobs	0.183	0.046	0.368	0.039	0.038	0.077	0.128	0.046
last job: technician or related	0.164	0.047	0.029	0.048	0.086	0.069	0.159	0.062
last job: whitecollar job	0.132	0.038	0.070	0.043	0.209	0.055	0.265	0.055
last job: seasonal worker	-0.231	0.052	-0.088	0.057	-0.226	0.065	-0.133	0.067
last job: parttime worker	0.078	0.053	-0.042	0.043	0.002	0.088	0.140	0.055
log last average real wage	0.068	0.019	0.013	0.025	0.056	0.027	0.093	0.034
health problems	0.112	0.039	0.201	0.050	-0.067	0.061	0.051	0.075
at least one child	0.135	0.027	0.197	0.034	0.150	0.034	0.266	0.045
days/91 employed last 3 years	-0.008	0.022	-0.020	0.027	-0.038	0.031	-0.081	0.037
days/91 employed last 3 years squ.	0.001	0.001	0.002	0.002	0.003	0.002	0.006	0.002
entitled to unempl. compensation	0.169	0.046	0.112	0.061	0.122	0.063	0.300	0.086
unemployment rate in community	0.005	0.004	0.005	0.005	0.012	0.005	-0.002	0.006
younger than 30 $\times$ $\tau_{t,Q}$	-0.325	0.165	-0.143	0.212	-0.322	0.250	0.924	0.348
30-34 years old $\times$ $\tau_{t,Q}$	-0.066	0.146	-0.079	0.193	-0.282	0.200	0.894	0.293
40-44 years old $\times$ $\tau_{t,Q}$	0.382	0.165	0.107	0.180	0.350	0.220	0.824	0.305
45-49 years old $\times$ $\tau_{t,Q}$	0.264	0.172	-0.363	0.184	0.171	0.230	0.804	0.306
50 years or more $\times$ $\tau_{t,Q}$	0.446	0.206	-0.104	0.237	0.466	0.261	1.242	0.337
no vocat. degree $\times$ $\tau_{t,Q}$	-0.183	0.103	-0.292	0.127	-0.669	0.218	-0.142	0.268
younger than 30 $\times$ Q_{t-1}	0.138	0.079	0.041	0.093	-0.025	0.110	-0.242	0.123
30-34 years old $\times$ Q_{t-1}	0.053	0.060	0.068	0.085	0.002	0.079	-0.172	0.104
40-44 years old $\times$ Q_{t-1}	-0.197	0.074	-0.075	0.079	-0.167	0.084	-0.203	0.112
45-49 years old $\times$ Q_{t-1}	-0.033	0.072	0.099	0.077	-0.143	0.088	-0.192	0.110
50 years or more $\times$ Q_{t-1}	-0.116	0.103	-0.059	0.121	-0.229	0.097	-0.318	0.117
no vocational degree $\times$ Q_{t-1}	0.054	0.046	0.061	0.055	0.247	0.099	0.158	0.099
constant	-3.477	0.186	-3.223	0.219	-3.483	0.261	-3.424	0.306
Individual Level Variances and Covariances								
Var(α_E)	0.617	0.028	0.880	0.047	0.553	0.039	1.047	0.087

<continued on next page>

Means and Standard deviations of Parameters from MCMC Estimation <continued>

Name	Male West Mean	SD	Female West Mean	SD	Male East Mean	SD	Female East Mean	SD
$\text{Var}(\alpha_Q)$	0.286	0.065	0.451	0.089	0.356	0.073	0.292	0.072
$\text{Cov}(\alpha_E, \alpha_Q)$	-0.046	0.026	-0.003	0.042	-0.053	0.032	-0.122	0.057
$\text{Var}(\alpha_E)/(\text{Var}(\alpha_E)+1)$	0.381	0.011	0.468	0.013	0.356	0.016	0.511	0.021
$\text{Var}(\alpha_Q)/(\text{Var}(\alpha_Q)+1)$	0.221	0.039	0.308	0.042	0.261	0.039	0.224	0.042
$\text{Corr}(\alpha_E+\epsilon_{t,E}, \alpha_Q+\epsilon_{t,Q})$	-0.032	0.018	-0.002	0.025	-0.036	0.021	-0.075	0.034
$\text{Corr}(\alpha_E, \alpha_Q)$	-0.109	0.060	-0.004	0.067	-0.120	0.069	-0.220	0.096

Notes: $t = 0, \ldots, 16$ indexes the quarters since the inflow. E_t indicates the employment status and Q_t the training status in period t. $\tau_{t,E}$ and $\tau_{t,Q}$ indicate the elapsed duration in employment/unemployment and training, respectively. D_t is a dummy equal to one if a participation in training occurred during any previous quarter since the inflow. tsb denotes the time since the beginning of the program. α_E (α_Q) denotes the individual specific effect in the employment (qualification) equation, $\epsilon_{t,E}$ ($\epsilon_{t,Q}$) the idiosyncratic error term in the employment (qualification) equation.

