Master Thesis

How effective are language trainings for immigrants? Evaluating active labor market policies by applying propensity score matching

08/08/2011

Supervisor: Martin Nordin

Felizia Hanemann
Pers.nr.: 870514–7620
Email: felizia.hanemann@googlemail.com
Mobile: 076-4089194

Abstract

Integration programs are important and at the same time controversial parts of an immigration policy. This paper contributes to this debate by evaluating the effectiveness of German language trainings for immigrants with respect to their labor market performance. To allow for possible sample selection bias, the propensity score matching method is applied. This method matches treated and control persons with similar characteristics such that the assignment to each group can be considered as random placement and treatment effects can be worked out by comparing the outcomes of the corresponding matching partners. The empirical analysis relies on a new administrative dataset containing extensive employment information. Contrary to expectations about the effectiveness of a policy program, the results show significantly negative effects of language trainings, both on the employment status and on earnings. There is some evidence that the effects are more pronounced in Eastern Germany and for female participants. In accordance to previous studies, the negative effects can partly be explained by locking-in effects, which indicate that the job search intensity is usually lower during and after the participation in a program. Data limitations and arising econometric problems in the matching procedure might also cause the effects to be negative.

Contents

List of Figures

List of Tables

List of Abbreviations

ALMP	Active Labor Market Policy
AMM-O	Arbeitsmarktmonitor Ostdeutschland
	engl.: Monitoring report of the labour market in East Germany
AMM-SA	Arbeitsmarktmonitor Sachsen-Anhalt
	engl.: Monitoring report of the labour market in the state Sachsen-Anhalt
ATE	Average Treatment Effect
ATT	Average Treatment Effect on the Treated
BA	Bundesagentur für Arbeit
	engl.: German Federal Employment Agency
BAMF	Bundesamt für Migration und Flüchtlinge
	engl.: Federal Office for Migration and Refugees
CIA	Conditional Independence Assumption
DiD	Difference-in-Difference estimator
ESF	European Social Fund
FDZ	Forschungsdatenzentrum
	engl.: Research Data Centre
GDP	Gross Domestic Product
GSOEP	German Socio Economic Panel
IAB	Institut für Arbeitsmarkt- und Berufsforschung
	engl.: Institute for Employment Research
IEBS	Integrated Employment Biographies Sample
JCS	Job Creation Schemes
NN	Nearest Neighbor Matching
PSM	Propensity Score Matching
SUTVA	Stable Unit Treatment Value Assumption

1. Introduction

Several barriers hinder immigrants from a successful performance in the labor market. The problems in finding jobs mainly result from cultural differences, lack of networks and educational issues. Among these, low language skills are the most prejudicial drawback for immigrants to enter the labor market. For this reason, many European countries have recently reformed their integration policy and implemented special language trainings for immigrants. Due to the huge amount of public expenditures spent on these language trainings, it is of strong interest whether and under which circumstances these programs are effective with regard to the labor market outcomes.

The research question of this thesis is thus whether language courses are suitable instruments to improve the performance of immigrants in the labor market. In order to answer this question, this thesis contains a microeconometric evaluation of German language courses for immigrants. In accordance to previous literature, the outcome of interest is the employment status after the participation in the training. To address regional differences in the labour markets, the effects are further estimated separately for West and Eastern Germany, as well as for male and female participants. In addition, the impacts of the trainings on the earnings of the participants is measured as an indication for the quality of the job.

The fundamental challenge in a microeconomic evaluation is to identify the treatment effect of the policy instrument so that it can be proved that the employment situation of an immigrant is affected exclusively by the participation in a language training. As based on individual data, it is never possible to observe both the participation and the non-participation effect for the same individual at the same time. That means that from observing an individual who became employed after participating in a language training it cannot be concluded that the course is effective as it cannot be observed whether the person would have found a job anyway, even if not participating in the training. It is further not possible to simply compare participants with non-participants as the treatment is not assigned randomly and thus selection bias could arise. To allow for the self selection into the training, a special econometric method called Propensity Score Matching (PSM) is applied. The aim of the method is to find an adequate control group consisting of so-called statistical twins of the treated persons. In other words, for every person of the treatment group, a comparable person with the same background characteristics has to be found. The similarity in the statistical background allows to directly compare the labor market outcome of both individuals by the matching estimator. The sign and the amount of the estimator can be interpreted as the average treatment effect of the participation in the language trainings. It is important to note that the PSM is reducing but not completely eliminating the potential occurrence of selection bias as PSM only controls for observed variables and there could still be unobserved heterogeneity left leading to biased results.

The described statistical method requires a rich dataset containing many observations and information about the socioeconomic background as well as about the labor market performance. In the past, the evaluation of German active labor market policies (ALMP) was mainly based on survey data like the German Socio Economic Panel (GSOEP) or the Labor Market Monitors of East Germany and Sachsen-Anhalt. These datasets consist of relatively small numbers of observations and thus do not provide enough heterogeneity in the characteristics. In 2010 a new and rich administrative data base

has been made available by the Federal Employment Office of Germany. The so-called Integrated Employment Biographies Sample (IEBS) is a merged data set collected through different administrative processes and thus contains more reliable data and especially more observations than conventional survey data. Besides the required information on the participation in ALMP programs and details on the individual employment histories, the dataset also contains a large variety of covariates including socio-economic characteristics.

The main finding of the evaluation is that the language trainings for immigrants cannot be proved to be successful in terms of an instrument of active labour market policies. The estimated average treatment effect of the participation on the employment status show negative results between -0.072 and -0.167 depending on the region and the observed time period. That means that the chances for a successful labour market performance even decline when participating in the language training. In accordance to this, the participation also results in a lower quality of the jobs, measured in earnings. The effects are slightly more negative in Eastern Germany and for female participants. The negative effects can partly be explained by locking-in effects and partly by severe econometric problems and general drawbacks of the matching procedure.

In previous research studies, the effectiveness of active labor market policies (ALMP) has not been evaluated satisfactorily mainly due to the lack of suitable data. Since the data set was made available last year, only a few studies used the IEBS so far. To the best of my knowledge, this is the first study examining the effectiveness of language trainings, consequently my master thesis will fill this gap by performing the PSM method based on this new data set.

The paper is organized as follows. The next chapter provides an overview of the current labor market situation in Germany and a description of the different kinds of labor market policies in general. The related literature about the evaluation of active labor market policies and about the impacts of language courses is presented in chapter 3. Chapter 4 gives background information on the integration program and on the data used for the analysis. The theoretical details of the corresponding methodology as well as the practical implementation of it is explained in chapter 5. Chapter 6 reports the results and interpretations of the estimates and discusses statistical modifications. Chapter 7 concludes the preceding observations.

2. Description of German labor market situation and policies

In Germany, more than 2.89 million unemployed persons are registered in June 2011, which corresponds to a rate of unemployment of approximately 6.9%. There has been a declining trend in the last months due to the economic recovery so that the current numbers are reassuring that Germany is on the right way compared to the last years, when the unemployment rate was twice as high as today. Also compared to the other European countries, Germany is just above the European average. Active labour market policies are an important element of the functioning of the labour market. Such political interventions are necessary to effectively regulate the labor supply and demand and to ensure social security for the people. The associated efforts can be divided into passive and active labor market policies. In Germany, the purpose of passive labor market policy is to assure a livelihood in times of unemployment by transferring monetary payments to compensate for the loss of income. The amount of the transfers depends on the

duration of unemployment, on the reasons of unemployment (bankruptcy of the employer, disability to work because of illness or accident etc.) and the family circumstances.

In contrast, active labor market policy is executed without direct financial aid, but with assistance measures to permanently reintegrate unemployed individuals 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 and public training programs. Most of the measures are geared towards special target groups, such as youth, persons with disabilities or immigrants. The reason why especially immigrants are subject to a lot of programs is illustrated in Figure 1.
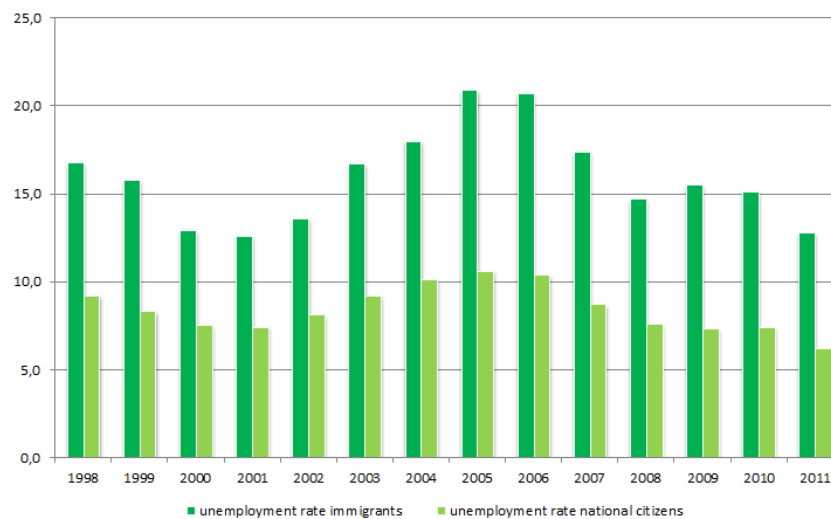
Figure 1: Unemployment rates between 1998 - 2011 separately for nationals and immigrants in Germany in percent [1]

The illustration shows that the unemployment rate for immigrants is significantly higher than the unemployment rate for German citizens in every year since 1998. Thus, special programs are needed to integrate immigrants in the native labor market and the German language trainings are said to be the most important basis for immigrants to find a job. If this is true remains to be seen in the end of the analysis. However, we can observe a tendency of the effectiveness of the language trainings in Figure 2.

The number of participants amounted to 28.000 in the year 2001 and it was almost reduplicated to 53.000 in the year 2002. The attendance stayed constant in the year 2003 and declined again in the year 2004 with approximately 40.000 participants. The other bar shows the number of former participants of the language course which are in an employment relationship six months after the training ended. This graphical analysis indicates that the effects of the language training are rather low as only 29% of the participants in the year 2001 were successful in searching for a job. In the year 2004 the achievement ratio even dropped to 13%. This decline can be explained by the overall negative situation in the labor market as the unemployment rate rose considerably between the years 2001 and 2004 as can be seen in Figure 1. However, even if we take 30% of successful participants as the basis, it is still questionable if this is a satisfying number for such a time-consuming conceived and cost-intensive program. To get an impression

---

[1]Source: Own elaboration with statistical data information from www.ec.europa.eu/eurostat
[2]Source: OECD 2005
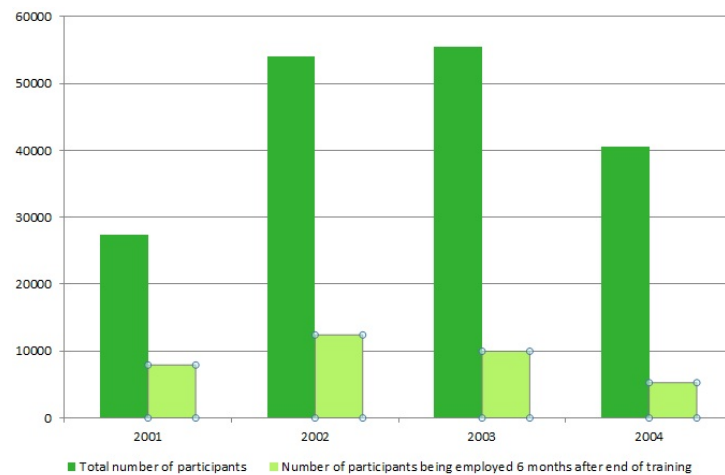
Figure 2: Total number of participants in the German language training between 2001-2004 in comparison to the number of participants being employed six months after the end of the training[2]

of the costs, Figure 3 provides an overview over the public expenditures spent on measures of ALMP and especially on educational training such as language courses or vocational trainings.
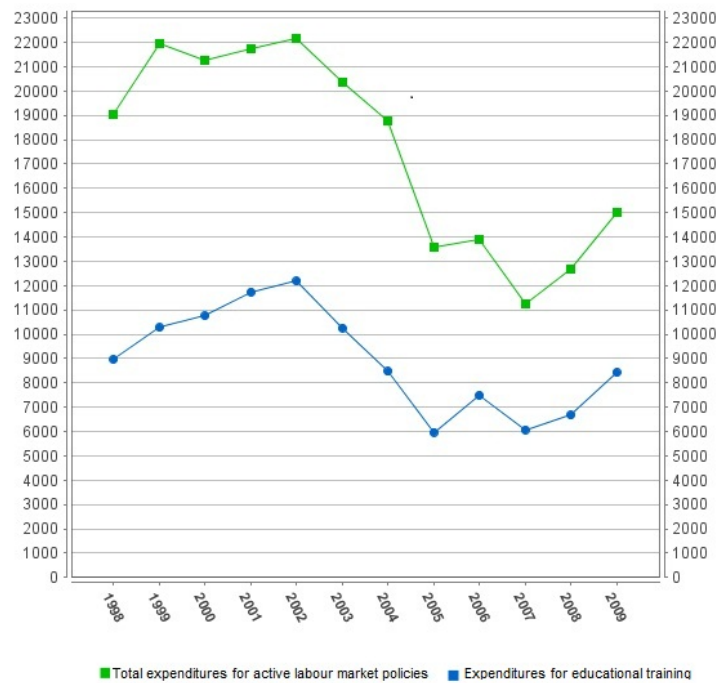
Figure 3: Total amount of expenditures for ALMP compared to expenditures for educational training in million Euro[3]

From an overall perspective, the public expenditures declined over time. The most remarkable change has been between the year 2004 and 2005 where the expenditures for all measures decreased by almost 5 billion Euro. After the year 2007, more money was spent again for labor policies so that it ended with 15 billion Euros in the year 2009 compared to the start of the record in 1998 when 19 billion Euros were spent on it. The expenditures on educational trainings form a big part of the overall expenditures, as in

---

[3]Source: Own elaboration with statistical data information from www.ec.europa.eu/eurostat

the year 2005 more than 8 billion Euro were spent, which is almost half of the total expenditures. The run of the curve concerning the expenditures on educational training is rather uninteresting as it follows the aggregated curve very closely.