Table 5.4: Classical ATT Aligned to Program Start

$t - t_s$	Male West Mean	SD	Female West Mean	SD	Male East Mean	SD	Female East Mean	SD
0	-0.134	0.009	-0.106	0.009	-0.139	0.010	-0.073	0.010
1	-0.141	0.012	-0.107	0.013	-0.169	0.014	-0.076	0.014
2	-0.073	0.014	-0.047	0.016	-0.116	0.015	-0.049	0.015
3	-0.002	0.014	0.040	0.018	-0.051	0.016	-0.010	0.017
4	0.046	0.015	0.115	0.020	-0.019	0.017	0.038	0.018
5	0.066	0.016	0.154	0.021	0.023	0.018	0.073	0.021
6	0.084	0.017	0.173	0.022	0.060	0.019	0.106	0.022
7	0.099	0.017	0.187	0.023	0.090	0.019	0.132	0.022
8	0.114	0.017	0.202	0.023	0.113	0.019	0.154	0.024
9	0.119	0.018	0.210	0.023	0.133	0.021	0.170	0.025

t_s denotes the quarter in which the program starts.

Table 5.5: Predicted Participation Rate of Participants if Postponing Participation (Aligned to Start of Program)

	Male West		Female West		Male East		Female East	
$t - t_s$	$\hat{Q}_t$	$\bar{Q}_t$	$\hat{Q}_t$	$\bar{Q}_t$	$\hat{Q}_t$	$\bar{Q}_t$	$\hat{Q}_t$	$\bar{Q}_t$
0	0.000	1.000	0.000	1.000	0.000	1.000	0.000	1.000
1	0.270	0.650	0.175	0.708	0.337	0.729	0.220	0.742
2	0.317	0.413	0.192	0.464	0.374	0.533	0.221	0.619
3	0.292	0.226	0.165	0.229	0.333	0.349	0.189	0.432
4	0.241	0.074	0.132	0.064	0.269	0.142	0.151	0.195
5	0.177	0.020	0.095	0.025	0.197	0.024	0.107	0.029
6	0.132	0.020	0.072	0.021	0.141	0.014	0.073	0.021
7	0.099	0.019	0.053	0.018	0.097	0.010	0.052	0.018
8	0.078	0.006	0.043	0.011	0.071	0.006	0.039	0.012
9	0.060	0.001	0.035	0.008	0.050	0.000	0.028	0.002
Total	0.687	1	0.523	1	0.767	1	0.626	1

t_s denotes the quarter in which the program starts. $\bar{Q}_t$ is the mean of the participation dummy of participants as observed in data, $\hat{Q}_t$ the mean as predicted under the waiting scenario. The row labeled "Total" gives the share of those who ever enrol into a program.

Table 5.6: ATT of Training versus Waiting Aligned to Program Start

	Male West		Female West		Male East		Female East	
$t - t_s$	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0	-0.134	0.009	-0.106	0.009	-0.139	0.010	-0.073	0.010
1	-0.141	0.012	-0.107	0.014	-0.169	0.014	-0.076	0.014
2	-0.058	0.013	-0.041	0.015	-0.090	0.015	-0.044	0.015
3	0.014	0.012	0.044	0.016	-0.020	0.014	-0.003	0.015
4	0.051	0.012	0.108	0.017	0.007	0.014	0.042	0.015
5	0.051	0.012	0.130	0.018	0.039	0.013	0.070	0.017
6	0.058	0.013	0.138	0.019	0.062	0.014	0.094	0.018
7	0.062	0.013	0.141	0.019	0.068	0.016	0.111	0.020
8	0.068	0.013	0.147	0.020	0.076	0.016	0.124	0.022
9	0.065	0.013	0.148	0.020	0.080	0.018	0.129	0.025

t_s denotes the quarter in which the program starts.