It is noticeable that the curve of the expenditures on labor market policies is anti-cyclic to the run of the unemployment rate of Figure 1. In times of increasing unemployment, the amount of expenditures decrease in the subsequent period and vice versa. Two opposing reasons can explain this circumstance: First, it could be the case that high unemployment is induced by a stagnant economic climate or an economic recession such that the government budget is weakened and savings in the expenditures of the subsequent year are unavoidable. According to the changes in German GDP illustrated in Figure 4, the economic growth was moderate or even negative in the years between 2001 and 2005. Shifted by two years, the unemployment rate was the highest between 2003 and 2007, possibly as a consequence of the bad economic situation. However, it is one of the purposes of labor market policies to stabilize the economic conjuncture. As a matter of fact, the expenditures for ALMP should be even higher in times of low economic growth but at the same time there is also the possibility to balance the fluctuations by anti-cyclic governmental investments creating new jobs and consequently the unemployment rates decrease. There is another possible explanation which speaks for a reason in the other direction: If the expenditures for the ALMP are reduced, less participants can be successfully integrated into the labor market and as a consequence, the unemployment rate increases. This explanation would militate for high effectiveness of ALMP. To further determine this effectiveness, the next chapter contains a literature review of the former evaluation approaches of ALMP.
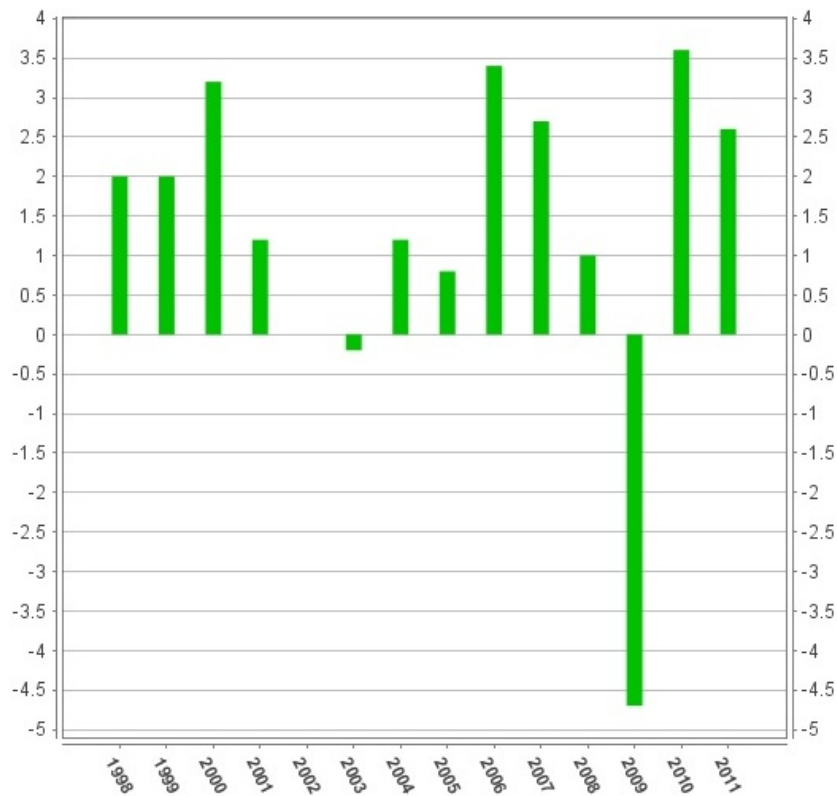

Figure 4: Yearly changes in German GDP[4]

3. Literature Review

The following literature review is divided into three parts. The first part gives an overview of general efforts and findings in the evaluation of active labor market policy, especially concerning Germany.[5] The second part introduces literature about the performance of immigrants in the labor market and the importance of language skills. The third part brings both topics together by reviewing previous and current studies about the effectiveness of language trainings for immigrants in Germany.

The first attempts to evaluate different aspects of ALMP in Germany have been made in the early 1990. During this time, researchers used to work with survey data containing only limited number of observations and time horizons because of the lack of rich administrative data sets. Most of the studies rely on the German Socio-Economic Panel, a longitudinal panel dataset of the households in Germany, or on the so called East German labor market monitor, a panel study on the population of the German state Sachsen-Anhalt. Most of these studies yield negative or insignificant results of the labor market measures. Hujer and Wellner (2000) for example evaluate the impact of professional trainings on the duration of unemployment and find only short-term positive effects in West Germany, declining in the long-term perspective. In East Germany on the contrary, there are no significant positive effects neither in the short-term nor in the long-term perspective. The authors point out that the empirical results can only be interpreted as tendencies due to few observations and insufficient information in the data. The investigation of Kraus et. al. (2000) supports these results by analyzing the effectiveness of public work programs and come to know that participants are on average worse off in finding jobs than non-participants. In contrast to these studies, Lechner and Eichler (2002) state that the participation in a program significantly reduces the unemployment rate and claim that their results are robust with respect to different specifications.

In the year 2004, first attempts were made to take advantage of administrative data. Speckesser (2004) used official and mandatory social insurance data containing information about registered employment status and active labor market programs. Speckesser applied a semi-parametric matching technique to measure the effectiveness of these programs and the empirical findings indicate that the estimated effects are in most cases negative, but insignificant. Similar studies followed by Lechner et. al. (2005), Fitzenberger and Speckesser (2007) and Fitzenberger and Völter (2007). The empirical findings depend on the extent of the program and the observed time period. All of these studies show negative effects in the short run, which are explained by the fact that the job search intensity is lower during the participation in a program ("locking-in effects"). However, in the long run, that means 3-8 years after the program started, significantly positive treatment effects on employment rates up to 10 percentage points can be noted for the majority of the programs.

The newest evaluation studies use the so-called Integrated Employment Biographies (IEB) data set. It is a special administrative data base containing large samples and providing extensive program and employment information. This data set forms the basis for the analysis in this paper, therefore further details about this data set are illustrated in chapter 3. Fitzenberger et. al. (2006) analyze the effects of three main training programs applying local linear matching based on propensity score. Similar

---

[4]Source: Own elaboration with statistical data information from www.ec.europa.eu/eurostat
[5]A more elaborated collocation of the important literature can be found in the appendix of the paper.

to the results of the previous data period, negative locking-in effects are observed in the short run and significantly positive treatment effects in the long run. Caliendo et. al. (2008) investigate job creation schemes on the same data base finding that those programs neither harm nor improve the labor market opportunities of participants. It should be noticed that the authors define the time periods as short (1-3 months), medium (6-12 months), and long (over 12 months) duration. Compared to this, Fitzenberger's definition of a long-term duration differs remarkably as he measures the treatment effect 25–31 quarters after the beginning of the treatment depending on the start date of the treatment. This different concept of short-term and long-term duration might cause the differences in the empirical findings.

To sum up the first part of the literature review, it is important to state that one severe challenge in the evaluation are the locking-in effects which seem to negatively influence all short run results. The long run results are very different depending on the data base, the evaluation method and the time horizon, so that no clear-cut conclusion can be drawn from the literature analysis in general for overall labor market programs. Furthermore, different results are found for different types of measures and for different subpopulations within a program. Lechner and Gerfin (2002) compare the effectiveness of different programs. They conclude that direct employment programs perform poorly and vocational training programs show mixed performance. Only temporary wage subsidies are proved to be successful in terms of increasing the chances of the labor market. Due to this heterogeneity in the effects of different programs, it is helpful to concentrate on the effects of a single program instead of questioning the entire bandwidth of ALMP. Thus, the following literature review is concentrated on the performance of immigrants in the labor market and the importance of language trainings.

In political discussions there is unanimity about the importance of solid language skills as necessary requirements for a successful integration of immigrants in the labor market. However, there are only few studies investigating the direct effect of language proficiency on the employment situation empirically. Problems in the econometric implementation might be a reason for this lack of empirical studies. The first severe problem is that self-reported information about the language ability in survey data may suffer substantially from measurement error. As a consequence, the variable reporting about the language ability error is correlated with the error term and the resulting estimators will be biased and inconsistent. As a second problem, the acquisition of further language skills may be endogenous with regard to the labor market outcome. This means that there is unobserved heterogeneity because some variables are omitted from the model which are positively correlated with both the language acquisition and the employment status of an individual. For instance it is conceivable that individuals who have the ability to easily learn a language might also have the ability to quickly familiarize with the requirements of a new job. This ability leads to better language proficiency and to greater chances of employment or higher earnings. Typical omitted variables are the overall learning ability, motivation, intelligence and other latent characteristics that are usually hard to measure. Leaving those variables out of the model again leads to a correlation between the language ability and the error term resulting in systematically biased and inconsistent estimators. These endogeneity problems challenge most research studies about the impact of language proficiency on earnings, but there are ways to overcome these problems as the following studies show.

The earliest studies have already been done in the early 80s for Hispanic Immigrants in the United States. Grenier (1984) analyzes wage equations including language attributes among the explanatory variables drawn from the Survey of Income and Education. Those language attributes have a significant effect on wages. Papers written by Rivera-Batiz in 1990 and by Chiswick in 1991 confirm this result by finding that English proficiency is a major factor determining the wages of immigrants in the USA. A more recent study about the effect of language acquisition on the immigrant's earnings was conducted by Berman et al (2003). The data originates from Soviet immigrants living in Israel and learning the Hebrew language. This study is of interest as it finds that the effect of language skills on earnings significantly depends on the occupation level. For immigrants being employed as IT specialists, 2/3 of the wage gap between natives and immigrants can be explained by improved language skills. For low skilled and low wage immigrants, there is no significant return to Hebrew proficiency.

Concerning Europe, a famous study was conducted by Dustmann and Fabbri (2003) using survey data from the UK. They analyze the impact of language proficiency on the employment status in two steps. First, they investigate the factors influencing the acquisition of language skills such as education, age and years of residence in the host country. In a second step, the authors analyze the extent to which language skills influence the labor market situation of immigrants. After implementing a reasonable identifying strategy to take into account the potential bias resulting from measurement error and endogenous selection, the authors come to the conclusion that "the best estimates suggest that fluency in English increases employment probabilities by about 22 percent points" (Dustman and Fabbri (2003), p. 696). Furthermore, by increasing language proficiency, earnings on average rise about 18-20%.

In their analysis of adequate measures of ALMP, Lechner and Gerfin (2002) conclude that language courses perform poorly. However, the authors stress that especially when evaluating language courses, the analysis might suffer from bias by some remaining selection on unobservables. Dustmann and van Soest (2001) put special focus on the possible biases through unobserved heterogeneity and measurement errors, which would lead to over- and underestimated results. They conduct an empirical study based on the GSOEP and find that the underestimating measurement error is larger than the overestimating unobservable error so that the results in general tend to be underestimated and thus they suspect language proficiency to be even more important than the existing literature has shown so far.

Further empirical evidence is given by politically motivated studies accompanying current projects rather than by strict scientific literature. Such a report has been done by Deeke (2006) for special language courses financed by the European Social Fund (ESF). The report indicates pessimistic valuations of the measure as still only 15% of the participants are in regular employment six months after the end of the language course. The results are only interpretable with reservations as they do not apply special econometric methods to face the evaluation problem of not being able to compare the outcomes of the participants to nonparticipants. The author argues that even if the language courses have a positive effect on the employment chances of immigrants, the proficiency is a necessary, but not sufficient precondition for a job as it cannot solve the ethnic discrimination problem occurring on the local labor markets.

A general study about the labor market situation of immigrants in Germany by the OECD (2005) states that a proper impact of language skills on the employment situation has not been proven without ambiguity as all results of previous studies are not very robust. The authors call for an extensive evaluation

system of the language courses to find out about the real effectiveness of the language courses not only with respect to the labor market outcomes like employment and wages, but also regarding the transfer of human capital and the overall integration of immigrants. Only then can be ascertained whether the language courses are efficient measures and are worth the political attendance and the financial expenditure.

4. Institutional Background and Data

4.1. Language courses as part of Active Labor Market Policy

Language capital is an important component of host country human capital. Immigrants without language skills have a decisive disadvantage in finding jobs when they compete against native applicants or other immigrants with better language skills. To avoid labor market discrimination based on the language skills and to facilitate the integration process, most countries provide special language courses as part of the active labor market policy.

The granting of German language courses has been regulated by the legislation of the Social Security Code III between the years 1998 and 2004. According to the law, ethnic German repatriates (according to §4 of Federal Displaced Persons Act) and their descendants (according to § 7 of Federal Displaced Persons Act) are entitled to a language training, if they have been employed for at least 70 days of the last year in their country of origin and if they are unemployed in Germany. The costs for the language trainings are reimbursed for 6 months at maximum with a weekly stint of at least 35 hours. Depending on the different institutions, the costs of the language training vary between €2,50 - €3 per hour, thus a total amount of around €2500 per participant has to be reimbursed by the Federal Ministry of Labor and Social Affairs for pure language trainings. Additionally, there is the possibility to get reimbursed for child care, teaching material and transportation costs. With a volume of 120 Million Euros[6] in the year 2000, the funding of language courses according to the Social Security Code has been the most extensive instrument for language trainings, which makes the measure interesting to evaluate.

There are strict rules concerning the compulsory attendance and absent times during the course. The costs are only reimbursed if the course is completed successfully by taking the final exam, which can be repeated once. This prospect of being reimbursed motivates the participants to pass through the whole training. This is important for the analysis because it assures that there are enough participants having attended the whole training and successfully learned the German language. As the exact number of days participated in the language course is given in the data, it is possible to measure the effect of the language trainings on the labor market outcomes without biases due to early course drop outs, which are excluded from the treatment group. From an econometric point of view, another advantage of the legal regulations is that the language training has priority over all other measures. That means that it is not possible to take part in any other measures from the active labor market policy before taking a language course if necessary. This regulatory fact can be exploited in the analysis as it excludes biases that would result if other measures despite from the language courses are taken at the same time. The simultaneous influence on the labor market success can at least be excluded in the short term analysis while the language training is still going on. In the observation period of 12 months in contrast, it cannot be guaranteed that the participants did not attend further courses after the language training which

---

[6]source: Krekeler, C. (2001), p. 10

might have a positive impact on the employment situation.

As a part of the reformation of the Immigration Act in the year 2005, the integration trainings have been restructured and extended to a wider group of persons being entitled to take part in language trainings. In the course of this, an additional program was implemented by the Federal Office for Migration and Refugees (BAMF-ESF-Program). Within this program, persons not having enough knowledge of the German language are eligible for funding of language courses irrespective of the duration of residence permit, of the citizenship and of their employment status in their origin country. The language courses put a special focus on professional purposes by combining language tuition, professional qualifications and practical exercises. Between 2000 and 2006 around 20 billion Euros have been provided by national funds and by the European Social Fund to finance the program. In accordance to the new funding period for EU structural funds between 2007-2013, the BAMF-ESF was extended by six years.

There are currently research projects implemented by the Centre for European Research in Mannheim to conduct an official evaluation of the BAMF-ESF-Program.[7] Yet, there is not enough data available and the evaluation will not bring satisfying results until the year 2013. Due to this fact and because of the advantages given by the regulatory conditions stated above, this study concentrates on the evaluation of the language courses according to the Social Security Code III in the time between 1998 and 2004. To avoid a mixed effect of different trainings, those individuals taking the training course sponsored by BAMF-ESF are excluded from the sample. The remaining data used for the analysis is presented in the next section.

4.2. Integrated Employment Biographies Sample (IEBS)

"This study uses the factually anonymous Integrated Employment Biographies Sample (IEBS) (Years 1993 - 2004)."[8] Besides the attractive fact that this dataset has been published only recently, the IEBS is a remarkable rich administrative data set containing information about employment biographies of around 1,4 million individuals. As the IEBS contains personal data, the application is subject to strong data protection regulations according to the German Social Security Code §75. The permission to use the full sample of the IEBS with the original data requires an extensive application and a binding contract with the Federal Ministry of Labor and Social Affairs. Due to this application process it can take several months until the data is made available. For this reason, the full data set is only suitable for long-running scientific projects. For less sophisticated and short-run projects there is another data set available called Scientific Use File[9]. It is a merged random sample of the full data set with information being effectively anonymized. This data set forms the basis for the following analysis and the "data access was provided via a Scientific Use File supplied by the Research Data Centre (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB)."[7]

The personal data is collected through different processes coming from the following sources:

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.

---

[7]See http://www.zew.de/en/forschung/projekte.php3?action=detail&nr=1045&abt=am
[8]Cited according to the citation guidelines of the Research Data Centre (FDZ), see
  http : //doku.iab.de/fdz/access/Zitierweisen_e.PDF
[9](IEBS-SUF V1) Version 1.0

The main feature of this data is detailed daily information on the employment status of each recorded individual.

The Benefit Recipient History: Includes daily spells of all unemployment benefit, unemployment assistance and subsistence allowance payments between January 1990 and June 2004. It also contains information on personal characteristics. In particular, the Benefit Recipient History includes information about the exact start and end dates of periods of transfer receipt.

Applicants Pool Base: Contains data on individuals searching for jobs in the period January 1997 to June 2004. Additional information about the labor market status and personal characteristics of an individual is given such as educational qualifications, nationality, and marital status, regional and local information.

Participants-in-Measures data: Contains detailed information on participation in public sector sponsored labor market programs covering the period January 2000 to July 2004.

For each record in any of the four sources, a spell was created containing the according information. There is information about the beginning and ending dates for each spell as well as information about the source the information comes from. The IEBS contains daily information on unemployment benefits, job search and employment details, for example daily remuneration, branch of industry etc.. Furthermore, a large variety of covariates including socio-economic characteristics such as gender, year of birth, education, nationality, place of residence, place of work, regional type is provided in the dataset. Most importantly, a large number of measures of active labor market programs are described in the data set. The information about those programs is very detailed, which makes it possible to account for program heterogeneity in an uniquely detailed way. In particular, the documentation allows an economic evaluation of job-creation measures, integration subsidies, long-term trainings, business start-up allowances and of German language courses.

Despite all advantages of the database, there are also several drawbacks which have to be considered in the analysis and interpretation of the results. The first disadvantage is that there is no information about direct program costs available in the data, therefore an extensive cost-benefit analysis with calculating net effects of the measures is not possible. Furthermore, although the documentation of employment and benefit recipient history already started in 1993, the participation in measures of ALMP and job seeking has not been recorded until the year 2000, thus the time period for the evaluation of ALMP is rather short. Moreover, missing direct information about registered unemployment constrains the data set as only information about the receipt of unemployment compensation from the German federal labor office is observed. Until 2004 these were unemployment benefits, unemployment assistance or income maintenance during further training. For this reason Fitzenberger and Wilke (2004) introduce different definitions with lower and upper bounds which can be used as proxies for the unemployment status. Despite these disadvantages, the data set is still the best available basis for microeconomic evaluation studies of ALMP concerning the data quality and extensiveness.

5. Methodology

5.1. Theoretical background

All microeconometric evaluation approaches have one aim: To find out whether a treatment changes the outcome situation of a person and if a special program is effective in accomplishing its objectives. In the following, the effect α_{it} of the treatment on an individual i at time t is defined as the difference between the potential outcome Y_{it}^T after participating in a language training and the potential outcome Y_{it}^C without participating. Thus, T refers to the treatment and C to the control group. This difference can be formalized by:

$$\alpha_{it} = Y_{it}^T - Y_{it}^C \tag{1}$$

Very often, an evaluation is dedicated to estimate the mean effect that a program would have on an individual drawn randomly from the population. This is called the Average Treatment Effect (ATE) and can be expressed by:

$$E(\alpha_t) = E(Y_t^T - Y_t^C) = E(Y_t^T) - E(Y_t^C) \tag{2}$$

where $E(.)$ stands for the expected value or the average.

If policy-makers have to decide about the expansion of a program to a wider population or to another target group, the parameter of interest is the ATE. For other research designs, it can be of interest to only consider the impact of those who actually participated in the program. An instance could be if policy-makers have to decide whether to abandon a program permanently or not. Then the Average Treatment Effect for the Treated (ATT) is the adequate estimation:

$$E(\alpha_t|D=1) = E(Y_t^T - Y_t^C|D=1) = E(Y_t^T|D=1) - E(Y_t^C|D=1) \tag{3}$$

where D is a dummy variable indicating treatment ($D=1$) or no treatment ($D=0$).

The problem of this research design is that one cannot observe how persons of the control group would perform in the labor market if they had passed through the treatment. Thus, the second part of the term $E(Y_t^C|D=1)$ is only hypothetical and cannot be estimated directly. A substitute for this unobservable situation has to be found, for which the following equation holds:

$$E(Y_t^C|D=1) - E(Y_t^C|D=0) = 0 \tag{4}$$

This equation expresses that the expected value of the mean outcome must be identical for two both groups if they share the same distribution of observable or unobservable characteristics. Finding this substitute which allows to estimate the ATT is known as the Fundamental Problem of Causal Inference which is described in more detail in the next section.

Fundamental Problem of Causal Inference

The most intuitive and simple approach to evaluate the effectiveness of a program is to simply compare the mean differences in the labor market outcomes of participants and nonparticipants. However, this approach is only suitable if the participants of the language training had been assigned to the treatment

group randomly.

Why is random assignment to the treatment and control group of such a great importance? If participants are not assigned randomly, one faces the risk to suffer from selection bias. Selection bias describes a systematic error which occurs if some individuals are more likely to be included into the sample than others due to some observable or unobservable characteristics. In terms of our evaluation the selection problem describes the fact that those with better labor market characteristics are more likely to participate in the training program and at the same time have higher chances to find a job. Not taking this selection bias into account would most probably lead to an over-estimation of the effectiveness of the program. According to Heckman (1979, p. 153), this bias in practice arises among other reasons because of self selection by the individuals into the treatment group. In technical terms selection biases arise if individuals are compared without having identical distributions of all characteristics, thus if equation (4) does not hold and $E(Y_t^C|D=1) - E(Y_t^C|D=0) \neq 0$. For this reason, the random assignment is indispensable for a credible evaluation study.

Random assignment to treatment is ex ante only given in experimental studies. Choosing the individuals who should be treated randomly through a lottery ensures that the potential outcome is independent of the treatment status, thus $(Y^T, Y^C) \amalg D$. With this random assignment, the treatment and control group are formed such that all characteristics are equally distributed. Therefore the groups can be considered as identical except for the treatment status: $E(Y^C|D=1) = E(Y^C|D=0)$. Experimental research designs are often viewed as the most robust evaluation approach. Unfortunately, those experiments are hard to conduct for complex policy programs, which forces most research designs to rely on non-experimental data.

Using non-experimental data however directly leads to the next challenge, called the Fundamental Problem of Causal Inference. This term stands for the fact that "it is impossible to *observe* the value of $Y_t(u)$ and $Y_c(u)$ on the same unit u and, therefore, it is impossible to *observe* the effect of t on u." (Holland 1998, p.947) In terms of our study this means that it is impossible to observe the labor market outcome for one individual u receiving the training $Y_t(u)$ and at the same time not receiving the training $Y_c(u)$ and, therefore, it is impossible to observe the effect of the treatment t on the individual u.

It is only possible to separately observe the labor market outcomes for individuals receiving the training and for those not receiving the training. These observable outcomes are called the factual outcomes. To truly learn about the actual effect of the training we need to compare this factual outcome to the so called counterfactual outcome, which is the outcome that would have resulted if the person had participated or respectively not participated in the program. The unobservability of the counterfactual outcome is a basic evaluation problem, which every evaluation study using non-experimental data has to overcome.

Group	Y^t	Y^c
Treatment (D=1)	observable	(counterfactual)
Treatment (D=0)	(counterfactual)	observable

There are different evaluation approaches[10], which all aim to obtain a credible estimate of the counterfactual and to use this to identify the program effect. This study is based on the Propensity Score

---

[10]For example: Difference-in-Difference estimator, Instrumental variables, Heckman selection estimator etc.

Matching (PSM) approach. The technique of matching was developed in the 1980s by Rosenbaum and Rubin, but it has been established as a useful tool for labor market policy evaluation only in the late 1990s. The intuition behind this method is easy: For every individual in the treatment group a statistical twin has to be found in the control group possessing the same or at least similar characteristics as the treated individual, such that the sample can be considered as randomly selected. The basic issues and technical details of the matching method are explained in the following.

Identifying assumptions

For all non-experimental techniques, assumptions have to be made to identify the causal effect of a policy or program on the outcome of interest. The method of matching is based on two 'identifying assumptions'.

A1) First, the so called Conditional Independence Assumption (CIA) claims that the outcome of both treated and untreated individuals must be the same if one controls for observable differences in characteristics. This can be expressed formally by:

$$(Y^t, Y^c) \amalg D|X \tag{5}$$

As II is a sign for independence, the assumption can be phrased as follows: The potential outcomes of the treated and untreated individuals are independent of the treatment status conditional on all relevant pre-treatment covariates X. This implies that selection is only based on observables and all covariates X both influencing the treatment status and the outcome are considered in the model. In other words, after controlling for X, the treatment is effectively assigned randomly and thus, the CIA assures the compliance of equation (4) and the avoidance of selection bias.

A2) The second assumption is known as the Common Support Condition stating that the treatment status must not be perfectly predictable, such that there is still variation between participants on non-participants. Individuals with the same characteristics must have a positive probability to be in the treatment as well as in the control group. In technical terms, this condition states that the probability to be treated lies between zero and one, conditional on each value of X, thus:

$$0 < P(D = 1|X) < 1 \tag{6}$$

At the same time, the converse probability of not being treated must lie between zero and one as well, such that the proportion of participants and non-participants must be positive for each value of X. This condition ensures that there is sufficient overlap between the covariates of the treated and untreated and that matching on these covariates is possible.

A3) Some studies also mention the Stable Unit Treatment Value Assumption (SUTVA) (Rubin 1978) as a requirement for the estimation of the ATT, which says that the impact of the program on one person does not depend on anybody else participating in the program. Formally, this means that:

$$(Y_i^t, Y_i^c) \amalg D_j \tag{7}$$

SUTVA implies that the potential outcome Y for individual i is independent of the treatment status D of any other individual j in the dataset.

The fulfillment of the assumptions A1 and A2 is called Strongly Ignorable Treatment Assignment (SITA) by Rosenbaum and Rubin (1983, p.43) as it ensures the necessary randomness of the sample. Under these conditions, Rosenbaum and Rubin (1983) suggest to use balancing scores $b(X)$, that is a function of the relevant covariates X such that the distribution of X given $b(X)$ is independent of assignment into treatment. This method solves the so called dimensionality problem, which means that it is nearly impossible to match all covariates of a high dimensional vector X separately as the number of matching cells increases exponentially, making exact covariate matching impossible. The use of balancing scores reduces the matching index to one dimension and simplifies the analysis.