Table 5.7: Predicted Participation and Employment Rates for Different Planned Program Durations: Male, West

	One quarter		Two quarters		Three quarters		Four quarters	
$t - t_s$	$\hat{Q}_t$	$\hat{E}_t$	$\hat{Q}_t$	$\hat{E}_t$	$\hat{Q}_t$	$\hat{E}_t$	$\hat{Q}_t$	$\hat{E}_t$
0	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
1	0.399	0.079	0.793	0.079	0.904	0.079	0.918	0.079
2	0.086	0.187	0.286	0.165	0.706	0.159	0.812	0.158
3	0.024	0.246	0.061	0.244	0.246	0.232	0.617	0.229
4	0.007	0.267	0.017	0.290	0.050	0.311	0.198	0.326
5	0.003	0.296	0.006	0.329	0.015	0.365	0.043	0.387
6	0.001	0.316	0.002	0.348	0.005	0.390	0.013	0.422
7	0.001	0.323	0.001	0.358	0.002	0.399	0.005	0.435
8	0.000	0.328	0.001	0.365	0.001	0.410	0.003	0.448
9	0.000	0.344	0.000	0.383	0.001	0.429	0.001	0.467

t_s denotes the quarter in which the program starts. $\hat{Q}_t$ and $\hat{E}_t$ are the simulated means of the participation and employment probability, respectively.

Table 5.8: Predicted Participation and Employment Rates for Different Planned Program Durations: Female, West

	One quarter		Two quarters		Three quarters		Four quarters	
$t - t_s$	$\hat{Q}_t$	$\hat{E}_t$	$\hat{Q}_t$	$\hat{E}_t$	$\hat{Q}_t$	$\hat{E}_t$	$\hat{Q}_t$	$\hat{E}_t$
0	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
1	0.416	0.070	0.825	0.070	0.922	0.070	0.931	0.070
2	0.099	0.166	0.311	0.151	0.748	0.147	0.836	0.147
3	0.030	0.246	0.071	0.253	0.269	0.249	0.647	0.248
4	0.011	0.292	0.021	0.325	0.059	0.357	0.219	0.377
5	0.005	0.328	0.008	0.372	0.019	0.422	0.051	0.452
6	0.002	0.350	0.004	0.398	0.007	0.456	0.016	0.501
7	0.001	0.364	0.002	0.415	0.004	0.475	0.007	0.524
8	0.001	0.375	0.001	0.428	0.002	0.490	0.005	0.540
9	0.001	0.386	0.001	0.440	0.002	0.501	0.003	0.552

t_s denotes the quarter in which the program starts. $\hat{Q}_t$ and $\hat{E}_t$ are the simulated means of the participation and employment probability, respectively.

Table 5.9: Predicted Participation and Employment Rates for Different Planned Program Durations: Male, East

	One quarter		Two quarters		Three quarters		Four quarters	
$t - t_s$	$\hat{Q}_t$	$\hat{E}_t$	$\hat{Q}_t$	$\hat{E}_t$	$\hat{Q}_t$	$\hat{E}_t$	$\hat{Q}_t$	$\hat{E}_t$
0	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
1	0.455	0.048	0.869	0.048	0.946	0.048	0.952	0.048
2	0.096	0.127	0.333	0.117	0.798	0.115	0.873	0.115
3	0.027	0.178	0.068	0.177	0.295	0.171	0.723	0.171
4	0.009	0.214	0.020	0.227	0.062	0.229	0.264	0.215
5	0.004	0.256	0.007	0.279	0.018	0.296	0.052	0.294
6	0.001	0.281	0.003	0.308	0.006	0.336	0.013	0.353
7	0.001	0.291	0.001	0.321	0.003	0.352	0.006	0.374
8	0.000	0.308	0.001	0.341	0.001	0.374	0.003	0.398
9	0.000	0.339	0.000	0.373	0.001	0.408	0.001	0.434

t_s denotes the quarter in which the program starts. $\hat{Q}_t$ and $\hat{E}_t$ are the simulated means of the participation and employment probability, respectively.

Table 5.10: Predicted Participation and Employment Rates for Different Planned Program Durations: Female, East

	One quarter		Two quarters		Three quarters		Four quarters	
$t - t_s$	$\hat{Q}_t$	$\hat{E}_t$	$\hat{Q}_t$	$\hat{E}_t$	$\hat{Q}_t$	$\hat{E}_t$	$\hat{Q}_t$	$\hat{E}_t$
0	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
1	0.443	0.039	0.899	0.039	0.959	0.039	0.961	0.039
2	0.075	0.083	0.352	0.076	0.866	0.074	0.919	0.074
3	0.019	0.123	0.062	0.123	0.345	0.119	0.823	0.119
4	0.006	0.159	0.016	0.175	0.064	0.191	0.323	0.198
5	0.002	0.195	0.006	0.219	0.018	0.249	0.064	0.260
6	0.001	0.217	0.002	0.242	0.007	0.279	0.019	0.309
7	0.001	0.237	0.001	0.263	0.003	0.302	0.008	0.340
8	0.000	0.263	0.001	0.290	0.002	0.330	0.005	0.373
9	0.000	0.287	0.001	0.314	0.002	0.354	0.003	0.398

t_s denotes the quarter in which the program starts. $\hat{Q}_t$ and $\hat{E}_t$ are the simulated means of the participation and employment probability, respectively.