One possible balancing score is the propensity score $p(x) = Pr[D = 1|X]$, that is the probability of participating in a program conditional on observed characteristics X such as demographic and labor market factors. Hence, instead of finding a matching partner for every single covariate, it is sufficient to compare whether the probability of participating in the training is similar for two individuals. Put it the other way around, having two individuals with the same propensity score ensures that the distribution of their covariates X is identical. Therefore, it holds that $E(Y^C|D = 1) = E(Y^C|D = 0)$, what means that those two individuals are identical except for their treatment status and they can be considered as randomly assigned to each group.

The propensity score can be estimated by a discrete choice model. Usually, the treatment variable is dichotomous so that we can leave multinomial models disregarded. Linear probability models are also not considered because of the known drawbacks that can lead to predictions lying below 0 or above 1, which makes no sense for probabilities. Therefore, the choice has to be made between a logistic or a probit regression model. Both models yield very similar results so that the decision is not too crucial, but usually the probit model is preferred. The first step in the calculation is to define regression equations with the outcome as dependent variable:

$$
\begin{aligned}
Y_{it}^T &= g_t^T(X_i) + u_{it}^T \\
Y_{it}^C &= g_t^C(X_i) + u_{it}^C
\end{aligned}
\tag{8}
$$

where the functions $g(.)$ describe the relationship between the potential outcome and the set of covariates and u denote the error terms. The latent factors determining the participation in the program can be described by the following index function:

$$
I_i = \beta X_i + \epsilon_i \qquad \text{with } \epsilon_i \sim N(o, \sigma^2)
$$

The participation in a program can then be observed by :

$$
D_i \begin{cases} 1 & if \quad I_i > 0 \\ 0 & if \quad I_i < 0 \end{cases}
\tag{9}
$$

Under this condition, the propensity score can be estimated by the probability of entering the treatment conditional on the characteristics X:

$$
Pr(Y = 1|X) = \Phi(X'\beta)
\tag{10}
$$

where Φ is the Cumulative Distribution Function (CDF) of the standard normal distribution. β is usually estimated by maximum likelihood.

Matching procedure

After having calculated the propensity score, the next step is to match a participant of the language training to a non-participant with the similar propensity score. There are different types of matching algorithms and all matching estimators contrast the outcome of a treated individual with the outcome of comparison group members.

Nearest Neighbor Matching: This matching algorithm is used most often. The statistical twins for the treated individuals are chosen from the control group according to the closest propensity score. One can decide whether an untreated individual can be used as a comparison more than once ('with replacement') or only once ('without replacement'). Matching with replacement leads to a trade-off between bias and variance, thus it must be decided from case to case which is the appropriate matching type depending on the richness of the dataset.

Caliper and Radius Matching: This matching algorithm is based on the nearest neighbor matching. Additionally, it imposes a tolerance level on how big the difference may be between the propensity score of the treated and untreated individual. By doing this, the risk of bad matches is avoided, but at the same time fewer matches can be performed, so it again depends on the richness of the dataset whether this algorithm is applicable.

Stratification and interval matching: A set of interval is used to divide the common support of propensity score, then the treatment and control cases are matched within each interval. The advantage of the stratification method over the nearest neighbor method is that the estimation is based on many more observations and thus provides better matches. On the other hand again, a rich data set is needed.

Kernel and Local Linear Matching: This approach uses weighted averages of all individuals in the control group to estimate the counterfactual outcome. The advantage of this approach in contrary to the other three above is that it uses more information and thus lowers the variance. At the same time, there is the risk of bad matches being used, so that there is again a trade-off between bias and variance.

The decision between the different types of matching algorithms has to be made based on the richness of the data. In general, one has to consider the trade-off between bias and variance. The more information is used to construct the counterfactual for the treated individual, the lower is the variance and therefore the matching quality increases. At the same time, possible bias increase resulting from poorer matches. Another decisive factor in choosing the matching algorithm is the handling of the common support condition. Implementing this condition ensures that the analysis is restricted to those treated which propensity score can be observed in the control group, too. Thus, all treatment observations whose propensity score is higher than the maximum or lower than the minimum propensity score of the control group have to be dropped. The importance of the common support condition is different for the matching algorithms, it is for example more important when matching with replacement than without replacement and it is of major importance when performing Kernel matching, because more observations are used to

construct the counterfactual.

The common support condition is one reason why matching is preferred to standard regression methods. Due to this condition only treatment effects within the common support can be estimated. This requirement is quite restrictive and creates robustness in contrast to traditional methods where estimates can even be produced in the absence of comparison results as the functional form fills in for missing data and extrapolates outside the common support. The functional form of regression methods indicates the second advantage of the matching procedure, where no assumptions of the outcome equation are needed as it is non-parametric. Regression methods in contrast need the imposition of assumptions about the functional form which cannot always be justified and incorrect functional forms may lead to biased results. Matching avoids potential misspecifications. The third reason in favor of the matching method is that one can even match on variables that are correlated with the error term of the outcome equation, because matching only assumes that the mean of the error term is the same for treatment and control group.
On the other hand, there are of course challenges and difficulties with the matching method as well. As the selection process is based on the observables, matching is only as good as the quality of the covariates. If there are unobserved variables, that simultaneously influence the assignment to treatment and the outcome variable, the matching outcome will still suffer from selection bias. Furthermore, a lot of information is needed to fulfill the common support condition, thus the method is very data hungry.

To sum up, the role of matching is to balance the distribution of relevant pre-treatment characteristics between the treatment and control group in non-experimental datasets. The aim is to minimize selection biases and to allow for heterogeneous treatment effects. The major steps of implementing the PSM are:

1. Estimation of the propensity score on the basis of a logit or probit model.

2. Choice of adequate matching algorithm.

3. Estimation of the treatment effects on the basis of the matched sample.

4. Check common support condition.

5. Sensitivity analysis.

These steps will be followed in the next section about the practical implementation of propensity score matching in order to evaluate the impacts of the language trainings on immigrants.

5.2. Practical implementation

The practical implementation of the methodology described in the previous section is simplified by the usage of two special Stata[11] packages calculating the propensity score and matching the individuals according the defined variables. The first available package named "pscore" is written by Becker and Ichino (2005) and the other one with the command "psmatch2" has been elaborated by Leuven and Sianesi (2003). As the latter one has been updated more recently and is often seen as the more flexible package, I decided to use psmatch2.

5.2.1. Model choice

The first step in calculating the propensity score is to define the treatment and control group and the relevant outcome variable. The population of interest in this study is defined by those individuals who receive unemployment benefits or unemployment assistance and who are eligible for participation in German language trainings according to the legal regulations stated in chapter 3. Given this definition of the treatment group, only immigrants can be considered to be in the treatment group. This self-evident sounding requirement caused the first challenge with the given dataset. Unfortunately, only the current citizenship is given, so that for example migrants in the second generation cannot be identified. Acquiring the German citizenship is only possible on request, there is no automatic naturalization after a certain time of residence. Several requirements have to be met before becoming a German citizen, as for example a valid residence permit, living permanently in Germany for eight years, no conviction, knowledge and conformity of the German constitution. As last requirement one has to prove sufficient language skills, for example by handing in a certificate of the successful attendance in a language course. For this reason, there are individuals in the data set having participated in the language training but indicating German citizenship. Leaving these individuals out of the analysis would lead to severe biases as these are probably the ones being most sophisticated and motivated to integrate not only in society but also in the labor market. However, as only the current citizenship is provided and not the country of origin, it is not possible to find a proper control group for these migrants when matching on the basis of nationality because other individuals with German citizenship but without migration background must of course be excluded from the analysis. Nevertheless, I decided to include those participants later acquiring the German citizenship and as a consequence not to match on the basis of the country of origin. By doing this, I accept possible bias arising from unobserved heterogeneity in the variable of the country of origin, such as ethnic labor market discrimination or cultural differences. Despite the fact that these biases are regarded to be smaller than the ones stemming from unobserved successful integrations, the decision to include the naturalized immigrants is also made for practical reasons. The total number of treated persons used in the analysis amounts to 904, as will be seen later in the descriptive statistics. More than 330 individuals attended the language course and already acquired the German citizenship. Excluding those from the analysis would lead to a loss of observations of about one third. Thus, including those individuals showing the german citizenship does not only prevents from severe biases, but it also increases the number of treated persons and thus leads to a more comprehensive sample.

   Another drawback is that the citizenship is no information which must be given in order to apply for social benefits, but the employers give this information for statistical reasons. As it is not strict administrative information, the stated citizenship could be incorrect or inconsistent. This problem can

---

[11]Data Analysis and Statistical Software, Version 11 is used in the analysis

partly be solved by using further information of an additional data set according to the instructions in the methodology report from the IAB.[12] I decided not to work on this basis as most of the heuristics presented in the report rest on probabilities and therefore one has to consider the trade-off between inconsistencies and the risk of adding wrong information.

Concentrating on the main body of the active labor force, it is necessary to exclude unemployed who were trainees, home workers, apprentices, or without previous employment. More precisely, this study focuses 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, following the example of Fitzenberger et. al. (2007). This is to avoid that the subpopulation consists of individuals who endogenously registered as unemployed. Thus, by concentrating on the inflow sample into unemployment, we focus on individuals who have already been attached to the labor market. The information from the employment period is necessary to construct the control group based on labor market relevant variables. Furthermore, the change from unemployed to employed status defines a natural time scale to align treated and non-treated individuals. To sum up, the treatment group is defined as those individuals recorded as participants of the German language training irrespective of the citizenship, having been continuously employed for at least three months in the past and being unemployed at the state of starting the language course without taking other measures of ALMP.

How should the control group look like? This should now be easy to answer after the detailed description of the treatment group. The control group consists of non-german citizens who became unemployed after being continuously employed for at least three months and who do not participate in a German language training. Unfortunately the last point is problematic again. So far, only those individuals who successfully finished the language course are considered in the treatment group. These can be identified because only those individuals are reimbursed for the costs of the language course and thus they are recorded as receiving those benefits. However, there are individuals who dropped the language course during the treatment and consequently the control group is a mix of drop offs and nonparticipants. Depending on how far an individual proceeded in a training, the participation might have an influence on the employment status and thus, this individual must not serve as a control observation. As the exact number of days of the language training is given in the data, it is possible to assign the drop offs to a group according to their total duration of training. Following algorithm is applied:

- If the language course lasts less than 2 weeks: Keep in control group as this course is unlikely to have much influence on the job search and because the data does not allow that too many observations are dropped.

- If the language course lasts between 2 weeks and 3 months: Drop observations as the course might have an influence on the job search, but the participation is too short to include it into the evaluation.

- If the language course lasts more than 3 months: These individuals can be considered as being treated as a three months full-time language training should have influences on the employment probability of an individual.

---

[12]$http://fdz.iab.de/188/section.aspx/Publikation/k070921f03$

This procedure takes the different levels of knowledge into account at best possible to avoid that all those individuals who started a language training have to be excluded from the analysis which would have led to a severe loss of variance in the control group. Still, there is one problem: The language courses started in 1998, but the data provides information only beginning from the year 2000. Consequently, it could be the case that individuals took the language course before the year 2000 and as a result the control group might include some individuals who unobserved passed through the whole treatment. There is no way to identify those individuals and therefore there is no alternative but to consider this circumstance when interpreting the estimation results.

The relevant outcome variable for the treatment and the control group is the employment status after the treatment or non-treatment. To demonstrate short-term as well as medium-term effects of the language courses, we control the employment status after 3 months, after 6 months and after 12 months of the start of the language course. The decision to use the start date and not the end date of the language course as the indicating start date of the measure depends on several reasons. First of all the end dates are defective according to Bender et. al. (2005). It is stated that the end dates of training programs suffer from measurement error that leads to biases. Furthermore if we would use the end dates as the starting dates of the measure, we would face an endogeneity problem, because motivated participants complete the course earlier and increase the likelihood of employment. At the same time, the end dates also imply the opposite effect for less motivated, less successful participants. The so called locking-in effects refer to the period a person participates in a measure and describe the fact that the job search intensity may be lower during the language training. For less successful participants, this period lasts longer and they are locked in the measure longer time, maybe resting on the feeling that they are productive at the moment and there is no need to look for jobs. Thus, the job finding rate is expected to be low during the participation. Because of these explanations, we use the start dates of the language trainings as the start date of the measure, from which on the time until employment is counted for participants.

For non-participants we have to simulate start dates as we do not have a natural start date of a measure. What are the requirements for the simulated start dates for non-participants? First of all, the period of active job searching should be similar in time for matched participants and non-participants to capture potential influences of business cycles and seasonal effects. Therefore the non-participants are required to remain unemployed until the program of the matched participant ends as we assume the participant to actively search for a job after finishing the measure due to the locking-in effects. To simulate the fictitious starting dates of the non-participants, I use an approach suggested by Lechner (1999). For this purpose, the start dates of participants are regressed on a set of time invariant personal and regional characteristics. The estimated coefficients plus a draw in the residual distribution are used to predict start dates for nonparticipants. Persons with simulated starting dates later than their exit date from unemployment or not in the same year as the starting date of their matched participant are excluded from the control group. This method aims to include both the seasonal consistency as well as the personal conformance such as labor market experience which could influence the duration of unemployment.

5.2.2. Variable choice

After having defined the treatment group, the control group and the outcome variable, the supporting covariates have to be elaborated. It is difficult to find all those observable characteristics that affect both the program participation as well as the outcome. The best a researcher can do is to consider economic theory and prior research concerning the program participation in order to find the best constellation of variables. Taking this into account, the following groups of variables are used for the calculation of the propensity score:

Personal characteristics: To control for background information which could possibly influence the labor market performance of an individual, we consider age, gender, education and highest graduation, marital status and the number of children.