Table 5.11: ATT of Planned Program Duration of One Quarter versus Two Quarters

	Male West		Female West		Male East		Female East	
$t - t_s$	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.022	0.006	0.016	0.006	0.011	0.006	0.007	0.007
3	0.002	0.006	-0.007	0.007	0.001	0.007	-0.000	0.008
4	-0.023	0.007	-0.033	0.008	-0.013	0.007	-0.016	0.009
5	-0.033	0.007	-0.045	0.008	-0.023	0.008	-0.024	0.010
6	-0.033	0.007	-0.049	0.009	-0.027	0.008	-0.025	0.011
7	-0.034	0.007	-0.051	0.009	-0.030	0.009	-0.026	0.011
8	-0.037	0.007	-0.053	0.009	-0.033	0.009	-0.027	0.011
9	-0.039	0.008	-0.054	0.009	-0.035	0.009	-0.026	0.012

t_s denotes the quarter in which the program starts.

Table 5.12: ATT of Planned Program Duration of Three Quarters versus Two Quarters

	Male West		Female West		Male East		Female East	
$t - t_s$	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	-0.006	0.002	-0.003	0.002	-0.002	0.002	-0.001	0.002
3	-0.012	0.006	-0.004	0.007	-0.006	0.006	-0.004	0.008
4	0.021	0.007	0.031	0.009	0.001	0.007	0.016	0.009
5	0.037	0.008	0.049	0.009	0.017	0.007	0.030	0.010
6	0.042	0.008	0.058	0.010	0.028	0.008	0.037	0.011
7	0.042	0.008	0.060	0.010	0.030	0.008	0.039	0.011
8	0.044	0.008	0.061	0.010	0.033	0.009	0.040	0.011
9	0.046	0.008	0.062	0.010	0.035	0.009	0.040	0.012

t_s denotes the quarter in which the program starts.

Table 5.13: ATT of Planned Program Duration of Four Quarters versus Two Quarters

	Male West		Female West		Male East		Female East	
$t - t_s$	Mean	SD	Mean	SD	Mean	SD	Mean	SD
0	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	-0.007	0.002	-0.004	0.002	-0.002	0.002	-0.001	0.002
3	-0.015	0.008	-0.005	0.008	-0.006	0.007	-0.004	0.008
4	0.036	0.013	0.052	0.015	-0.012	0.012	0.023	0.016
5	0.059	0.012	0.080	0.015	0.014	0.012	0.041	0.016
6	0.074	0.012	0.103	0.015	0.045	0.013	0.067	0.017
7	0.078	0.013	0.109	0.015	0.052	0.013	0.077	0.018
8	0.082	0.013	0.111	0.016	0.057	0.014	0.083	0.019
9	0.084	0.013	0.112	0.016	0.060	0.014	0.084	0.020

t_s denotes the quarter in which the program starts.

List of Tables

List of Figures

Lebenslauf

Marie Elina Waller
geboren am 07. Juni 1978 in Hamburg

10/2003 - 12/2009	Promotion in Volkswirtschaftslehre an der Universität Mannheim, Mitglied des Center for Doctoral Studies in Economics (CDSE)
10/2003	Diplom in Volkswirtschaftslehre, Georg-August-Universität Göttingen
04/2001 - 08/2003	Hauptstudium in Volkswirtschaftslehre an der Georg-August-Universität Göttingen
09/2000 - 02/2001	Studium in Volkswirtschaftslehre an der Universität Poitiers, Frankreich
04/1999 - 07/2000	Grundstudium in Volkswirtschaftslehre an der Georg-August-Universität Göttingen
10/1997 - 03/1999	Grundstudium in Französischer Philologie und Betriebswirtschaftslehre an der Georg-August-Universität Göttingen
06/1997	Abitur am Gymnasium Wilhelm-Raabe-Schule, Lüneburg

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig angefertigt und die benutzten Hilfsmittel vollständig und deutlich angegeben habe. Entlehnungen aus anderen Schriften sind ausdrücklich als solche gekennzeichnet und mit Quellenangaben versehen.

Freiburg, den 01. Juli 2009

Marie Elina Waller