Influences on employment situation: We consider the current disability and health status and past health problems, especially whether they constrain the individual in finding and performing a job. The records of any transfer payments such as unemployment benefits, unemployment assistance etc. are regarded. Furthermore we have to consider whether the individuals are demanding further measures of active labor market policy despite the language course, which makes it hard to evaluate which measure the success in finding a job is based on in the long run.

Regional information: The data set offers many variables to identify local labor markets. It is possible to differentiate between place of residence and place of work, between the different federal states of Germany, between different cities and regions within the federal states and furthermore information about the job market situation of the region is given.

Labor market characteristics: In order to be able to consider the history of labor market performance, we use information on the type of occupation and industry of the jobs before unemployment, if available the reason why the employment relationship ended, the total length of employment and unemployment, daily wages and whether the last job was a full-time or a part-time job.

A more detailed listing of the used variables can be found in Table 1.

As most of the variables are not ordinal, it is necessary to create dummy variables for all characteristics except for the total duration of employment and unemployment and the daily wage. These dummies are indispensable for a reasonable interpretation but they also lead to problems of multicollinearity and drop outs as described later in the analysis section. The variable schooling contains three dummies according to the German school system, which is divided into three parts after finishing primary school and provides three possible degrees. Qualification is also measured by three dummy variables, indicating whether the individual continued by absolving vocational training in a company or at school or studying at university. The information about the conditions of the place of residence is very reliable as the different employment agencies are divided into different categories concerning the local employment situation. Thus, when recording any employment status, the agency automatically provides information about the region, which are summarized into the five categories as stated in the table below. Similarly,

Name	Definition
Gender	1 if male, 0 if female
Agegroup	age in 3 groups: 1 "below 30", 2 "31-50", 3 "over 50"
Foreigner	1 if German citizenship, 0 otherwise
Schooling	0 if no degree, 1 Secondary General School (degree after 9th grade), 2 Intermediate Secondary School (degree after 10th grade), 3 Grammar School classes A-level (degree after 13th grade)
Qualification	0 no schooling degree, 1 vocational schooling degree, 2 special vocational schools, 3 university degree
Married	1 if married, 0 otherwise
Child	1 if at least one child, 0 otherwise
Health	1 if no health problems mentioned with impact on placement, 0 otherwise
Pasthealth	same categories as health, but referring to health status at the beginning of previous employment
Disabled	1 if disabled, 0 otherwise
Benefits	benefits received before unemployment in following groups, especially subsidies for employment relationships, measures to encourage self-employment, measures for further training
Lack of motivation	1 if the person did not appear regularly at the labor office, on lack of cooperation, if other measures or programs have been abandoned at an early stage, if job proposals received by the employment office were not approved by the individual etc., 0 otherwise
Region	classification of districts of residence according to local labor market conditions into following groups: 1 urbanized district with good employment situation, 2 urbanized district with high unemployment, 3 rural district with good employment situation, 4 rural district with high unemployment, 5 districts in Eastern Germany with bad employment situation
Occupation	last employment in following categories: 1 gardener, animal breeder, fishermen, 2 miners, 3 manufacturing, 4 technical professions, 5 provision of services, 6 other labor forces
Industry	industry of last employment in following categories: 1 agriculture, building industry, 2 processing trade, 3 retail market, 4 services, 5 administrative units and insurances, 6 organizations and private households
Bluecollar	1 if last employment was a bluecollar job, 0 otherwise
Part-time	1 if last employment was not full-time, 0 otherwise
lnwage	log wage on daily basis of last employment
totempl	total duration of employment
totunempl	total duration of unemployment

Table 1: Variable definitions

the different kind of occupation and the industry of the last employment have been reduced from more than 50 different types to 6 dummy variables. These categories have been built according to the German standard industry classifications.

There are different point of times the variables are exposed. There are for example time invariant variables such as the birth year which can be simply transferred from the first record of the person in the data to all other spells. The information about the labor market characteristics must be taken from the last employment before becoming unemployed. This is the reason why the persons in the treatment as well as in the control group were required to have been in an employment contract for at least three months before becoming unemployed. The personal and regional information potentially influencing both selection into the program and the employment outcome should be measured at the beginning of the language training, so that we use the start dates for the participants and the simulated start dates for the non-participants to record details about the regional and personal background.

Even if matching is a powerful method to simulate experimental research, it is no miracle maker. It is not possible to match unmeasured variables, thus it is still not possible to control for all potential selection bias that could arise when the assignment to the treatment is not independent. In short again, selection bias can arise when some of the covariates determining the probability of a participation also influence the process of finding a job. For example it could be the case that motivated individuals have a higher probability of entering a training program and at the same time also have a higher probability of finding a job. Even if one can never be sure to have eliminated all bias, there is a way to at least narrow down possible selection bias arising from unobservable factors as motivation. For this purpose, I created a variable which should proxy soft factors influencing the participation and the job search such as motivation and stamina. Information from employment spells can be used indicating whether a person did not regularly appear at the labor office, if other measures or programs have been absolved or if it has been abandoned at an early stage, the number of job proposals received by the employment office but not approved by the individual etc..

Descriptive Statistics

Table 2 presents the descriptive statistics for the entire sample and separately for the treatment and control group.

The full sample consists of 66.114 individuals. 904 individuals meet the requirements of the treatment group, which is a tiny number in consideration of the full sample size, but it is still high enough for a representative treatment group. The entire control group includes 42.168 individuals. At this point, it is important to note that this number refers to the control group before matching. The control group so far includes all potential nonparticipants meeting the requirements to be in the control group. Thus, it is not justified which of them can serve as a suitable counterfactual conditioning on their characteristics and common support condition is not verified yet. Consequently, the number of individuals in the control group will decrease after matching. The high number of 42.168 individuals potentially serving as a statistical twin allows for extensive matching procedures such as nearest neighbor matching without replacement or caliper matching.

Male individuals form the bigger part of the full sample with 61,8%. Among the participants of the language trainings on the other hand, only 44,9% are male, thus more female than male individuals take

	Entire Sample		Treated=1		Treated=0	
	Mean	SD	Mean	SD	Mean	SD
Personal characteristics						
Gender	0.618	0.486	0.449	0.478	0.568	0.494
Age						
Agegroup 1 (below 30)	0.517	0.435	0.254	0.449	0.509	0.431
Agegroup 2 (31-50)	0.382	0.302	0.507	0.500	0.424	0.494
Agegroup 3 (over 50)	0.101	0.486	0.239	0.427	0.067	0.250
Foreigner	0.108	0.324	0.366	0.389	0.149	0.356
Influences on employment situation						
Schooling						
Degree after 9th grade	0.291	0.454	0.222	0.416	0.389	0.487
Degree after 10th grade	0.087	0.282	0.275	0.447	0.122	0.328
Degree after 13th grade	0.121	0.326	0.332	0.471	0.161	0.367
Qualification						
Vocational schooling degree	0.130	0.337	0.146	0.353	0.172	0.377
Special vocational schools	0.010	0.100	0.039	0.193	0.015	0.122
University degree	0.049	0.218	0.291	0.454	0.073	0.260
Family	0.329	0.470	0.222	0.416	0.339	0.474
Child	0.257	0.437	0.410	0.492	0.338	0.473
Health	0.047	0.212	0.020	0.141	0.047	0.211
Pasthealth	0.143	0.350	0.021	0.144	0.134	0.341
Disabled	0.020	0.141	0.000	0.000	0.016	0.127
Benefits	0.034	0.181	0.012	0.110	0.001	0.020
Lack of motivation	0.111	0.105	0.170	0.376	0.079	0.269
Regional information						
Region						
Region 1	0.029	0.169	0.139	0.347	0.038	0.190
Region 2	0.279	0.449	0.269	0.444	0.287	0.452
Region 3	0.223	0.416	0.233	0.423	0.233	0.423
Region 4	0.191	0.392	0.196	0.398	0.186	0.388
Region 5	0.273	0.226	0.162	0.368	0.251	0.434
Labor market characteristics						
Occupation						
Group 1	0.409	0.149	0.052	0.222	0.319	0.466
Group 2	0.166	0.051	0.054	0.227	0.155	0.362
Group 3	0.277	0.448	0.067	0.389	0.079	0.271
Group 4	0.058	0.235	0.083	0.289	0.107	0.309
Group 5	0.524	0.499	0.167	0.359	0.075	0.264
Group 6	0.033	0.179	0.010	0.282	0.098	0.300
Industry						
Agriculture, building industry	0.006	0.077	0.002	0.047	0.007	0.082
Processing trade	0.007	0.081	0.001	0.033	0.007	0.086
Retail market	0.007	0.086	0.000	0.000	0.010	0.098
Services	0.017	0.129	0.025	0.127	0.023	0.149
Administrative units	0.002	0.043	0.000	0.000	0.002	0.050
Parttime	0.775	0.417	0.583	0.515	0.648	0.478
lnwage	22.377	10.451	8.28	4.393	10.167	4.282
totempl	251.074	887.561	36.083	124.996	252.773	890.794
totunempl	24.067	90.139	66.076	72.424	37.710	110.560
Number of observations	66.114		904		42.168	

Table 2: Descriptive statistics

advantage of the offered language trainings. These numbers are consistent with the current findings of the BAMF about the language courses implemented in Germany since 2005 (Engler and Scheidler, 2009). The share of women has been greater than 60% in both 2005 and 2006. According to the report, these numbers are not surprising as women are defined as a special target group of the integration courses. This is due to the fact that women are traditionally responsible for the education of the children and the higher their proficiency in the German language, the better they can support their children in their educational and professional qualifications. Furthermore, it is still more often the case that families are immigrating to Germany because the husband was offered a job and not the wife. Consequently, while the husband is working, the wife takes part in the language course irrespective of whether she is aiming to find a job or to receive social benefits, what is only possible after proving sufficient language skills. Another important reason behind the high participation rate of women is the offer of free child care during the language courses as 41% of the participants have at least one child compared to 25.7% in the full sample.

Most of the participants in the language trainings are between 31-50 years old with a share of 50.7% compared to 25.4% being below 30 years and 23.9% being over 50 years old. The study of the BAMF explains this by the fact that 60% of the participants have already been living in Germany for several years and are trying to catch up with the pent-up demand being accumulated over the years. In our full sample, 10.8% of the individuals are registered as non-german citizens. This is slightly more than the official proportion of foreigners in comparison to the total German population which amounted to 9% in the year 2010. This number serves only as a comparable illustration as the number in our sample is aggregated over several years. According to the requirements for the participation in the language course, one expects the share of the foreigners in the treatment group to be 100%. This is not the case as individuals having successfully completed the language trainings can apply for the German citizenship and only the latest registered citizenship is given in data set.

The descriptive statistics of the schooling and qualification variables are quite astonishing. An almost unrealistic share of 32.6% of the participants of the language training reached the highest schooling degree compared to a number of not more than 12.1% of the full sample. The same applies to the professional qualification as 29.1% of the treatment group graduate with a university degree, whereas only 4.9% of the full sample accomplish the same degree. Because of these unusual results, these variables have to be handled with care in the analysis. At the same time, the variables are of great importance as the numbers indicate positive selection into the language trainings. More precisely, the numbers point out that higher educated individuals are more motivated to acquire language skills and it must be assumed that they also have better chances in the labour market. Omitting these variables would probably lead to drastic selection bias and therefore it is indispensable to include those variables into the probit regression model.

As a last aspect, the labour market characteristics of the last employments should be considered. On average, the daily wage aggregates to 22.38€ in the full sample, whereas individuals in the treatment group only get paid by 8.28€ per day. While the total duration of continuous employment on average amounts to 251 days, the individuals of the treatment group have been employed continuously for 36 days on average. Thus, the fluctuation in jobs is higher for the treatment group, which also explains that the

differences in the duration of unemployment are not that large with 24 days in the full sample and 66 days in the treatment group.

To sum up, most of the numbers of the descriptive statistics are consistent with the expectations and with findings from other studies. However, some numbers are rather less distinctive such as the occupation and industry variables or some numbers are even questionable with regard to their validity such as schooling and qualification and consequently must be accepted with reservation.

6. Analysis and Empirical results

6.1. Propensity Score Estimates

Table 3 shows the results for the probit regression for different specifications of models. Unfortunately, three problems led to rather parsimonious models. The first problem arose due to the fact that almost only dummy variables were used in the analysis. Therefore a lot of variables were dropped due to multicollinearity. The second problem has to do with the fact that some variables were dropped because they predicted the failure perfectly. For example, there was no individual being disabled and participating in the language course. Heckman et. al. (1998) describe this circumstance as a problem of the common support. If P(x)=0 or P(X)=1 for some values of X, then it cannot be matched conditional on these X values as persons with these characteristics either always or never receive treatment. Thus, the common support condition fails as persons with the same characteristics cannot be observed in both the treatment and the control group.

The third and gravest problem is that most of the variables did not pass the balancing test. To point out the importance of this balancing test, it is necessary to look back to the theoretical background of the PSM. As stated in chapter 3, the propensity score is one possible balancing score which can be used to reduce the matching process to a one dimensional level. Because this balancing property is the primary purpose of the propensity score, it is of prime importance to check whether the covariates have the same distribution for the treatment and comparison groups at each value of the propensity score. To put it formally, the object of the test is to verify that

$$D \perp X \mid p(X) \tag{11}$$

where X is the set of covariates that are believed to satisfy the conditional independence assumptions. The balancing test should assure that after conditioning on $p(x)$, no further variables can be added such that further conditioning on X would provide new information on D and improve the estimation results. Furthermore after the matching process, there should be no significant differences between the distribution of the characteristics of the treatment and control group.

The practical approach to satisfy the balancing property is described extensively by Dehejia and Wahba (2002) and it has been the guide for the writers of the stata package pscore, where an algorithm for the balancing score has been included. Applying this algorithm allows to specify the significance level of the balancing test to either 1%, 5% or 10%. Furthermore one can adopt the number of blocks in which all observations are stratified such that there is no significant difference in the estimated propensity score between the treatment and control group for each block.

	Model 1	Model 2	Model 3	Model 4	Model 5
Personal characteristics					
Gender	-0.292***	-0.114***		dropped	0.079
Age					
Agegroup 2 (31-50)	0.372***	0.483***		0.412***	0.594***
Agegroup 3 (over 50)	0.653***	0.942***		0.930***	1.078***
Influences on employment situation					
Schooling					
Degree after 9th grade		0.166***		dropped	0.166
Degree after 10th grade		0.124***		dropped	0.657
Degree after 13th grade		-0.251***		dropped	0.581
Qualification					
Vocational schooling degree		dropped		dropped	dropped
Special vocational schools		0.329***		0.558***	0.725**
University degree		0.562***		0.712***	0.697***
Married		-0.186***		-0.133***	0.082
Child		0.112**		dropped	0.128
Health		-0.024*		dropped	-0.665
Pasthealth		-0.729**		dropped	-0.679
Disabled		dropped		dropped	dropped
Benefits		0.323*		dropped	0.925
Motivation		-0.079		dropped	dropped
Regional information					
Region					
Region 1			-0.484***	dropped	-0.748***
Region 2			-0.026	-0.024	-0.627*
Region 3			0.027	0.027	-0.677*
Region 5			-0.203***	-0.203***	-0.785**
Labor market characteristics					
Occupation					
Group 2			-0.553***	dropped	
Group 3			-0.720	dropped	
Group 4			-0.682*	dropped	
Group 5			-0.887***	dropped	
Group 6			-0.856*	dropped	
Industry					
Processing trade			-0.776**	dropped	
Retail market			-0.692***	dropped	
Services			-0.953***	dropped	
Administrative units			-0.602**	dropped	
Parttime			dropped	dropped	dropped
lnwage			-1.832***	-1.763**	-1.713***
totempl			dropped	dropped	dropped
totunempl			dropped	dropped	dropped
Number of observations	66.114	46.214	48.133	46.309	23.301
R^2	0.0446	0.1275	0.0813	0.1643	0.1591

The stars in this and all following tables indicate the level of significance as follows:
* lowly significant at 10% level | ** significant at 5% level | *** highly significant at 1% level
The reference person is a female worker younger then 30 years old, having no educational or qualification degree, not being married and having no children. The person is living in a rural district with high unemployment and is working as a gardener, animal breeder or fisher in the agriculture industry.

Table 3: Probit regressions according to different model specifications

To find the model specification which satisfies the balancing property is one of the hardest challenges in the method of PSM. The best way to find an appropriate model is to start with a parsimonious specification to estimate the propensity score. The next step is to identify those variables showing a large difference in means between the treatment and the control group. If there are no significant differences, what means that all covariates are balanced in all blocks, further covariates can be added. If one covariate is not balanced in only one special block, one might try to split the block and then test again with the smaller blocks. If instead the covariate is not balanced in more than one block or even in all blocks, then one should modify the model used in the probit regression by adding more interaction with other variables and higher order terms and then test the new model again. This would lead to a less parsimonious model as demanded by the authors of the pscore package: "If the means of one or more characteristics differ, (...) a less parsimonious specification of h(Xi) is needed" (Becker, Ichino p. 360). On the other hand, Bryson, Dorsett and Purdon (2002) advocate to leave out variables that do not satisfy the balancing property because over-parameterized models lead to problems with the common support. Furthermore, although the inclusion of non-significant variables will bias the estimates, it can increase their variance since either some treated have to be excluded from the analysis or control units have to be used more than once. Thus, leaving out those non-significant variables can be a proper way to increase the randomness of the selection process and to reduce the variance of the matching estimates (Augurzky, B. and Schmidt, C. 2000).

To sum up, there are no clear guidelines about how to specify the correct model and there are both reasons for and against including the full set of covariates. Thus, it is basically up to the researcher to find an economic plausible specification of the model. According to the method of specific-to-general modeling, model (1) in Table 3 includes just a few variables concerning the gender and the age of the individuals. The results are highly significant, but the pseudo R^2 is modest and it is obvious that more variables are needed to overcome unobserved influences. Model (2) additionally includes background information that influences the current labour market situation. Two variables were dropped because of multicollinearity, but the remaining results are fairly significant and the show the expected sign of coefficients. Model (3) only controls for the regional differences and the labour market characteristics of the previous employment. The signs are mainly negative. In some cases this is expected, for instance if having worked in service significantly decreases the probability of participating in the language course by 95.3% compared to those having worked in the agricultural industry. This is plausible because proficiency in language is much more important when working in the service than in agriculture and provides the service workers a solid language knowledge so that they do not need to improve their skills. However, there are also astonishing results, for example that living in region 3 (rural district with good employment situation) increases the probability of participating in the language courses by 2.7% compared to the reference group living in region 4 (rural district with high unemployment). These astonishing results indicate that the estimation might be biased because of unobserved heterogeneity in the left-out variables. Consequently, the next step is to add more explanatory variables to the model. The specification of the model (4) includes all available variables but the results are very parsimonious as more than two third of the variables have been dropped due to the problems described above. A reasonable amount of variables can be kept in the model if the binary variables about the industry and the occupation are left out of the model as done in (5). However, this specification does not look satisfying either as only few variables have significant impact on the participation probability and the number of usable observations dropped to 23.302 individuals. Furthermore, some variables are implausible again as for example finishing school with the highest degree

increases the probability of participating in the language course by 58,1% compared to those not having graduated from school at all. This is implausible as the share of high-skilled immigrants having passed 13 years of education in their home country is very small compared to the total number of immigrants and those immigrants having graduated at a school in Germany must have enough proficiency in language.

All of the five specifications show considerable drawbacks. By reason of plausibility the propensity score estimates of model (2) are used for the matching. This model excludes the labour market characteristics but at least some variables are significant and they show the expected sign of coefficients. The pseudo R^2 of the probit in this model is quite low with 0.1275. This number designates how well the included covariates X explain the participation probability and this low number speaks for a rather weak specification, which must be kept in mind for the further interpretation.

6.2. Treatment effects

The estimated average treatment effects of the language trainings for immigrants are shown in Table 4.

	Different Matching Methods			
	Nearest Neighbor	Stratification	Kernel*	Radius
Employment after 3 months	-0.076 (0.011)***	-0.075 (0.012)***	-0.163 (0.012)***	-0.134 (0.032)***
Employment after 6 months	-0.081 (0.011)***	-0.082 (0.011)***	-0.165 (0.015)***	-0.107 (0.031)***
Employment after 12 months	-0.079 (0.013)***	-0.078 (0.012)***	-0.167 (0.011)***	-0.148 (0.033)***

Standard errors in brackets; *bootstrapped standard errors

Table 4: ATT on employment status

The effects were evaluated for different point of times by observing the employment status first after 3 months, after 6 months and after 12 months. Furthermore, different matching methods were used to ensure that the best identification strategy is used. It is noticeable that the NN method and the stratification method yield similar results and on the other hand the Kernel and Radius method yield similar results, which are twice as high as those from NN and the stratification method.

The results show significantly negative results on the 1% level irrespective of the method and the time horizon. Taking the results of the NN method as a basis, the ATT connotes that 3 months after the start of the program, the participants had a 7.9% lower employment probability than they would have had if they had not participated in the program. As we used the start date of the measure as the start date of the effect evaluation, it is important to note that a lot of participants might still be under treatment after 3 months. Therefore, a typical explanation for the negative ATT are the locking-in effects. As already explained above, locking-in effects describe the fact that the job search intensity may be lower while participating in a program. This applies especially for full-time measures like the language training. First, participating in a language course implies high opportunity costs because 35 hours a week are spend on learning German instead of searching for job offers, writing applications and interviewing for jobs. Second, participants might not be interested in finding a job at short notice during the language course as this would force them to stop the course. Without a certificate of the passed language training, they are not reimbursed for the costs of the training and they are further not allowed to make use of any other measures or social welfare benefits in case that they get unemployed again. Summing up, the job finding rate is expected to be low during the participation in the language course.

The maximum duration of the language course by law is 6 months. Hence, when controlling for the ATT after 6 months, we expect that some of the motivated and capable participants already finished the course earlier and some of the participants are just about to finish it. By all means we expect both type of participants to have solid knowledge of the German language, which should have a positive impact on the job search. However, the negative ATT even became more pronounced with -8.1%. According to Schneider and Uhlendorff (2006) this is not surprising as it can take several months to overcome the locking-in effects. First of all, it is possible that the participants need some time for an orientation period and for preparing their résumés and other references. The second explanation especially applies to the regulatory fact mentioned earlier that the successful participation in the language course is the basic requirement for all further measures of ALMP. Thus, after completing the language training, many participants may continue with further educational programs which adds more time to the locking-in effect. The third explanation is a mathematical one: at the end of the training, the differences in the state probabilities between participants and non-participants are high because of the locking-in effect. To catch up these differences, even higher effects are needed after the end of the measure such that the former lock-in effect can be compensated and that the overall ATT a positive one. Due to this, it is possible that positive effects can only be observed after a longer observation period. So far, there is no consensus about how long it takes to overcome the locking-in effect so that it cannot be stated whether the negative effect after 12 months of -7,9% is still a consequence of the locking-in effect or if the participation in the language courses is actually negatively influencing the employment probabilities.

The findings are consistent with a number of other evaluation studies, which show negative effects in the short run and positive effects only after several years. Unfortunately, it is not possible to conduct such a long run study about the impacts of the language trainings on the basis of the IEBS data set. The registration of the language training participants only started in the year 2000. The latest information of both the language trainings as well as the employment situations are recorded in the year 2004. Hence, to evaluate the impacts of the language training after an outcome period of 12 months, we can only use language course information from the years between 2000-2003 as no employment effects can be observed for those participants starting the course in the year 2004. It would be possible to extend the outcome period to 24 or even 36 months, but this automatically leads to a reduction of the observation period. For instance, in case that the outcome period should be extended to 2 years in order to allow for potential locking-in effects, it would only be possible to observe those individuals who participated in the training between the years 2000 and 2002. A drastic number of treated individuals would get lost in doing so and one has to ponder about the trade-off between a longer outcome period and less heterogeneity in the characteristics. As the heterogeneity has not been very extensive in the first place as indicated by the low R^2 and the dropped variables of the probit estimation, a extension of the outcome period to 24 months does not add up. Furthermore, it could be the case that even the time period of 24 months is not enough to overcome the locking-in effects as some studies only show positive effects after 8 years.

As a matter of fact, with this data basis it is only possible to conduct a short run analysis of at the longest 12 months. Some studies show an alternative way to control for locking-in effects in short run investigations, that is to use proportional hazards models to estimate the probabilities of the transition from unemployment into employment. Such a method implicates more restrictive identifying assumptions, but on the other hand it allows to directly prove positive effects of a program as the transition probabilities

are independent of the duration of the observation period. As a consequence, it is possible to identify locking-in effects and program effects separately. Unfortunately, estimating such transition probabilities for instance by estimating semi-parametric Cox regressions goes beyond the scope of the discussion but it would be a reasonable way to allow for the common locking-in effects.

Treatment effects and regional differences

For reasons of practicability, the variables defining the regional differences have not been taken into account in the model specification (2). However, the regional differences in the overall labour market conditions might have an influence on the participation probability as well as on the chance to find a job. To leastwise cover the latter influence, the ATT is estimated separately for Eastern and West Germany. Even more than 20 years after the German reunification and after many attempts for structural adjustments, the regions in the Eastern part of Germany still suffer from the lack of infrastructure and from low economic expansion resulting in very limited labour supply. The bad future prospects cause an outflow of young people moving to the western parts of Germany. As a consequence, Eastern Germany is afflicted with severe skilled worker shortage and therefore only little investments or capital expenditures are made. This convolving development is currently non-controllable and therefore the results of a separate estimation for the Eastern and Western part of Germany is of interest. The results of the ATT gained through the NN method are presented in Table 5.

		Nearest Neighbor
West Germany	Employment status after 3 months	-0.071 (0.014)***
	Employment status after 6 months	-0.077 (0.012)***
	Employment status after 12 months	-0.074 (0.014)***
Eastern Germany	Employment status after 3 months	-0.094 (0.025)***
	Employment status after 6 months	-0.100 (0.026)***
	Employment status after 12 months	-0.097 (0.026)***

Standard errors in brackets

Table 5: ATT for Eastern and West Germany

The differences between the Eastern and Western part of Germany are not as big as expected, but highly significant. While the probability of finding a job after participating in the language training is reduced by 7.1%-7.7% in Western Germany, the probability is reduced by 9.4%-10% in Eastern Germany. In terms of our analysis, the small difference is a good sign as it indicates that the regional differences do not have such a great impact and hence leaving out the variables indicating the regional differences might not lead to significant unobserved heterogeneity.

As a further exploration, the ATT are estimated separately for male and female individuals in Table 6. Here, the differences are as follows: While the reduction of the probability of being employed for males lies between significant 1.6%-4.5%, participating in the language training seems to influence female individuals more negatively by a reduction in the employment probabilities by 10.9%-12.7%. This can be explained by the fact that more female immigrants participate in the language courses and some of them do not to learn German with the objection of finding a job. Instead, one can often find the traditional role allocation in immigration families, where the women stay at home and raise the children so that the women just want to learn German for the purpose of a better integration or for the receipt of social

		Nearest Neighbor
Male	Employment status after 3 months	-0.027 (0.020)*
	Employment status after 6 months	-0.045 (0.018)***
	Employment status after 12 months	-0.016 (0.021)***
Female	Employment status after 3 months	-0.113 (0.014)***
	Employment status after 6 months	-0.109 (0.013)***
	Employment status after 12 months	-0.127 (0.014)***

Standard errors in brackets

Table 6: ATT for male and female

benefits. As a consequence they do not even apply for jobs and this negatively influences the employment probabilities for female participants. As the share of female participants is high, this circumstance might also contribute to the overall negative effects of the language trainings.

Treatment effects on earnings

Recent papers have introduced another interesting dimension to the evaluation literature by investigating the impact of ALMP on the quality of the jobs. By quality it is meant that the job requirements fit to the skills of the workers, that the jobs are longer lasting and that the earnings are higher. Due to the short time horizon of the available data after the participation in the language course, the duration of the job after the training cannot be measured and there is not enough information to examine the job matching. Hence, only the impact of the language course on the upcoming earnings of the participant can be investigated. The estimates of the nearest neighbor matching can be found in Table 7.

The outcome variable in that case is the daily wage of on employer. Similar to the ATT on the employment probabilities, the language course participation also negatively influences the earnings of the following job. In case that the participant successfully finds a job after 3 months, the earnings will be 64.1 cents lower compared to the earnings of the matched non-participant. The impacts on earnings after 6 and 12 months are similar with -63.5 cents and -66.6 cents. These numbers are not too big so that the negative effect can still be explained by the locking-in effects because the earnings differences are driven by the employment dynamics as zero earnings are recorded in case of unemployment.

|	Nearest Neighbor
Impact on earnings after 3 months	-0.641 (0.052)***
Impact on earnings after 6 months	-0.635 (0.048)***
Impact on earnings after 12 months	-0.666 (0.054)***

Standard errors in brackets

Table 7: ATT on earnings

Estimating the impact of ALMP not only on the employment probabilities but also on the earnings, enables to investigate an important supplementary research question: Whether the programs are desirable from a society's point of view and if the huge amount of money spent on in can be justified from an economic perspective. This question can only be answered by a cost-benefit analysis. For this purpose, a comparison between all the expenditures and all the benefits of the program would be needed. The difficulty with this is that the benefits of the programs are hard to express in monetary terms. One possibility is to use earnings as the outcome variable instead of the employment status and to consider

the positive chance in earnings as the benefit of the program. In our case, a cost-benefit analysis does not make sense as the impacts of the language course on earnings within the time period of 12 months is negative and thus, the net social benefit of the program would always be negative. However, cost-benefit analyses are important tools to determine the viability of the program and Kluve (2006) postulates that future research must concentrate on this kind of evaluation. A pioneer work for example has been done by Jespersen et. al. (2008) who investigate the costs and benefits of Danish ALMP. They find substantial positive long-term treatment effects and therefore the programs provide social surpluses even when taking the costs into account.

6.3. Quality of the match

After estimating the propensity score, the next step is to verify the quality of the match by controlling the region of common support between the treatment and control group. In practice, all those treatment observations were deleted whose propensity score is smaller than the minimum and higher than the maximum propensity score of the control group. In our case, the region of common support lies between [0.0362683, 0.2912773]. A straightforward way to verify that the common support condition is fulfilled is a visual comparison of the density distribution of the propensity scores in both the treatment and control group as presented in Figure 5. The common support condition demands that the densities of the treatment and the control group are more similar after matching. The figure on the right shows treated individuals in green and untreated individuals in red and the common support criterion is satisfied as there are enough observations in the common support. The cut of control observations which do not fall into the common support area is illustrated by the blue shaded area in the right figure. The loss amounts to approximately 1.4 percent, which is rather small and thus poses few problems according to Bryson et. al. (2002) as the subpopulation can still be considered to be representative.
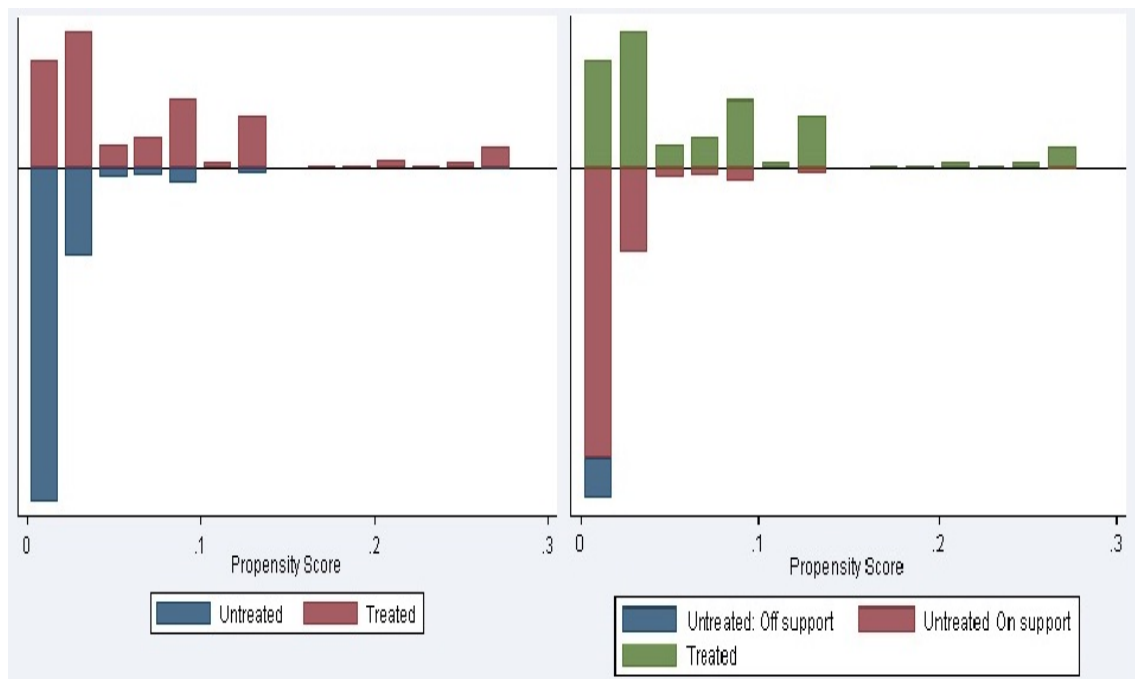
Figure 5: Propensity score distribution before and after matching

The graphical comparison of the treatment and control group after the matching process indicates a weak adjustment of the density distributions as only few individuals of the control group with the lowest propensity score are excluded. For higher values of the propensity score, the distribution is unchanged which does not speak for a high quality of the match. Furthermore, it is noticeable that the densities of the treatment group are always higher than those of the control group, except for the lowest propensity score. Especially when it comes to the higher values of the propensity score, the densities of the control group are rather thin. This signifies that many treated individuals with high propensity scores are matched with the same individual of the control group, which could be a major source of biases. One possibility to avoid this problem is to impose a trimming rule, which excludes a certain percentage q of the treatment observations at which the density of the control observations is the lowest. In our case, the percentage q should have been chosen such that the treatment observations in the interval $[0.15, 0.3]$ are excluded from the analysis since the densities of the counterfactuals are very low in this area. This trimming method was first suggested by Heckman et. al. (1998) and has been used often as a supplement to the maximum-minimum comparison. Again, the application of the trimming method depends on the data situation and as the treatment group in this study is already small and lacking of heterogeneity in the characteristics, it is not advisable to further narrow down the treatment group by a trimming rule.

As the graphical comparison does not provide satisfying results, further quality checks are preferable. There are numerous ways to examine the quality of the matching procedure and the analysis itself. Unfortunately, limited temporal data access[13] hindered me from performing further methods to validate the quality. If the time frame had been more extensive, I would have calculated the standardized bias suggested by Rosenbaum and Rubin (1985) as a quality indicator for the matching procedure. This standardized bias is defined as the difference of the sample means in the treated and the matched control group as a percentage of the square root of the average of the sample variances in both groups. Successful matching demands that the standardized biases are smaller after matching than they had been before matching. There is no clear rule about how big the reduction in biases should be to indicate successful matching, but according to Caliendo and Kopeinig (2008) in most empirical studies a bias reduction below 3% or 5% is seen as necessary and achievable.

In a similar approach, t-tests could have been conducted to verify the equality of means for both the treatment and the control group. Differences in means before matching are natural, but after matching there should not be significant differences in the means as the covariates should be balanced satisfyingly. Unfortunately, the actual improvement of the bias reduction is not visible and again there is no guideline about which values are necessary to indicate a good matching quality.

Finally, a further indicator for the matching quality is the pseudo R^2. For this purpose, I would have estimated the propensity score again on the already matched sample and then the pseudo R^2 are compared before and after matching. As the aim of matching is a similar distribution of the covariates between the treatment and control group, the R^2 should be fairly low after matching.

It is furthermore common to perform a sensitivity analysis as a last step of the propensity score matching. Sensitivity analyses are time-consuming and could not be performed in this context because of the restricted data access mentioned before. However, to get an idea of the potential examination strategy at least the aim and the different versions of the sensitivity analysis are presented in the following.

---

[13]Due to the legal regulations data access was provided only from 20.-24.7.2011 at the IAB.

The aim behind this analysis is to assess the robustness of the results. Therefore the analysis should cover all alternative specifications and possible techniques of the matching process in order to control how sensitive the results react on variations. Possible variations in this study can be derived from all decisions which had to be made during the practical implementation of PSM. For example, it could be of interest how the results would change if only individuals indicating foreign citizenship are included in the treatment group. As a reminder, I decided to also take into account language course participants indicating German citizenship in order to avoid biases. Furthermore, I decided to assign course drop outs to the treatment and to the control group according to the total duration they participated in the training. It might be interesting to find out how sensitive the results react if all drop outs are excluded from the sample. Further possible variations could be to use different matching methods, to calculate the ATT for all presented probit model specifications and to use Cox regressions or to try out different combinations of observation and outcome periods to minimize the occurrence of locking-in effects but maximize the validity of the results.

7. Conclusion

The core question of the paper is whether language courses are effective with regard to the labor market outcomes. As the analysis shows, this question cannot be answered satisfactorily. The results show negative and significant average treatment effects of the language trainings, which means that acquiring addition language skills decreases the probability of being employed. One cannot conclude from this result that language trainings are ineffective as the data only allowed a short-time analysis and the locking-in effect might be the reason for the negative results. Furthermore, the high share of female participants not being active in the labour market afterwards might contribute to the negative results as well as those individuals in the control group who unobserved attended a language training before the year 2000. A further explanation for the negative results are the econometric problems which occurred during the analysis and which raise doubts about the robustness of the result.

To sum up, these problems started with the fact that only the current citizenship is recorded and thus it cannot be matched on the basis of the different nationalities. A lot of variance got lost here and increased the hazard of bias by unobserved influences. Further challenges occurred when specifying the variables of the probit regression, namely multicollinearity, variable drop outs and unsatisfied balancing properties. The resulting model is very parsimonious and only limited in its capacity to minimize the selection biases. Last but not least one has to consider the limitations of the matching method. Even if this method is the best method known for the purposes of the evaluation of ALMP due to its advantages stated in section 4.1.3, this method is no miracle maker as it is only as good as the quality of the covariates.

Sustained focus of future research must therefore lie on the occurrence and avoidance of potential biases, which cannot be eliminated by the matching technique so far. Future language course evaluations for instance must take into account that in contrast to other measures which transfer specialized knowledge, language as the object of transfer is a type of human capital where the access is easier as we are surrounded by language in the every day life. Therefore, the level of knowledge differs considerably between the participants of the language course. Not taking these language differences into account may cause further biases as these unobservables have influences on the success of the language course and

on the job search. Therefore, further research requires a rich data set containing a lot of heterogeneity in the characteristics allowing for all econometric approaches and preferably also information about the motivation of a person and the current proficiency level of the language.

It is hard to formulate policy implications resulting from the analysis, not only because the robustness of the results is questionable, but also because the effectiveness of the program with regard to the employment situation is not the only aim of the language trainings. A complete cost benefit analysis would be needed in order to be able to give well-grounded policy advices and to answer whether the amounts of public expenditures can be justified or not from a society's point of view. According to Caliendo and Hujer (2005), the ideal evaluation process consists of the following steps: "First, the impacts of the programme on the individual should be estimated (microeconometric evaluation). Second, it should be examined if the impacts are large enough to yield net social gains (macroeconomic evaluation). Third, it should be answered if this is the best outcome that could have been achieved for the money spent (cost-benefit analysis)." Until now, most of the studies concentrated on the first step, because cost-benefit analyses are rather difficult to conduct reliably as the expression of non-monetary terms is always done partly arbitrarily. Especially when it comes to external effects, it is difficult to valuate the impacts in monetary terms. Concerning the language courses, it is important to mention that the purpose of the language courses should not only be the integration of immigrants into the local labor market, but the language courses also make a huge contribution to the social integration of the immigrants. This is directly linked with the social recognition of a person, with building up social networks and the identification with the host country. These aspects are not to be underestimated and therefore it is not sufficient to evaluate the justification of a program only on the basis of pure economic outcomes and therefore there is a strong need for further research in this field of active labor market policy.

However, before being able to perform comprehensive cost-benefit analyses for the fiscal legitimization of such policies, the fundamental research must achieved to be reliable. Unless the results of the microeconometric evaluation studies are not robust in statistical terms, it does not make sense to proceed to the macroeconometric evaluation and much less to cost-benefit analyses. As the literature review and the practical guides of propensity score matching show, there are still too many drawbacks and unambiguity in the results to rely on it. This study contributes to the need for fundamental research of the effectiveness of language trainings. Even if the results are not as might be desired, studies of that kind are necessary to analyze the effectiveness of such programs as a first step to answer whether the expenditures for the policies can be legitimized from a society's point of view.

Literature

Augurzky, B. and Schmidt, C. (2000): The Propensity Score – A Means to an End, IZA Discussion paper, No. 271, Bonn.

Becker, S. and Ichino, A. (2002): Estimation of average treatment effects based on propensity scores, *Stata Journal*, StataCorp LP, vol. 2(4), pp. 358-377.

Bender, S., Biewen, M., Fitzenberger, B., Lechner, M., Miquel, R., Osikominu, A., Waller, M.

and Wunsch, C. (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 und SIAW St. Gallen.

Berman, E., Lang, K. and Siniver, E. (2003): Language-skill complementarity: returns to immigrant language acquisition, *Labour Economics*, Elsevier, vol. 10(3), pp. 265-290.

Bryson, A., Dorsett, R. and Purdon, S. (2002): The use of propensity score matching in the evaluation of active labour market policies, Working Paper No. 4, Department for Work and Pensions.

Caliendo, M. and Hujer, R. (2006): The Microeconometric Estimation of Treatment Effects - An Overview, *Journal of the German Statistical Society* vol.90(1), pp. 197-212.

Caliendo, M., Hujer, R. and Thomsen, S. (2008): The Employment Effects of Job Creation Schemes in Germany - A Microeconometric Evaluation, in: Millimet, D., Smith, J. and Vytlacil, E. (eds.), *Advances in Econometrics*, vol. 21: Estimating and Evaluating Treatment Effects in Econometrics, pp. 383-430.

Caliendo, M. and Kopeinig, S. (2008) , Some practical guidance for the implementation of propensity score matching, *Journal of Economic Surveys*, Vol. 22(1), pp. 31- 72.

Chiswick, B. R. (1991): Speaking, Reading, and Earnings among Low- Skilled Immigrants, *Journal of Labor Economics*, vol. 9(2), pp. 149-170.

Deeke, A. (2006): Berufsbezogene Sprachförderung für Arbeitslose mit Migrationshintergrund, Erste Ergebnisse aus der Begleitforschung zum ESF-BA-Programm, IAB Forschungsbericht Nr. 21/2006.

Dehejia, R.H. and Wahba, S. (2002): Propensity Score Matching Methods for Non- Experimental Causal Studies, *Review of Economics and Statistics*, vol. 84(1), pp. 151-161.

Dustmann, C. and van Soest, A. (2001): Language Fluency and Earnings: Estimation with Misclassified Language Indicators, *The Review of Economics and Statistics*, vol. 83(4), pp.663-674.

Dustmann, C. and Fabbri, F. (2003): Language proficiency and labour market performance of immigrants in the UK, *Economic Journal*, Royal Economic Society, vol. 113(489), pages 695-717.

Engler, M. and Scheidler, A. (2009): Deutschland: Bilanz der Integrationskurse, *Migration und Bevölkerung*, Netzwerk Migration in Europa e. V.,
$http://www.migration-info.de/migration\_und\_bevoelkerung/archiv/ausgaben/ausgabe0907.pdf$.

Fitzenberger, B. and Wilke, R.A. (2004): Unemployment Durations in West-Germany Before and After the Reform of the Unemployment Compensation System During the 1980s, ZEW Discussion Paper No. 04-24.

Fitzenberger, B., Osikominu, A., and Völter, R. (2006): Get Training or Wait? Long Run Employment Effects of Training Programs for the Unemployed in Germany, IZA DP No. 2121.

Fitzenberger, B., Osikominu, A., and Völter, R. (2007): Which Program for Whom? Evidence on the Comparative Effectiveness of Public Sponsored Training Programs in Germany, IZA DP No. 2885.

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

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

Grenier, G. (1984): The Effect of Language Characteristics on the Wages of Hispanic American Males, *Journal of Human Resources*, vol.19(1), pp. 35–52.

Heckman, J. (1979): Sample Selection Bias as a Specification Error, *Econometrica*, vol. 47(1), pp.153-161.

Heckman, J., Ichimura, H. and Todd, P. (1998): Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme, *Review of Economic Studies*, vol. 64, pp.605-654.

Holland, P. (1986): Statistics and Causal Inference, *Journal of the American Statistical Association*, vol. 81(396), pp. 945-960.

Hujer, R. and Wellner, M. (2000): Berufliche Weiterbildung und individuelle Arbeitslosigkeitsdauer in West- und Ostdeutschland: Eine mikroökonometrische Analyse, Mitteilungen aus der Arbeitsmarkt- und Berufsforschung, 33, pp. 405–420.

Jespersen, S., Munch, J. and Skipper, L. (2008): Costs and benefits of Danish active labour market programmes, *Labour Economics*, vol. 15(5), pp.859-884.

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

Kraus, F., Puhani, P. and Steiner, V. (2000): Do Public Works Programs Work? Some Unpleasant Results from the East German Experience, in: S. Polaschek (ed.), *Research in Labour Economics*, JAI Press.

Krekeler, C. (2001): Sprachförderung für Spätaussiedler: Ein erfolgreiches Auslaufmodell, *Deutsch als Zweitsprache* vol. 0.

Lechner, M. (1999): Earnings and Employment Effects of Continuous Off-the-Job Training in East Germany after Unification, *Journal of Business and Economic Statistics*, vol. 17, pp. 74-90.

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

Lechner, M. and Eichler, M. (2002): An Evaluation of Public Employment Programmes in the East German State of Sachsen-Anhalt, *Labour Economics*, vol.9, pp.143-186.

Lechner, M. and Gerfin, M. (2002): A Microeconometric Evaluation of the Active Labour Market Policy in Switzerland, *The Economic Journal*, vol. 112, pp. 854-893.

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

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.

Leuven, E. and Sianesi, B. (2003): PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing, http://ideas.repec.org/c/boc/bocode/s432001.html, Version 4.0.4 10nov2010.

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

Rivera-Batiz, F. L. (1990): English Language Proficiency and the Economic Progress of Immigrants, *Economics Letters*, vol. 34(3), pp. 295-300.

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

Rosenbaum, P.R. and Rubin, D.B. (1985): Constructing a Control Group Using Multivariate Matched Sampling Methods that Incorporate the Propensity Score, *The American Statistican*, vol. 39, pp. 33–38.

Rubin, D. B. (1978): Bayesian inference for causal effects: The Role of Randomization, *Annals of Statistics*, vol. 6, pp. 34–58.

Schneider, H. and Uhlendorff, A. (2006): Die Wirkung der Hartz-Reform im Bereich der beruflichen Weiterbildung, *Zeitschrift für ArbeitsmarktForschung*, vol. 39(3), pp. 477-490.

Speckesser, S. (2004): Using Social Insurance Data for the Evaluation of Active Labour Market Policy: Employment Effects of Further Training for the Unemployed in Germany, Working Paper, University of Mannheim.

A. Overview on microeconometric evaluation studies of German ALMP

Year	Author	Title	Program	Data	Observation period	Defined Outcome	Empirical Method	Results
1995	Steiner, Kraus	Haben Teilnehmer an Arbeitsbeschaffungsmaßnahmen in Ostdeutschland bessere Wiederbeschäftigungschancen als Arbeitslose?	Job Creation Scheme	AMM-O	1990-1992	Employment rates	Hazard rate model	Zero effects in the long term
1997	Eichler	Arbeitsbeschaffungsmaßnahmen (ABM) in Sachsen-Anhalt: Vorläufige Ergebnisse einer Untersuchung von Beschäftigungschancen von Teilnehmern im ersten Arbeitsmarkt	Job Creation Scheme	AMM-SA	1993-1996	Probability of unemployment	Probit model and DiD	Negative effect on employment probabilities
1997	Hübler	Evaluation beschäftigungspolitischer Maßnahmen in Ostdeutschland	Various instruments of ALMP	AMM-O	1990-1994	Employment probabilities	Multinomial logit model	No significant effects
2000	Bergemann, Fitzenberger, Schultz, Speckesser	Multiple Active Labor Market Policy Participation in East Germany: An Assessment of Outcomes	Job Creation Scheme; Training	AMM-SA	1990-1998	Employment effects	Conditional DiD	Significant negative effects on first treatment; insignificant, negative effects on second treatment
2000	Hujer, Wellner	Berufliche Weiterbildung und individuelle Arbeitslosigkeitsdauer in West- und Ostdeutschland: Eine mikroökonometrische Analyse	Training	SOEP	1990-1993	Employment probability; duration of employment after measure	PSM	No significant effects
2000	Kraus, Puhani, Steiner	Do public works programs work? Some unpleasant results from the East German experience	Public work programs	AMM-O	1990-1994	Reemployment rates	Hazard rate model	Negative effects for both male and female
2000	Lechner	An Evaluation of Public-Sector-Sponsored Continuous Vocational Training Programs in East Germany	Vocational training	GSOEP	1990-1994	Unemployment rate	PSM with Caliper	No positive effects in the first years after training
2002	Eichler, Lechner	An evaluation of public employment programmes in the East German State of Sachsen-Anhalt	Public employment programmes	AMM-SA	1992-1997	Unemployment probabilities	PSM with NN; DiD	Significant reduction in unemployment rate
2002	Jaenichen	Lohnkostenzuschüsse und individuelle Arbeitslosigkeit. Analysen auf der Grundlage kombinierter Erhebungs- und Prozessdaten unter Anwendung von Propensity Score Matching	Wage subsidies	Prefiguration of Measure-Participation-Data	1999-2001	Employment status	PSM with NN	Significant reduction of unemployment rates

Year	Author	Title	Program	Data	Observation period	Defined Outcome	Empirical Method	Results
2003	Reinowski, Schultz, Wiemers	Evaluation von Maßnahmen der aktiven Arbeitsmarktpolitik mit Hilfe eines iterativen Matching-Algorithmus. Eine Fallstudie über langzeitarbeitslose Maßnahmeteilnehmer in Sachsen	Trainings	Micro-census Sachsen	1989-2000	Departure from unemployment	Mahalanobis-Matching; Proportional hazard model	Negative effects
2004	Caliendo, Hujer, Thomsen	Evaluation der Eingliederungseffekte von Arbeitsbeschaffungsmaßnahmen in reguläre Beschäftigung für Teilnehmer in Deutschland	Job Creation Scheme	Prefiguration of Measure-Participation-Data	2000-2002	Departure from unemployment	PSM	Results differ depending on subgroup, but mainly results are negative
2004	Hujer, Thomsen, Zeiss	The Effects of Vocational Training Programmes on the Duration of Unemployment in Eastern Germany	Vocational training	Administrative Data	1999-2002	Duration of unemployment	Bivariate mixed proportional hazards model	Vocational training prolongates the unemployment duration in Eastern Germany
2004	Speckesser	Using Social Insurance Data for the Evaluation of Active Labour Market Policy: Employment Effects of Further Training for the Unemployed in Germany	Training	Social Insurance Data	1993-1995	Employment effects	Kernel matching on PSM	Significantly negative effects immediately after the beginning of the treatment with a positive trend at the end
2005	Bergemann	Do Job Creation Schemes Initiate Positive. Dynamic Employment Effects?	Job Creation Scheme	AMM-SA	1990-1999	Probability of new or reemployment	PSM with NN; CDiDHR	Only short-term positive effects for female
2005	Caliendo, Hujer, Thomson	Identifying Effect Heterogeneity to Improve the Efficiency of Job Creation Schemes in Germany	Job Creation Scheme	Measure-Participation-Data	2000-2002	Employment rates separately for sociodemographic groups, regions and sectors	PSM with NN	Positive effects for female participants, negative for male
2005	Lechner, Wunsch, Miquel	Long-Run Effects of Public Sector Sponsored Training in West Germany	Training	Measure-Participation-Data	1993-2002	Employment and unemployment rates, monthly earnings	PSM with NN	Programs have negative effects in the short run and positive effects over a horizon of about four years
2006	Caliendo, Hujer, Thomsen	Sectoral Heterogeneity in the Employment Effects of Job Creation Schemes in Germany	Job Creation Scheme	Measure-Participation-Data	2000-2002	Employment effect	PSM	Results show that JCS are unable to improve the re-integration chances of participants into regular employment.

Year	Author	Title	Program	Data	Observation period	Defined Outcome	Empirical Method	Results
2007	Fitzenberger, Speckesser	Employment Effects of the Provision of Specific Professional Skills and Techniques in Germany	Training	Administrative data	1993-1997	Employment rates	Local linear matching	Negative lock-in effect right after the beginning of the program and significantly positive treatment effects on employment rates a year after the beginning of the program
2007	Fitzenberger, Völter	Long-Run Effects of Training Programs for the Unemployed in East Germany	Public sector sponsored training	Administrative data	1993-1994	Employment and benefit recipiency	PSM	Positive medium– and long–run employment effects for the largest program, no significant effects for the others
2008	Caliendo, Hujer, Thomsen	The Employment Effects of Job Creation Schemes in Germany - A Microeconometric Evaluation	Job Creation Schemes	IEBS	2000-2004	Employment effects	PSM	Negative or insignificant employment effects for most of the analyzed groups

Declaration of authorship

I, Felizia Hanemann, hereby declare that this present thesis was prepared independently, without help from others, and without using anything other than the named sources and aids. The texts, illustrations and/or ideas taken directly or indirectly from other sources, quoted verbatim or paraphrased, have without exception been acknowledged and have been referenced in accordance with academic guidelines.

The present work is being submitted for the degree of Master of Science in Economics to the Lund University. It has not been submitted before for any degree or examination at this or any other University.

_____   _____
Place, Date   Signature